Nuno Luis Madureira

Key Concepts in Energy

Springer

Key Concepts in Energy

Nuno Luis Madureira

Key Concepts in Energy

Springer

Nuno Luis Madureira
Instituto Universitario de Lisboa
Lisboa
Portugal

ISBN 978-3-319-04977-9 ISBN 978-3-319-04978-6 (eBook)
DOI 10.1007/978-3-319-04978-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014932284

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

Abstract This book provides an encompassing, comprehensive, and cross-disciplinary overview of energy issues clustered into eight concepts. The concepts are designed to provide a bird's eye perspective of the economy, technology, and history, without losing sight of the contextual dynamic of business conjunctures, scientific trajectories, and national policies. Overall, it strikes a balance between the normative content of ideas and their troublesome, unpredictable, and contingent application by scientists, engineers, and businessman.

1.1 Introduction

1.1.1 The Time of Energy

This book provides an encompassing, comprehensive, and cross-disciplinary overview of energy issues clustered into eight concepts. The concepts are designed to provide a bird's eye perspective of the economy, technology, and history, without losing sight of the contextual dynamic of business conjunctures, scientific trajectories, and national policies. Overall, it strikes a balance between the normative content of ideas and their troublesome, unpredictable, and contingent application by scientists, engineers, and businessman. As time matters and as the time of energy proves a rather intricate web of expectations, decisions, and habits, we pay particular attention to processes and to outcomes.

Energy usage entails at least three different measures of time. At one extreme is the time of consumers, who expect to be served instantaneously simply by activating a switch or by lifting the pump in the gas-filling station. The gesture is taken for granted, because consumers have repeated it so many times and always with the same result that only when extraordinary events disturb the routine, such as blackouts caused by short circuits and overloaded electricity mains, queues, and shortcomings provoked by strikes, rationing acts issued under rising international

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_1,

tensions or infrastructure damages wrought by natural disasters, only in such disturbing circumstances do people actually realize that there is nothing guaranteed about energy supplies. From a technical standpoint, the disruption of consumption happens because energy involves network chains and has high storage costs. As a rule, the larger the dependence upon the network and the higher the storage costs, the more instantaneous consumption becomes. The main example is the operation of electrical systems where, at each instant in time, the amount of electricity generated and put on the network must exactly equal the amount taken off and consumed. Although the picture is changing quickly, up until now, there has been no way of storing electricity economically and efficiently. For this reason, the operators of electricity networks have learned to develop new scheduling algorithms for hourly generation and the transmission of electricity, for example, whenever some TV broadcast attains all-time peak ratings, and the population joins together in front of the television at exactly the same time on a cold winter night. To match this singular peak, some extra power generation must be instantaneously called upon. The phenomenon become noticed in the 1960s, when black and white TV sets spread throughout entire societies generating mass audiences for popular TV programs such as "Coronation street," "Z Cars," "Gunsmoke," "Bonanza," "The Andy Griffith Show," "Carosello," "Cómicos y Canciones," and not to mention the "Miss World" contest. As a rule, the peak in electricity demand tended to occur immediately after these programs ended, boosted by the simultaneous switching on of additional lights and electrical appliances such as kettles.

The second measure of energy duration is conveyed by the lead times for energy transformation, generation, and distribution. Any change in the existing infrastructure has to be anticipated some years previously, through forecasting techniques, so as to meet future demand. The gap between the decision to invest and the completion of the new productive unit may take 3 years in the case of an oil refinery, natural gas power station, or natural gas pipeline, 4–5 years for a coal power station, 5–8 years for an hydroelectric plant, and 6–9 years for a nuclear power station. In this context, it is likely that those who sign the first contracts are not the same parties as those later hosting the ribbon-cutting ceremony. Likewise, owing to the gap, energy infrastructure lead times are beset with the symmetrical problems of idle capital when the course of events creates underutilized generating capacity (the unanticipated slowdown in energy consumption after the first oil shock) and disturbing market bottlenecks when the course of events yields capacity shortages (gasoline scarcity and refinery outages after hurricane devastation). Aside from these non-trivial circumstances, the time pattern of energy production hinges on the dynamics of business cycle, the interest rates prevailing, the stability of the regulatory and ownership frameworks, and on the opportunities for technological innovation. Shaped by medium-term planning, this duration has moreover increasingly incorporated active community participation through public debate on the environmental and amenity impacts.

Finally, the exploitation of natural resources is embedded in the economic long term. From the outset, the prospection and discovery of underground deposits has been a specialized activity in which amateur and professional geologists replaced wildcatters, speculators, small land leasers, and men of fortune. With the advances in drilling technology, professionalization proceeded further, opening the way for the takeover of natural resources by large corporations that transformed the activity into a multinational business by the close of the nineteenth century. The trend toward concentration, prompted by mergers and fusions, reinforced the geostrategic standpoint of decision-makers, ingraining the exploitation of coal, oil, natural gas, and uranium deeper into the long haul. Recently, globalization brought business concentration to new heights, boosting international oligopolies in the areas of equipment supply, and engineering and consulting services, and further reinforcing the competitive reach of the biggest oil, gas, and electricity companies. Besides the discovery of untapped deposits, the activity of exploitation promoted the improvement of in situ geological knowledge, reducing the uncertainty about the ultimate recoverable resource. This process in which knowledge appears as a by-product of economic activity holds regardless of the differences in duration and in the productivity of underground reserves. Naturally, deposit size, seam thickness, and permeability and porosity of rocks are key factors for the economic feasibility of recovery. The wide variation in lifecycle extraction can be illustrated by the difference between the largest conventional oil field in the world, Al-Ghawār, in Saudi Arabia, with its 65 years of uninterrupted activity (and counting), and the shorter duration of some labor-intensive coal mines in historical European pits, which were exhausted in <10 years. Whatever the case, forecasts of future production tend to be much more precise in this upstream activity than in other branches of the industry. Equally significant is the fact that natural resource exploitation has proven quite sensitive to the political economy of international relations and affected by wars, guerrilla warfare, revolutions, coups, and governmental churn.

All of the above time measures—current consumption, future production, and long-term extraction—are entwined in the daily supply of energy goods and services. In the starkest terms, they correspond to the accounting of useful, final/secondary, and primary energy, which means that the more we move back in the process of energy transformation and conversion, receding from final consumption to downstream and to upstream activities (miles travelled to gas and to crude oil), the larger will be the inertia within the techno-economic system and the greater the reliance on long-term commitments. It is the cross-temporality of expectations, decisions, and habits that deeply roots energy concepts in history.

1.1.2 Winners and Losers

It was mostly during the nineteenth century that mankind systematically began to tap into the legacy of carbon and carbon and hydrogen compounds accumulated beneath the earth over the course of 360 million years. Courtesy of the practice of drawing assets from this stored energy bank, developed nations were each year able to add fresh inputs to their renewable resources of muscular force, biomass, wind, and water power. What is more, large-scale usage of fossil fuels did not simply mean additional “raw inputs” for productive activity but was historically associated with a far-reaching technological–economic shift characterized by a continued increase in the useful work (technically called exergy) extracted from each unit of input, a change in the composition of economic activities, a change in the distribution of population, and enhanced flexibility in energy supplies.

The beginning of the new civilizational era spearheaded by hard coal, brown coal, petroleum, and later by natural gas posed new challenges for economic organization as it shifted the focus from the useful power of traditional energy carriers toward the chain of energy conversions produced within machines (the reciprocal transformations of heat, mechanical work, and electrical magnetism), thus linking the fuel inputs with the useful energy obtained. Thanks to this, physicists and entrepreneurs realized that processes hitherto separated could be brought together under the same roof: Motion, heat, light, electricity, gravitational force, and chemical energy were but different physical manifestations of the “ability to do work.” Each one presented its own advantages and its own drawbacks. The trade-off between advantages and drawbacks could be abridged in the idea that energy entailed different qualities. As fossil fuels progressively replaced the demand for renewable energy sources, fostering the ability to do work with ever fewer resources, two types of downsides were increasingly targeted: The first was the impairment of the landscape and the unruly emission of “coal smoke” and water pollution and the second was the doubt about the future availability of fossil fuels and the risks of running out of energy resources.

Whatever doubts there might be on pollution or security grounds, the debate was always framed by the regional or national territory. Sometimes, it aroused passionate campaigns in pledges to defend the wilderness, the scenery, public health, resource conservation, efficient public management, and techno-scientific sound solutions. On other occasions, as in late comers to industrial development, environmental degradation was not even recognized as a problem but rather as the negative side effect of heavy industrialization, of the execution of megaprojects in infrastructures or military investments designed to catch up with the West. Only in the 1980s did generalized concerns about acid rain set in motion the call for a more multilateral approach and transnational system of organization designed to face boundless environmental problems. Since then, global warming turned the world into the global village that Marshall McLuhan had foreseen. Under the twin pressures of CO_2 emissions abatement and biodiversity conservation, energy became the key issue for medium-term sustainability. To overcome upcoming temperature increases caused by atmospheric concentrations of carbon dioxide and radioactively active gases, a wide range of opinion and scientific research arrayed around the agenda of the transition to a low carbon economy. What the expression meant is a substitution process, away from fossil fuels toward renewable energy or new energy technologies such as hydrogen fuel cells or electrical batteries. Soaring oil prices and forward investments of large corporations pushed, in recent years, some of these technologies to cutting-edge development and to the brink of commercial viability. The noteworthy point, however, is the resumption of technological paths that have been overlooked for decades or left in the lurch of marginal R&D.

Things formerly considered technical failures, dead ends or outright uncompetitive ventures, witnessed a sudden resurrection. Ultimately, paths not taken in engineering and business proved to be feasible at a different point in time, with different cost incentives.

Traditional windmills attained their peak during the nineteenth century in the Netherlands, Britain, and Scandinavian countries, and experiencing a second wave of diffusion of smaller, easy to operate and repair, water-pumping Halladay-type windmills, in the new world (USA, Australia, South Africa). The largest types reached a capacity of 7.5–30 kW, but fell into what appeared an irredeemable disuse with the advent of cheap fossil-fueled engines and widespread rural electrification. After decades of oblivion, the rebirth of wind turbine technology based on lightweight blades, a rotor that rotates the generator and a gear box enhanced the capacity to extract energy from the wind in the range of 500–4,500 kW. Moreover, world wind capacity has doubled every 3 years since 2001.

By the same token, the phasing out of electrical vehicles after their manufacturing peak before World War I foreclosed ongoing improvements in electrical batteries, the development of a standard charging plug by electrical car companies, the installation of networks of recharging stations, and competitive promotional exhibitions in car races. Years of technological eagerness in which outstanding cars proved possible to drive 130 miles on one battery charge, well above the epoch standard of around 25 miles, were brought to a halt by the blow struck by the internal combustion engine. Up to the 1970s, the electrical car was no more than an odd curiosity in museums and its industrial manufacture dropped to zero. The staggering success of electric hybrid vehicles from 1997 onward, particularly in Japan, North America, and Europe, brought to full blossom a commercial technology whose lockout had seemed insurmountable. Based in the development of nickel metal hydride and lithium batteries instead of the former prototypes of lead–acid batteries, the new electrical hybrid vehicles point to autonomy in the average range of 85–125 miles with the ability to fully charge a battery pack in under 10 min, regardless of the greater lifespan. Additionally, the major cost component, the battery, has decreased greatly in price over the last decade, and this trend is expected to continue, fostering their overall competitiveness with gas-fueled cars.

The same reasoning applies to technologies that were deemed uneconomical and have witnessed a return to the R&D agenda in recent years. That is the case of the coal-to-liquids (coal liquefaction) technology, in which coal is transformed into a synthesis gas ("syngas"), mainly consisting of carbon monoxide (CO), hydrogen, and carbon dioxide (CO_2), via coal gasification before it is converted into synthetic fuels or synfuel. Synfuels are used as an alternative to oil and allow the production of petroleum, diesel, synthetic waxes, lubricants, and chemical feedstock. Promoted by Nazi Germany to escape its geostrategic weaknesses in the access to international oil sources, the production of synthetic fuels reached the unprecedented levels of near 90 % of Hitler's needs, despite the prohibitive and uneconomic costs of the industrial process. Since then, the South African Coal Oil and Gas Corporation (Sasol) has been operating the world's only commercial coal-to-liquids plant. Though still at a conceptual stage, recent research has proved the economic feasibility of coal liquefaction processes, provided that the installation of carbon capture and storage technologies might offset the high levels of carbon dioxide (CO_2) emissions.

Under this eager conjuncture, former dead ends imparted a fresh start for technological developments and decades of standstill gave way to shifts that unfolded at breakneck speed. Time, scientific advances, and markets turned chronic energy losers into winners. As it turned out, evolution seemed to refute the idea of progress as a move ahead toward the untested and recast it more as the "eternal recurrence" of identified technical possibilities. The lesson to be drawn from the bouncing back from failure and from the overcoming of technical deadlocks is that one cannot assess the relevance of events just by their immediate consequences. A strict consequentialist view of engineering and business progress entails the risk of neglecting the real open possibilities, the contingencies, the turning points, and the branching points that actors and technologies had to face across history. All the more so, when time operates a poignant devaluation of the relative epochal significance of things with few discernible consequences. As a rule, discontinued paths of development tend to be expunged from their institutional, documental, and statistical "footprint" to become nearly untraceable.

The ensuing pages provide a balanced view of the winners and losers in energy's technological development, of leading sectors and market niches in the economy, of cutting edge innovators, and late followers in science and engineering, highlighting what happens with discontinued paths of development. The concepts of "technological hybridization," "last gasp," and "probable reserves" exposed respectively in Chaps. 4, 5, and 6 further illustrate this perspective.

1.1.3 Key Concepts

The eight concepts addressed in this book build up an understanding of the central ideas to energy in the contemporary world. Each chapter corresponds to a single concept and is presented independently, although formally repeating the same structure: An initial section introduces the normative definition of the concept, along with analysis of its theoretical foundations and practical implications; the ensuing sections describe the historical emergence of the idea; the last section finally provides a comprehensive overview of the advances in current research. From a thematic point of view, each organized chapter delves into topics such as energy measurement and the thermodynamic explanations for energy transformation (Chap. 2), the life cycle of technological innovation, consumption, energy efficiency, and resources depletion (Chaps. 3–6), and alongside the issues relating to economic calculus (Chaps. 7–9). Broadly speaking, this boils down to three core areas: scientific breakthroughs; technological transitions; and prices and costs. The sequence moreover reflects the historical timeline so that the book is arranged to properly progress in chronological order. It is therefore reasonable to expect more references to coal at the beginning and more references to nuclear energy at the end. Likewise, microeconomic developments are watched closely in the opening chapters, while macroeconomic policy deserves more attention in the final chapters.

Indispensable to the replication of knowledge, definitions represent points of arrival in the scientific debate. Easy to understand and to quote, they provide the normative digests of rules, taxonomical specifications of properties that members of a class tend to possess, prescriptions or proscriptions for action, and stylized facts drawn from laboratory experience. Between the emergence and shaping of concepts and their hardening into "point-of-arrival" definitions, there is however a commonly bumpy journey. Usable and "ready-made" definitions generally wipe out a long history of qualms, confrontations, and gatherings that preceded its unquestioned recognition. To mention only the most high-profile cases referred to in the forthcoming pages, concepts put forward by men of science like James Prescott Joule, William Stanley Jevons, or Marion King Hubbert triggered vigorous rebuttal from other experts and from industrial interests. It is this sparkling context to shaping and establishing the concepts of energy that the core of each chapter discloses. To widen the perspective on what is at stake here, a balanced appraisal of the different strands of thought is put forward alongside fairly detailed bibliographic references. By this means, each chapter is intended to work as a repository of updated information on each topic and serve as a basic guide to the reader.

Since the book rests substantially on original research, it correspondingly owes a lot to those who have made this possible supporting the project through different means. I am deeply grateful to the Calouste Gulbenkian Foundation for the two grants awarded for archival research, to the Foundation of Science and Technology and the Fulbright Commission for fellowships funding, and to the Center of European Studies for the recruitment of a research assistant at Harvard University, Yenning Qin, who has done an outstanding job in data collection. Bruno Cordovil has also carried out deep reaching research in the National Archives, London, which proved of great assistance. I am furthermore grateful to Dr. Philippe Dreano and his assistance in accessing and consulting the EDF Archives at Blois, France, as well as the archivists from the British Petroleum Archives at the Warwick University. The revision of the final text has had the fortunate aid of Rachel Evans and the precious improvements suggested by Kevin Rose. I am also indebted to Nuno Pimentel and to the anonymous referees of the Environment and History, Journal of Global History, and Energy Policy journals that provided me with the benefit of their expert comments. This appreciation is extended to the White Horse Press, Cambridge University Press, Elsevier, and Chicago University Press for the permission to reproduce published articles and images. Notwithstanding all this helpful support, the author is the only responsible for any flaws that readers might find in the text.

Articles herein partially reproduced following due permission:

- The anxiety of abundance: Stanley Jevons and coal scarcity in the nineteenth century. (2012) Environment and History 18(3): 395–421.
- The iron industry energy transition (2012) Energy Policy. Special Issue: Past and Prospective Energy Transitions, Roger Fouquet & Peter Pearson (50): 24–34.
- Oil in the age of steam (2010) Journal of Global History 5(1): 75–94.
- The European debate on rate systems in the interwar period (2010). Energy Policy 38(8):4703–4711.

Chapter 2
Primary Energy/Useful Energy

Abstract While the useful energy delivered by an animal, a windmill, or a worker captured most attention in the preindustrial era, the fossil fuel economy of the nineteenth century shifted the focus toward the energy conversion cycle. Only by following the order of transformations and only by measuring just how much was conserved at each stage could the aggregation of different energy qualities produce meaningful results. This chapter shows how the invention of new units of measurement, such as horsepower and kilocalorie, together with the principle of energy conservation (First Law of Thermodynamics), paved the way for a system of energy accounting based on four input cycles: primary, secondary, final, and useful energy.

2.1 Energy Measurement

2.1.1 Input-Cycle Categories

During industrialization, energy fell under the measuring rod of money. The ability to do work was detached from the indivisibilities of everyday life and turned into a separate commodity, tradable over ever longer distances. Contrasting the traditional fixity of the historically longstanding power sources such as man and animal power or wind and water power, the steam and internal combustion engines ushered in an era of greater mobility where fuel sources did not necessarily overlap with the sites of power generation. Reflecting this commodification and furthermore the underlying need to measure the potential energy contained in the chemical energy of fossil fuels along with the work effectively performed per unit of time, industrial societies devised the solution of gauging the flow of energy through its contribution at different moments in the industrial cycle. Four major sequential "cuts" fell under consideration:

- The first, **total primary commercial energy**, accounts for the raw inputs supplied to the economy and represents the sum of all market energy sources arriving at the entry of energy systems during a single year and expressed in terms of

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_2,

production equivalents. This comprises fossil fuels and renewables but not fuel wood on the grounds that much of wood consumption still bypasses the scope of formal market transactions.

- The next, **secondary energy** designates all sources of energy that result from the transformation of primary sources in any one single year and generally termed energy carriers.
- The third, **final commercial energy** accounts for secondary energy products and services delivered by different energy markets to final and intermediate consumers, and again over a single year, and comprises all transformed products such as charcoal, coke, coal, natural gas, gasoline, fuel oil, and electricity, among others.
- The fourth, **useful energy** corresponds to the energy effectively made available to the user in terms of the services delivered through end-user equipment over the course of one single year and expressed in terms of consumption equivalents for the work performed by mechanical power, lighting, heat generation, and travel mileage.

Much of the scope of the input-cycle categories stems from a hands-on approach in which the sequence of categories mirrors the transactional structure in effect in the real economy. Thus, this involved efforts to identify primary energy with upstream activities, accounted for by physical units and not only by energy units; secondary and final energy with the midstream and downstream handling of raw materials by industrial wholesalers and distributors that select, clean, transform, and transport the fuels to their final markets; and, useful energy with the decentralized consumption of heat, mechanical power, light, electrical current, and chemical energy. Owing to their capacity to replicate the logistical pattern of the industrial structure, input-cycle categories incurring low information costs were consequently acknowledged as the official statistical criterion by key international organizations.

One final advantage is that any losses incurred throughout the aforementioned processes are easily identified. It became expected that a significant amount of fossil fuel contained energy would get lost in transportation and in transformation into secondary products and thereby diminishing the volume of final energy relative to primary energy inputs. Likewise, impairments in final distribution in addition to conversion losses of end-user equipment also cut down the amount of useful energy relative to final energy. In the starkest terms, this process serves to define energy efficiency as the differential between primary energy and useful energy.

However, input-cycle categories are anything but problem free. Globally, the further we move away from the classical pattern of fossil fuel combustion, the less engaging the model becomes. It suffices to consider, for instance, the awkward fit of renewable energies within the parameters of the input-cycle framework. What is the "primary energy" of a wind turbine? And just what is the "primary energy" of a hydropower dam? How should we measure them? The standard solution energy economics conceived for these cases display the drawbacks of aggregating different energy qualities: measuring a dam's primary energy by envisioning it as if a coal power station and representing the kilowatt hour produced by the hydraulic turbines in thermal equivalents for the amount of coal required generating that amount of power. In other words, think black on white (the expression "white coal" constituted a token of the kinetic water stream in hydropower stations). Clearly, one facet to the statistical paradox stems from the fact that input-cycle categories were never designed to grasp intermittent power sources, renewed by the passage of time and, ultimately, by the solar cycle, able to convert mechanical force or heat into electrical current. They had instead been drafted to explain how a tradable organic material might be burned in the presence of oxygen to release the products of carbon dioxide, water, and thermal energy. Whenever energy quantities related to kinetic energy (e.g., hydro, wind) differ in quality from energy quantities related to potential chemical energy (e.g., fossil fuels), the solution envisaged is to express everything under the common formal framework of reference for chemical energy.

The same problem arises in the statistical accounting of nuclear power, further worsened by institutional discrepancies: while some international organizations record primary nuclear energy as the amount of electricity produced through atomic fission, others record primary energy based upon the actual nuclear fuels in use. As the two methodologies are not directly comparable, they readily become a source of error (Lightfoot 2007). Once again, the alternative lies between reasserting the analogy of nuclear and coal power stations (accounting for the energy content of electricity) and estimating the raw inputs supplied to the economy (accounting for the natural uranium consumed). Beyond the large discrepancies between the methodologies, the most serious issue derives from potentially misrepresenting the effective efficiency of energy systems since one dominant technology (generally chemical energy based) subsumes others technologies for statistical purposes. We would note, for instance, that when the energy produced by water falling in a dam is estimated in terms of equivalence to coal burning stations, the kinetic efficiency of turbines is swapped for the standard chemical conversion efficiency of steam boilers. On balance, there is a trade-off between the sheer handiness of decomposing the energy cycle into the statistical profile of primary/secondary/final/useful energy and shaping every energy process according to the yardstick of chemical energy.

To overcome these input-cycle category shortcomings, Ayres (1998), Ayres et al. (2003), Ayres, and Warr (2005) suggest a whole new methodological framework. As an alternative to basing them on the industrial logistical cycle, they propose singling out the useful part of energy, or "energy," as the core factor of analysis. In some way, this corresponds to inverting the traditional view as it begins where classification should supposedly end and then proceeds backwards along the cycle: the end node is nothing other than the production of energy for services employed in manufacturing, transportation, space heating, lighting, cooking, and so forth. Along with the statistical landscape of the different types of useful energy, Ayres and Warr also undertook research on the primary work of production equipment, disclosing their breakdown and efficiency in energy usage. In fact, this outline by equipment type enabled the authors to show just how much energy is lost in the process of converting it into useful work. Overall, six major statistical categories are considered: muscle work, electricity, mechanical drive, low temperature heat, medium temperature heat, and high temperature heat (Warr et al. 2010). Each form of energy allocated to types of useful work might be further specified, for instance, taking into account the separate contributions of steam engines, electric motor drives, and internal combustion engines in automobiles, trucks, and buses, within the more general category of "mechanical drive."

The inclusion of useful work as a factor of production representing the productive component of energy inputs seeks to pinpoint the impact of energy quality upon economic growth. Since part of the total available consumed energy is actually wasted and does not contribute to growth, it thus makes statistical sense to weigh the other part, the working "lot" or energy.

2.1.2 The Measuring Rod of Money

Overhanging the above considerations is the deep intertwinement of useful work and technology. The way individuals behave, consume, and save hinges upon the catalog of accessible technological choices but also on the infrastructures of knowledge, habits, and physical networks. No less important, it hinges upon available capital. Given this association, the long-lasting disregard that the useful facet of energy consumption has experienced proves somewhat startling. The ensuing pages explain this oversight. Taking a broader view, the key reason is that the costs involved in assembling data on useful energy are considerably higher as this requires mapping end-user production equipment, which is likely to present different temporal layers and competing operational conversion efficiencies. Another way of expressing this is to outline the "end uses as the set of useful functions expressed by a given society, in its various compartments and on different scales" (Giampietro and Sorman 2012: 8). While primary energy typically portrays indicators from national energy accounting processes, useful energy stands for local, decentralized, and consumer-centered indicators. As such, this raises far more specification problems than the bulk account provided by primary physical gradients or secondary energy carriers.

To a large extent, nineteenth-century machinery, science, and institutions led first to the uprooting of energy measurements and ultimately to the globalization of standard units. Step by step, energy ceased being measured in terms of units of a certain commodity obtained from the application of power for a fixed timespan (bushels of milled flour per day, pounds of water leveled per hour, miles of journey covered). Step by step, mechanical power became assessed independently from its practical applications, standing for the force needed to accomplish work rather than as the materialization of useful work itself. As these units sprang from being local conventions, based on discretionary and negotiable standards, to represent the new edge of commercial, national, and universal standards, they became increasingly disembodied from the technologies of everyday life.

Uprooting means detachment from personal settings and local habits. In the modern world, measurement required specialized "bazaar" skills with ample room for negotiation, compliance, influence, coercion, and fraud. The simple determination of the volume of a heap of grain or the weight of a particular bushel had long since proven a tricky business in most of the town hall markets across northern Europe. Kula (1986) testifies to how the regime of negotiation and the discretionary characteristics of weight and volume units were clearly set to favor local interests over central powers and those holding privileged statuses over commoners. In the same vein, Pollard (1983) identifies the persistence of the British tradition to market coal by volume rather than by weight, a procedure that allows for greater variability in measurements, enhancing the trader's leverage and his power to establish beforehand how much coal a basket, a tub, a score, or a wagon should have. Such a preference for units of volume add further margins for haggling, permitting both parties to the transaction to adjust the deal in terms of prices as much as in terms of quantities. Trust and local knowledge were integral parts of the business.

The advent of the steam engine ushered in a change from fuzzy negotiable measurement units. Energy was progressively disconnected from the actions of everyday life, from community habits and turned into a separate commodity, with prices pegged to objective and unbiased measurement units of power. This marked the beginning of the process described above as the "measuring rod of money." Not by accident, the first commercial quantity unit used to gauge mechanical force was dubbed "horsepower," that is, an artificial muscle engine fed by coal rather than by hay.

2.2 Horsepower and "Calories"

2.2.1 The Power of Royalties

Once entrepreneurs succeed in commercializing breakthrough technologies, the catalog of final energy measurements gets increased. Entrepreneurs are compelled to move their products or services to final markets and to convince the customers about the power they are selling. Under these circumstances, it is no surprise that the specification of outputs is generally much more accurate, detailed, and highlighted than the information on inputs. Ultimately, power measurement units can be squared with some marketing scheme and serve as a motto for sales. Such close linkage between selling and the creation of final energy measurement units began with James Watt's invention of horsepower in the eighteenth century.

The walk undertaken by the Scottish engineer in the green fields of Glasgow nearby his house, on a day-off in May 1765, turned out to be one of the most productive walks ever taken by mankind. For the last few months, this self-educated engineer and instrument maker had been working on a solution to improve the atmospheric engine that had come into his hands, trying different ways to economize on steam and reduce the sources of heat loss. Watt was particularly puzzled by the discovery that about three-fourths of the amount of steam used at each stroke of the piston's engine was wasted by the inherent dissipation of coal. The short stroll did not tear the technical quandaries from his thoughts and, after making some short headway into the field, he stumbled on what appeared to be a solution: "having gone as far as the herd's house, the idea came to my mind that as steam was an elastic body, it would rush into a vacuum, and if a communication was made between the cylinder and an exhausted vessel, it would rush into it, and might be there condensed without cooling the cylinder" (Henry 1902). Instead of piecemeal improvements in mechanical engineering tried and tested for so long, Watt envisioned the scope for a breakthrough and an entirely new device. Whereas in the old Newcomen atmospheric engine, it was necessary to heat up the cylinder with steam during the up-stroke and then cool it down by opening a valve that let fresh water pass through a pipe into the cylinder, creating a vacuum for the down stroke, it now became evident that the best way of reducing fuel consumption was to save steam by keeping the piston cylinder constantly hot. This might be achieved by separating the working area of the piston cylinder from a new cooling area in an "exhaust vessel," where the steam would "rush in" and condensate.

Enlightening as the glimpse in the Glasgow greens may have been the fact remained that it still took over a decade to set up a satisfactory working engine. Watt quickly discovered that he had got himself into something far more difficult than he had ever thought: it was an endeavor that required full-time dedication; a highly skilled team of craftsmen able to build, fit, and insulate the engine's components; long-term capital investment; the defense of property rights through patent litigation; and not to mention commercial insight. Furthermore, the early attempts to build a prototype produced disappointing results owing to flaws in the construction of the condenser and steam piston. With the project at risk and the inventor near bankruptcy, a turnaround occurred with the acceptance of a proposal from an experienced Birmingham manufacturer, Mathew Boulton, to take an interest in the patent's development. The new associate not only enhanced the technical and economic basis of the venture but paved the way for one the most productive "joint ventures" in the history of business.

Watt moved to the outskirts of Birmingham in 1774. Once the fundamentals for the commercialization of working machines were set, orders for new engines with separate condensers began pouring in. With Boulton taking charge of general management, business relations, the payment of all expenses, and the advances of stock in trade, the Scottish engineer was unleashed from a good number of day-to-day concerns and able to dedicate himself to raising his invention to still higher standards. In the years ahead, he devised a parallel motion gear that allowed steam to be introduced into both ends of the cylinder to push the piston both up and down; a transmission mechanism that converted the reciprocating motion mostly used for water pumping into a rotary motion in effect in textile mills, breweries, and other industries; a system for shutting down the admission of steam into the cylinder part way, thereby enabling the use of steam power expansively; and a regulatory device that provided closer control over the number of strokes per minute and boosted the smoothness of the mechanical force generated (Hills 1993).

In the meantime, steps were also taken to secure exclusive rights to Watt's registered patent of the separate condenser for an additional period of 24 years (1776–1800), followed by five new patent rights which covered the latest innovations. Sheltered behind this recognition sanctioned by Acts of Parliament, Mathew Boulton devised a strategy to branch out the sources of income: they could earn money either from selling state-of-the-art equipment manufactured under their own supervision or from selling licenses to apply the basic discoveries behind the equipment, thereby extracting a rent from the patent's exclusivity rights. As Watt's engine provided fuel savings when measured up against an improved atmospheric Newcomen engine, Boulton could make his case on the commercial ground of charging an annual royalty based on comparative operational advantage. He, therefore, stipulated that one-third of the value of fuel savings should revert to the owners of the patent in the form of an annual "premium." Through this ingenious scheme, Boulton and Watt might still recover their development costs even while not selling the complete machine but only certain component parts. On the other hand, the greater demand for all engine types, whether atmospheric or steam based, the more justified and less risky became the decision to charge manufacturers switching to up-to-date technologies with an additional payment (an average Watt engine generated another 40 % of the purchase price through 10 years of royalty payments) (Kanefsky and Robey 1980; Tann 1998). The stiff competition meant that manufacturers and mine leasers could choose between simpler and cheaper atmospheric engines but with their higher levels of coal consumption or switch to the more complex separate condenser engine and pay the respective premium for lower fuel expenditure. In either case, the noteworthy point is how the dynamic economic "boom" of British industry after 1770 brought the issues of fuel saving and power measurement right to the forefront of industrial concerns.

The most acceptable way of singling out the amount to be annually paid for a Boulton and Watt engine consisted of setting the royalty in direct proportion to the engine's respective power. The new pricing scheme thus changed the meaning of measurement: whereas estimating power initially served as a planning device to satisfy customer needs for mechanical force, matching each machine to its load, within the new business orientation, it, moreover, became a commercial financial instrument tied to the value of royalties. Furthermore, since the power capacity determined the scale of payments, a corresponding scale of mechanical force with comprehensible, unambiguous, and verifiable units had to be established. Driven on by the increasing need for accurate measurements, Watt worked out a standard figure for his engines based on the cylinder diameter and on the stroke, and assuming an average for atmospheric pressure. To keep a record of the number of strokes, he also developed a "counter" logging the seesawing of the main beam (Hills 1993: 88). The final step was simply to link up the design of the machines with a system of power units. However, just what units to consider? This could be a particularly cumbersome issue in the case of rotative engines, whose power was not easily observable in terms of a weight being lifted or water being pumped in the course of a day from a given depth. The rotary motion had different applications to the reciprocating-pumping ones and the installation of rotative engines involved seldom pulls and gears, with larger losses in transmission and friction. In these circumstances, measuring power by some indicator of the useful energy seemed unsuitable. On the other hand, a measure of the final gross energy delivered was available and had been put into practice for several years: Boulton and Watt knew that whenever an engine was introduced to replace horses in a mill or in a brewery, a direct comparison was made between the size of the engine and the number of horses (Hills and Pacey 1972). Hence, the option for building engines in terms of horse substitution surfaced naturally as the best possible commercial standard and as also providing a clear blueprint for gauging the royalties due for payment.

2.2.2 Blood Horses and Mechanical Horses

Assessing steam horses by the standard of blood horses proved to be a shrewd commercial choice. The new technology was folded back onto the old, thereby rendering it more recognizable or, as a contemporary observer might add, reducing the information search costs for clients in remote locations. Certainly, any reference to "horsepower" in the context of eighteenth-century British industry was not a free standing outline but involved some shared knowledge about the type of horse supposedly found in factories: in all likelihood, to James Watt, the mechanical unit was the embodiment of a heavy horse, farm bred and sold for heavy draught work at the age of four or five and in service until around the age of fourteen. This also focused on the strongest specimens among the breeds, those able to deliver 40–50 % more productive power than a mule or a light horse during a full day's work. In short, the engineer pulled out his horsepower unit from the finest available among the thoroughbred blood horses.

With burgeoning demand for rotative engines from 1783 onwards, the loose representation of "horsepowers" needed to be unambiguously specified. Once again, James Watt resorted to real-life working conditions and drew upon the example of a horse attached to a mill that walked in an endless circle with a 24-foot diameter. By multiplying the speed times versus the force, he came up with 32,580 foot-pounds per minute, subsequently rounded off to 33,000 foot-pounds per minute. Although well above other historical estimates for the average force exercised by a heavy animal, which ranged between 22,000 and 25,000 foot-pounds per minute (with a single exception of 27,500 foot-pounds per minute) (Thurston 1894; Hills 1993; McShane and Tarr 2003), Watt's outsized figure remained the standard for gauging the absolute force of "horsepower" through to current days. This means that, in real terms, a heavy draught horse from the English shires, French percherons, or German rheinlanders was only able to deploy a force of 0.7–0.8 mechanical horsepower/hour, while a light draught horse would attain significantly less, reaching 0.5 horsepower/hour (Thurston 1894; Smil 1994). We would note that the point here is not the theoretical amount of blood horsepower capacity, but rather the average amount developed throughout a day's work. As every experienced horse-drawn teamster was aware, after 4 h of continuous haulage, the power of the horses dropped significantly, therefore leading to the provision of frequent stops for water and rest so as to avoid premature exhaustion. Though the best equine species could develop the equivalent of three mechanical horsepower for brief periods, the real force deployed in a 6–8-h workday (urban–non-urban haulage) would be about four times less, in the range of 0.7–0.8 HP (horsepower). Watt's oversized estimate of the power generated by horses is even more striking since steam engines were operated for between 10 and 12 h per day, the customary manufacturing working timetable.

However, and revealingly, a private business unit of measurement settled in Birmingham sprang up over large parts of the world as a trustworthy convention, shaping transactions and shaping the scale of observations made by economic actors. Yet, Watt's account did not come even close to a reproducible standard as other assessments of horses as prime movers arrived at very different figures to Watt's 33,000 foot-pounds per minute. Overall, this was a convention without agreement: a blueprint that restated the commercial superiority of the Boulton and Watt engines while providing an acceptable gauge for charging royalties. It was a business and marketing standard as much as a descriptive experimental measurement.

In explaining the passage from local units to this global unit of measurements, two major factors stand out: market power and the shock of innovation. The first stemmed not only from Boulton and Watt's technological superiority but furthermore from the active stance taken in defending their own patent rights, with the business partnership sparing no efforts to protect their sources of income against illegitimate imitators. By the early 1790s, an organized team of industrial spies had been deployed out working in the field, crossing the most remote spots of Lancashire to collect affidavits and inspect possible pirate engines while also keeping their investigations close secrets so that solicitors in London could prepare court injunctions against the pirates (Musson and Robinson 1959). Raising the Boulton and Watt flag, the next generation of Watt Junior and Matthew Robinson Boulton, took hold of this type of surveillance–punishment operations. These sorts of endeavors strengthened the identity linking the engine with the separate condenser and the business firm which put it into production, at least through to the expiry of the patent in 1800. And since the unit of measurement, horsepower, was imbued in the machine through technical and financial specifications, it spread unaffected alongside Boulton and Watt market power. Likewise, the damage of piracy was smallest in the international arena (Tann and Breckin 1978). Watt's engine benefited furthermore from the shock of innovation and novelty getting the attention of entrepreneurs and scientists through correspondence, visits and introductions, universities, academies, and foreign publications. The search for information combined with the eagerness to understand the engine's details proceeded ahead of adoption, thus creating an absorption gap that effectively underpinned the assimilation of technical principles, such as the horsepower unit (Tann and Breckin 1978). Hence, this impregnation proved so strong and convincing that the principle of "horsepower" persisted as a self-evident convention for describing mechanical work.

2.2.3 The Changeover to Heat

Alongside the penetration of the steam engine into manufacturing, mine, and transportation, there also came about a renewed interest in its potential as a scientific and experimental device. It was particularly among French engineering schools that such interest took hold. With a solid reputation for mastery in several domains, the French engineers perceived the steam engine as a puzzling innovation: the British had invented a machine that used coal to produce the "fluid" or "substance" of heat and, moreover, devised a system for circulating this "substance," thereby pushing the piston to generate a continuous mechanical force. As the largest European importer of Boulton and Watt Engines by the end of the revolutionary wars (1815), they were fully aware of the machine's productive efficiency and its wide scope of potential applications. The spurt of tests, research, and theory that ensued consequently focused on practical improvements to its component parts. The optimization path ranged from very specific mechanics, such as determining the optimum piston position or the maximum potential cylinder speeds, through to more ambitious modifications such as searching for a better heat "carrying" medium (for instance, alcohol instead of water) (Kerker 1960). However, the most ambitious agenda was set by the chemistry chair at the Conservatoire des Arts et Metiers in Paris, Nicolas Clément-Desormes.

Distinguished by a highly successful line of manufactures, Clément-Desormes encountered the steam engine prior to his Conservatoire appointment in November 1819. Unlike his colleagues, Clément-Desormes seemed not so interested in optimizing mechanical work but rather in extending the technological possibilities of energy production. From his perspective, the crux of the matter lay in the transition between "states of matter": what maximum effect could be obtained from a given mass of different working substances under various conditions in terms of temperature and pressure? What might be the maximum amount of energy obtainable from 1 kg of coal? What might be the maximum amount of force obtained through the expansion of steam given an absence of any heat transfer? Under his scientific perspective, the hypothesis of an ideal steam engine began to surface: nothing less than a machine bound by scientific engineering to enable operation at or near the theoretical maximum levels of efficiency (Fox 1970).

Before adventuring into such an ambitious program, some preliminary steps in measurement had nonetheless still to be taken. From the conversion's sequence point of view, the coal input could be calculated in kilograms. Likewise, the stroke capacity per unit of time could be estimated in the force deployed in kilogrammeters of work, which was more precise for "laboratory" estimations than the commercial horsepower unit. However, somewhere in the middle, at the heart of the engine, lingered the largely immeasurable steam sequence. To map the full batch of operations inherent to coal-steam mechanical workings, Clément-Desormes had to figure out a unit for the heat delivered by steam. He accomplished this by inventing the "Calorie."

Coined after the French word "calorique," or heat substance, the "calorie" described the amount of heat needed to raise the temperature of 1 kg of water by

1 °C. Following the definition of that which is now called the kilocalorie, all the transformations within the steam engine could be represented in a straightforward manner, through algebraic relations between outcomes and original conditions (for instance, work as a function of the temperature and pressure of the piston and the condenser). In the ensuing decades, the study of heat, closely tied to the steam engine paradigm became the keystone in the advance of physics (Hartman 1982). However, contrary to the commercial recognition of horsepower, the calorie remained a "private" category for scientific research, confined within the walls of universities and conveyed through hand-written notes and dictionaries of science, acknowledged mostly by chemists, physicists, and their students. It was, moreover, largely bound to the French academy and any non-nationals who might be in touch with French teaching. In the sense of being a unit of heat, the calorie only entered the German vocabulary in 1848 and the English in 1863 (Hargrove 2007). The consequence was that other heat measurement standards were adopted in the interim and standards which shied away from the metric system and the "French" calorie unit (henceforth referred to by the modern designation of kilocalorie—kcal). On the other side of the channel, British physicists devised their own unit outside of the metric system, adopting as their standard the amount of energy needed to heat one pound of water 1 °F (to recall, the calorie's definition—the amount of heat needed to heat 1 kg of water by 1 °C). Later on, this standard came to be known as the British Thermal Unit (BTU). As the British thermal unit represented 0.251 of the "French" kilocalorie, the scale of measurement applied by British physicists allowed for representing heat changes in smaller units.

In any case, for those studying in Clément-Desormes' classes, a new conceptual insight was acquired on the steam engine. The engineering components were henceforth linked through a set of linear transformations disclosing the energy potential contained in fuels and the final energy actually obtained: students learned that the heat content of 1 kg of charcoal (7,050 kcal) could be applied to evaporating water, spending 650 kcal of charcoal for each kilogram of water converted into steam, which would finally generate a total amount of 300,000–400,000 kg-m of work. Thanks to the common denominator of the kilocalorie, the sequence of transformations became measurable, comparable, and abbreviated; and thanks to this, Watt's industrial engine could be envisioned as a conversion machine. Final energy was by this means related with primary energy and, power at hand was related with the sources of power.

2.3 Omnipresent Energy

2.3.1 Energy Conservation

Just as horsepower entrenched the rod of money in nineteenth-century measurement, the kilocalorie amplified the experimental scientific approach. With the usage of the steam engine outside the industrial milieu as a laboratorial tool, the conversion of heat into kinetic energy became the overriding issue for European physics. Staying close to the path opened up by James Watt, French engineers attempted to unveil the principles behind engine conversion processes. Twenty-five years later, another researcher, however, proposed the inverse possibility of converting mechanical work into heat. In the event that the amount of heat obtained might be approximately the same as the amount of kinetic energy applied, then the energy conversion could be deemed indifferent to whichever energy switch was employed. Thanks to this standpoint, the issue of relative conversion efficiency (Clément-Desormes) was transformed into the groundbreaking theme of energy conservation.

Hitherto, the overwhelming problem had been measurement itself. The capacity to single out small, almost unobservable variations in temperature proved to be a decisive step toward studying the conversion of energy and dealing with the challenging problem of heat dissipation. Only by resorting to a highly controlled experimental environment and extremely precise instruments could such experiments succeed. Science turned toward the measurement of infinitesimal variations well ahead of progress in the state-of-the-art in scientific instrument making. In these circumstances, groundbreaking discoveries able to pave the way for the understanding of heat in terms of its electrical, chemical, or thermal properties had to be undertaken by someone capable of mastering the construction and calibration of scientific instruments alongside skillful expertise in laboratorial observation. In either field, James Prescott Joule, the son of a prosperous Manchester brewer, proved one of a kind.

Around the time Joule turned adult, the British brewing industry was undergoing change with the spread of new ideas for scientific brewing alongside a more critical stance toward the old craftsmanship traditions. Disclosing a sharp eye for business, Joule's father foresaw the potential of a scientifically minded approach to the delicate process of heat conversion in the germination, fermentation, and mashing of malt. In 1834, he determined that his two sons should give up on the private tuition provided by a resident master to begin studying chemistry under the supervision of Manchester's most famous chemist, John Dalton. In the years ahead, the young James Joule had to split his time between the brewery and chemistry and mathematics lessons, becoming the assistant to his father in managing and supervising the brewery while also attempting to bring about improvements to the productive processes. Most of all, it was the practical dabbling around with mashing heats and more efficient and faster machines that attracted his attentions (Sibum 1998).

At the age of nineteen, Joule also commenced independent experiments in a laboratory installed in his father's house. Although his purposes were rather practical, motivated by the intention of improving an existing electromagnetic engine, he soon got himself involved in hard scientific measurements. Since all measurements drew upon the registration of very small changes, the observer had to be a true authority in the science of observation. Besides a "trained and practiced eye" in reading temperature scales from the most accurate instruments available through to the right timing, the observer had furthermore to control a set of disturbing laboratorial incidents such as the reaction of the thermometers to radiation from the body, variations in room temperature, or the reaction of thermometers to their own displacement (Sibum 1995). As with much of Joule's work, the thermometrical skills acquired in the brewery milieu, in conjunction with the personal interest taken in the construction of scientific instruments in addition to direct acquaintances with skilled instrument makers combined to result in the utter mastery of laboratorial conditions.

Upon assuming electrical action as the prime agent in energy conversion, the experimental physicist shifted his emphasis and became increasingly interested in the study of friction as a perhaps even more important alternative agent for the transformation of mechanical work into heat. A new apparatus was built consisting of a copper vessel with water inside and within which a paddle wheel was fitted. The paddle wheel was set in motion through machinery connected to weights and pulleys that measured the weights' distance while descending. By this means, the mechanical action exerted to rotate the wheel (distance × weight) could be estimated and corresponded to its effect in raising the water temperature. Friction, the conversion agent, turned this mechanical work into higher levels of temperature, thus enabling the assessment of the mechanical value of heat. As usual, Joule's experience bore its own stamp with the creation of a unique controlled laboratorial environment able to ensure the accurate handling of equipment and thermometers. With the apparatus working close to standard conditions, it became possible to gauge how much force (pounds per meter) was needed to raise the temperature of 1 kg of water by 1 °F. Furthermore, once this sort of "quantitative equivalences of natural powers" was ascribed in a reliable and trustworthy manner, a common trait began to surface among factors hitherto separated: the conservation of the same amount of energy in the force that moved the paddle wheel as in the water's temperature.

Irrespective of their differences, "natural powers" like electromagnetism, mechanical force, heat, or light all appeared interlinked through an endless chain of interconversions: "In the conversions nothing is ever lost. The same quantity of heat will always be converted into the same quantity of living force. We can therefore express the equivalency in definite language applicable at all times and under all circumstances" (Joule 1847a: 10). Drawing upon previous experiences, Joule remarked that "the attraction of 817 lb through the space of one foot is equivalent to and convertible into, the living force possessed by a body of the same weight of 817 lb when moving with the velocity of eight feet per second, and this living force is again convertible into the quantity of heat which can increase the temperature of one pound of water by 1 °F" (Joule 1847a: 10, b). No waste or loss actually occurred throughout these changes since the destruction of living force is only apparent. According to the British physicist, it thus proves absurd to suppose that the powers with which God has endowed matter might be destroyed: whenever some force is annihilated by percussion, friction, or by any other means, an exact equivalent of heat is released. Underlying this physical realism is the conviction that all phenomena derive from the manifestation of a single substance whose identity is constantly maintained and thereby ensuring that the total amount of energy in the universe remains constant.

Hence, the principle of the physical conversion of energy sources, which had prevailed throughout the steam engine era paradigm, with fossil fuels being converted into heat and heat into mechanical work, now appeared somehow illusory. Closer inspection, grounded in laboratorial micromeasurements, unveiled the persistence of energy behind these physical conversions. Continuity, indestructibility, and conservation henceforth became the keys for understanding life on earth. The unity of nature, described in terms of the "conservation of natural powers," resulted from the regularities observed in energy transformations.

In hindsight, one may say that Joule's concept of "conservation" stood one step ahead of Nicolas Clément-Desormes principle of "conversion," even though the ideas advanced by the young brewer physicist sounded strange to contemporary ears who at first gave a cold reception to its papers. However, this early distrust quickly dissipated and was replaced by broad scientific recognition. In reality, the principle of the conservation of energy soon became reified as the cornerstone of a new science called thermodynamics and encapsulated under the heading of the First Law of Thermodynamics. After the turn of nineteenth century, a second version of the law of conservation of energy would restate the principle of the indestructibility of energy, adding nonetheless that at each transformation, part of the energy is transformed and part is dissipated (Falk 1985). Hence, while energy is not actually destroyed, there is a loss in the final utility obtained. Subsequent to these discoveries, accounting for the conserved and for the dissipated proportion established, the theoretical grounds for assessing changes throughout input cycles. What was at stake here involved the measurement of the proportion of energy input that attained its final destination?

The First Law of Thermodynamics turned scientific interest toward the common characteristics of the "powers with which God has endowed matter." Energy henceforth appeared as the unifying field for all mutually convertible energy forms, from simple motion to electricity, passing through mechanical work, chemical action, heat, and light. Within this conceptual framework, the units of measurement could be detached from their historical utility and totaled up so as to produce a global balance of everything able to produce work. To the extent that a certain amount of heat could be expressed in terms of the work necessary to move a body, with a certain weight, under a certain velocity, for a determined distance, the heat as physical phenomenon was necessarily dematerialized and turned into "energy." This new dematerialized entity remained aloof to effective usage and instead constituting a convenient outline for singling out a whole new subdivision of society. Once the overarching concept was reckoned as a common characteristic of all mutually convertible energy forms, this spilled over into everything else: "It was in the cyclone that devastated the land, in the cooling zephyr of a summer's eve, in the awful rolling of the thunder and in the lightning's flash, as in the rustle of leaves and in the gentle cooing of the dove" (Atwater 1887b: 398). If energy was everywhere, it had furthermore to be present in men. The thermodynamics of living beings consequently became the next step in research.

2.3.2 *Man as a Living Machine*

The steam engine offered a vantage point for the study of man and physicists and engineers soon noticed that there was a lot to learn by looking at human beings from the perspective of machines. This proved an entirely new insight, which set apart things once considered the core essence, including intelligence and will, to focus exclusively on the body's automatic system. Beyond the suggestive power of the man–engine analogy, thermodynamics set out a whole new research agenda for the singular "self-contained prime mover," describing his "furnace" his "mechanism of work and energy development," his "mechanism of transmission of power peculiarly and exactly adapted to its purposes," and the controversial issue of heat transformation (Thurston 1894: 34–37). Major problems, however, stood in the way of the spread of thermodynamics. Although the analogy seemed to work well up to a certain point of general description, just as scientists reached the key causal explanations of human metabolism, the theory began to falter. By the end of the nineteenth century, it still remained impossible to agree on whether or not heat was the direct product of the oxidation of food, the result of the oxidation of worn muscle and other tissues, or the product of thermal and other interactions occurring within the body.

A second troublesome question was the absence in animal temperature systems of differences similar to those characterizing and limiting the action of thermodynamic engines. Humans and animals do not display anything like a phase of heat–steam generation followed by the cylinder's expansion and consequent cooling. According to one important prevailing stream of physics, the production of motive power only ever occurs when it proves possible to produce a difference in temperature and "destroy the equilibrium of the caloric" (heat). It thus becomes the temperature difference between two reservoirs that determines the amount of work extractable by a heat engine. Converting heat-based thermal energy into kinetic energy—i.e., doing work, in effect requires recourse to the power of steam to move the piston down through the cylinder in a working stroke, thereby expelling steam into either the atmosphere or a condenser so as to cool down the core nucleus of the engine. The discovery that heat flows from bodies of high temperature to bodies of low temperature laid the ground for the Second Law of Thermodynamics (Kerker 1960; Uffink 2007). As far as animal metabolisms were concerned, no equivalent heat contrasts or heat flows were observed throughout mechanical work production. Resorting to "reductio ad absurdum" arguments, to perform work in consonance with the thermodynamic laws of heat engines, the working parts of the human body would have to reach a temperature of around 140 °C (284 °F), therefore way above the boiling point of water, before then cooling the entire body so that a new cycle of work could begin (Thurston 1894: 47). Without this steaming of the body, the transformations of energy could not really be understood as "thermodynamic." The accumulation of unanswered questions and the rule breaking of basic scientific principles accrued the doubts as to the true nature of living machines.

Finally, there was also the problem of energy conservation. Human beings, just like animals, were seen as agents of energy conversion wedded to the daily transformation of food into mechanical work. This feature had long been understood, and Joule himself had praised the efficiency of the "animal frame" vis-à-vis the best engines of his time. However, unlike the steam engine, the relationship between the amount of food and the quantity of muscular force deployed by man was neither linear nor deterministic. Human beings seemed more like machines in which the coal supply had been disengaged from the movement of the piston. Clerks, for instance, feed themselves with large amounts of foodstuff but accomplish little physical force. From this, it would appear odd to look at muscular force as a means of conserving the energy contained in food. The logical follow-up to this apparent dissipation of energy was to pose the question of whether intellectual work could likewise be a means to conserving the chemical energy of food. Nevertheless, such endeavors ended in deadlock, for no amount of experimental evidence could confirm just how much food was needed to accomplish a given quantity of intellectual labor. Globally, the energy spent on thinking remained a riddle unsolved by nineteenth-century scholarship.

On closer examination, the human body-steam engine analogy violated the Second and the First Laws of Thermodynamics. Moreover, no less problematic was the measurement of inputs. In line with the engine analogy, food would stand for fuel, the body for a heat converter, and muscular force for power. In this context, scientific research stuck to the line of revealing the chemical basis for fuel nutrition, but some questions remained unsettled, especially the methodology for appraising the chemical energy of food: should the heat in proteins be deemed the exclusive source? Was it not preferable to measure the carbon dioxide formed during metabolism? How about estimating the energy expenditure of a man's body mass while engaged in hard physical effort? (Carpenter 2003). With several hypotheses out in the open, the scientific community became wary that a new paradigm for food classification was looming over them. Whatever the path taken, the new developments would hinge upon the measurement of interchangeable chemical components rather than flavor; upon the economy of muscular effort rather than upon subjective utility; upon technical rationality rather than upon moral and religious judgments.

Despite the great strides made by European laboratories in the field of nutritional chemical analysis, the synthesis of a new scheme to measure the energy value of food was undertaken by the American Wilbur Atwater, while practically in parallel with the work of Rubner in Germany. A chemistry professor at Tennessee and Maine University, Atwater developed his research in an experimental agricultural station that specialized in analyzing the chemical composition of food. The groundwork commenced over 1878 and 1888 underpinned the future growth of his laboratory as a solid institution endorsed with funds, scientific equipment, and overarching project goals. Aware of the German advance in the field of physiology and nutritional chemistry, Atwater took leave of absence to visit Munich whereupon he learned firsthand of the procedures in effect for studying human nutrition by means of digestibility, nitrogen balance trials, and the transformation of heat by calorimetric combustion (Carpenter 1994).

At that stage, European physiology was at a critical juncture on the path to attribute energy values to foods. After his return to the USA, Atwater followed in the footsteps of the German-Munich school and embarked on an ambitious project to standardize American foods according to the protein and energy supplied. Later, in a series of articles published in the popular review "Century Magazine," the chemist decided to extend the analysis by adding the fuel value of proteins to the fuel value of carbohydrates and fats. For each food material such as beef, canned salmon, butter, or rye flour, he summed up the heat contained in the main nutrients, by unit of weight, and consequently arriving at the potential energy available in one pound of beef, one pound of canned salmon, etc. (Fig. 2.1). Applying data from human digestion experiments, the author devised the final "coefficients

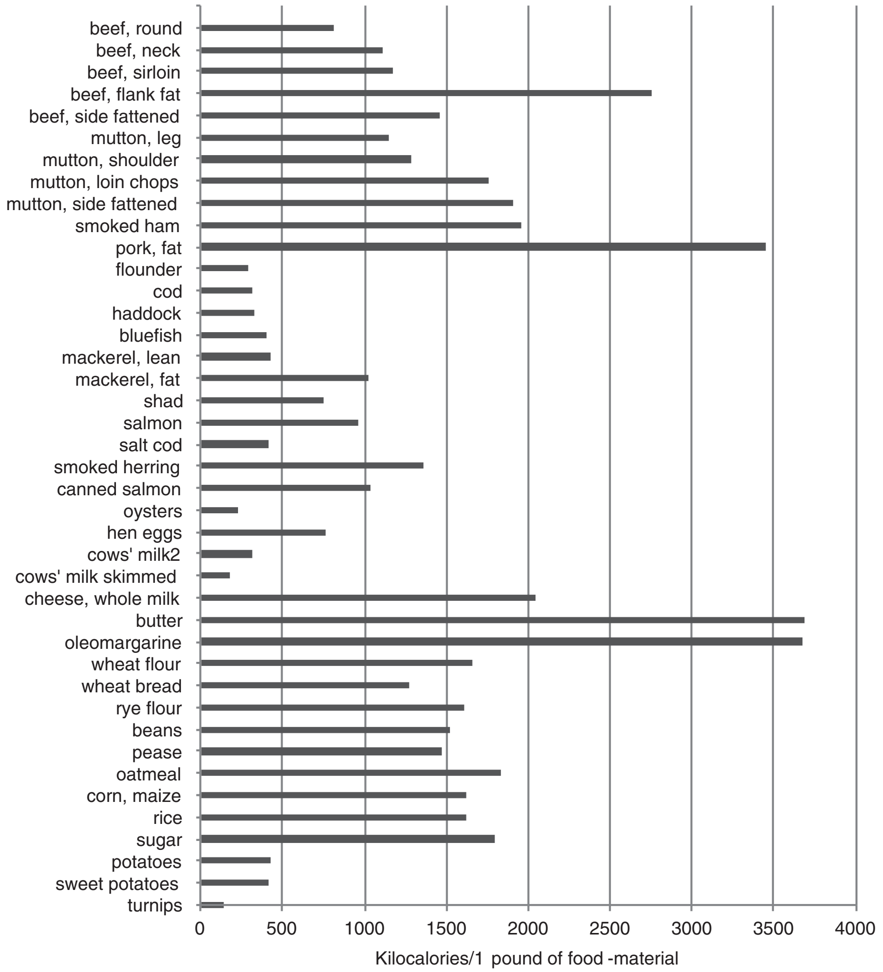


Fig. 2.1 Food energy potential. (kilo) calories in the nutrients of one pound of food material (Atwater 1887b)

of availability," defined as the intake minus fecal excretion divided by intake. By this means, he set forth the concept of "metabolizable energy," or the difference between the gross energy (as measured in the calorimeter), and the energy contained in feces and urine (as also ascertained by the calorimeter).

Lacking any better way to ascertain the energy value of proteins, carbohydrates, and fats, Atwater drew upon the European experiences and particularly laboratorial experiences with the device dubbed the calorimeter (Fig. 2.2). These consisted in measuring the heat produced during the combustion of organic substances (previously dried out) with the oxygen consumed in their burning. The chemical

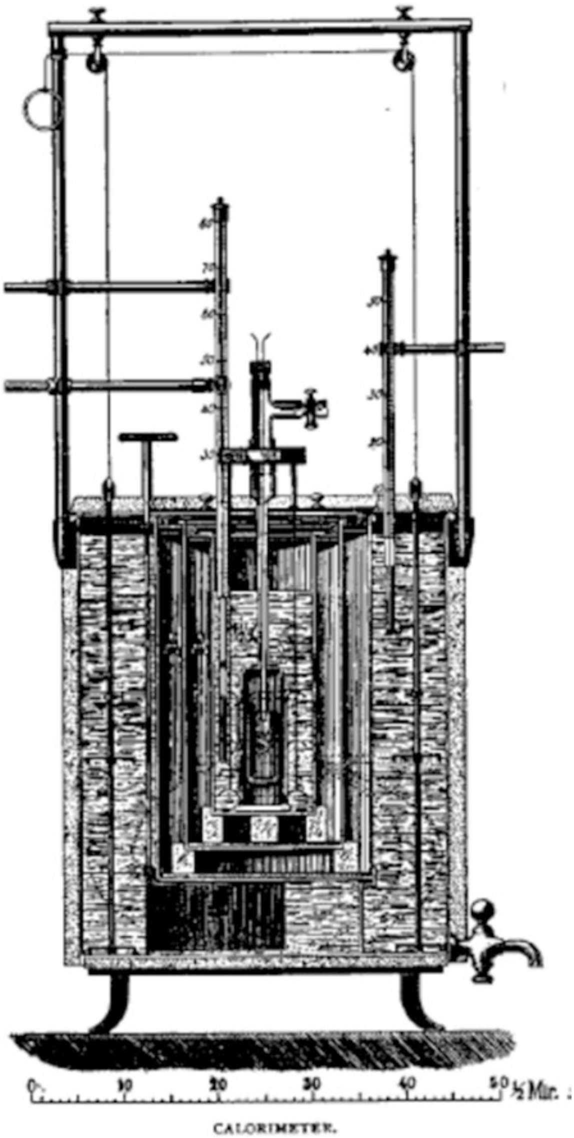


Fig. 2.2 The calorimeter (Atwater 1887b)

energy released through combustion was conveyed from an inner calorimeter container to a cover enclosed in a larger cylinder with water, so that the rise in water temperature, as shown by the attached thermometer, unveiled the amount of energy contained in the organic substance.

Like Watt's steam engine or Joule's paddle wheel device, the calorimeter was fundamentally an energy transformation machine: it converted the potential chemical energy of food into heat. Motivated by the desire to provide nutritious yet inexpensive food for people accustomed to doing physical work, Atwater's decided to present his conclusions to the American public by dubbing the energy value of foods as "calories." This decision prompted the reappearance of the old thermal unit across the Atlantic and completely refurbished with a new meaning. For a better understanding of Atwater's idiosyncrasy, it is important to note that, in the interim, the "calorie" unit had barely been applied by German and American experimental research. Moreover, Atwater defined the calorie not by the Clément-Desormes general gauge of the amount of heat needed to raise the temperature of water by 1 °C but, instead, through its equivalent unit of mechanical force as the potential energy needed to support a given amount of physical work against gravity, calculated as 1.53 foot-tons (Hargrove 2006, 2007).

The food calorie provided a fit in the range of hundreds or thousands of energy units per pound of food, thereby making it substantially easier for consumers to comprehend the nutritional scales. For someone concerned about spreading useful scientific knowledge throughout every layer of the population, the convenience and pedagogical clarity bestowed by quickly identifying foodstuffs on an ordinal scale (three and four digits for each food material) certainly constituted a strong justifying criterion. In any event, even the American physiologist with a strong profile as a publicist was far from imagining the success his tables would attain. At the time the article was published, he was not even sure that the heat produced by human metabolism could be equivalent to the heat produced by the laboratory combustion of organic materials (Atwater 1887a, b). The digestive system of individuals and the burning core of the calorimeter might not turn out so different after all. In such a case, the thermal values measured in laboratories might be susceptible to extrapolation as nutritional guidelines for human beings (broadly speaking, this hypothesis later came to be confirmed although with some adjustments) (Bijal 2009). In spite of the persistence of these and others scientific caveats the "calorie" quickly won a large audience among the American public. Hence, a food calorie actually refers to a kilocalorie, or 1,000 cal. That is, 1 food calorie equals 1 kcal or the amount of energy needed to raise 1 kg of water.

2.3.3 Overturning the Energy Value of Food

As a unit related to the fuel values of foods, the "Calorie" was promoted and disseminated by the Farmers' Bulletin issued by the United States Department of Agriculture, whereupon it found its way into handbooks for clinical and medical practitioners, advertisements, popular receipts, details on cereal boxes, emblazoned across foodstuff packages, newspaper articles, and books dealing with weight reduction (Hargrove 2007). In one fell swoop, food values rippled from the anonymity of academic footnotes to the spotlight of society. And the more the calorie spanned across popular culture, the more it became a reality in itself and a reality detached from energy measurements, almost as if an appropriate attribute of the foodstuff.

The enthusiasm generated by Atwater's tables stemmed from the degree to which a standardized, uniform, and quantitative perspective toward nutritional elements met a preexisting necessity. To the extent that the essentials for fueling humans were properly ascertained, the study of man could borrow at will from the study of thermodynamics and mechanical engineering. Still more importantly, by putting energy values on food, scientists fostered major changes in social relations: on the one hand, they allowed public powers to step into domains formerly reserved to personal choices, local culture, and morals; on the other hand, they unleashed an array of decentralized assessments about food—the personal diets and dietary plans that themselves went onto become an industry.

Scientific food measurements came to the fore at a time when the very notion of individual freedom was being undermined by principles of scientific-based government and technocratic rationalization. Both in the USA and in Europe, engineers and scientists envisioned techniques for the management of populations based upon statistical evidence and the extrapolation of general principles and social laws. Chemists and physiologists specializing in human nutrition began to characterize food as "fuel for the human machine" and nutrition science as a system of physiological "book keeping" susceptible to increasing the cost-effectiveness of human labor. Practical governmental usage of energy–food tables advanced swiftly in the USA, buoyed by ongoing trends to disseminate mutually dependent modes of management and knowledge (Taylor's efforts to establish a methodical science of workplace time and motion), but also by progressives advocating the redefinition of the state's obligations to incorporate the satisfaction of the nation's dietary needs (Desrosières 1998; Cullather 2007. "Progressivism" is further developed in Sect. 6.2). Caught by the call for active reforms, nutrition was brought into the fold of government and equated as standardized food management. Initially, federal interference was limited to an assessment of social and industrial relations by launching surveys, scientific standards of living for the working class and indexes of food consumption. After World War I, there came a bolder turn with recourse to the (kilo) calorie by military planners as a normative gauge to marshal resources and provide relief to stricken areas of Europe. Finally, the inter-war period saw the further imbuement of scientific food management by economic policy with its further spread across Europe and into their imperial dominions.

Within the sphere of population management, the calorie enabled a more rational equilibrium between "fuel" and physical effort, with a global dietary standard set in 1935 by the Health Organization of the League of Nations with the minimum amount of 2,500 calories set per day for a laboring adult (Grigg 1981; Cullather 2007). However, in the private realm, the unit took on a different hue, becoming instead a practical guide to cutting down on uncontrolled fueling, that is, to weight reduction.

While initial scientific concerns tended to underscore the working class food situation, excess weight proved the all-time popular theme, particularly among the middle classes. Due to the fact that Atwater's measurements tapped the heat produced during the combustion of organic substances, his ranks of energy values displayed a bias toward food materials with large proportions of fat nutrients. In terms of the calorimeter, fats constituted the greatest energetic resources anyone could consume. As the Farmers' Bulletin of 1894 explained: "The fats have weight for weight, about two and one-fourth times the potential energy of either the protein or the carbohydrates. Water has no potential energy. Hence the food materials which have the fattest and the least water have the highest fuel value. Butter and fat pork consist almost exclusively of fat. They lead the other food materials in fuel value. Lard, suet, and olive oil have even less water, and hence exceed the butter in this respect. Oleomargarine has about the same composition, fuel value and food value, as butter" (Atwater 1894).

From a thermodynamic standpoint, fat pork assured the best output for a given unit of input. In contrast, fruits, leafy vegetables, and fish registered low-level nutritional values and could scarcely be classified as foods because little muscular work would result from consuming these staples (which explains their absence from the food materials listed in Fig. 2.1). Ranking foods by their fuel value alone for this reason implies a partial viewpoint in which human qualities were extolled through the thermodynamic analogy with heat engines. In Atwater's own words, "the potential energy represents simply the fuel value of the food and hence is only an incomplete measure of its nutritional value" (Atwater 1887b: 400). However, in spite of these warnings, the word "calorie" became entrenched in the very methodology that prompted its revival, looming pragmatically as the single definitive measure of food value up until the "vitamin revolution" of the 1920s (which rehabilitated the in the meanwhile devalued fruits and vegetables). Within this framework, counting "calories," on a daily basis, could either signify an appropriate dosage of fuel for working people or a dosage of "fats" to blue collar employees. Since the ascending caloric scale of food materials was also a scale of fatness, reversing the whole scheme enabled the transformation of the energy values ranking into a blueprint for weight reduction. Through the effects of social usage and social appropriation, nutritional tables became far more useful in their inverted form and serving as practical diet plans. The "calorie" consequently thus became associated with weight reduction rather than with the amount of heat generated. In the view of the middle classes, energy intake ought later to be "burned off" by the mechanical work exercised in gymnasiums.

2.4 Energy Aggregation

The discovery that different qualities of energy could be aggregated under the same hood and summed up was one of the most remarkable achievements of nineteenth-century science. Two critical steps proved essential to accomplishing this aggregation.

In the first place, new measurement units had to be invented to grasp the diverse transformations of kinetic, thermal, and chemical energy. The mechanical work of pistons was sold to end users under the caption of horsepower; the potential of primary chemical energy contained in coal was assessed through the kilocalorie just as the secondary heat generated by steam; the primary fuel value of food along with the secondary metabolized value of its intake were, in turn, estimated according to the heat unit of the food calorie. Briefly, measurement moved from final to secondary and primary energy, encompassing the different qualities of energy. Science aimed at discovering the fundamental physics to the processes behind the functioning of machines and, afterward, behind human metabolism.

In the second place, transformation was explained through the Law of Energy Conservation. The outcome of mechanical work undertaken by a steam engine was nothing other than one way of conserving the energy contained in heat. Its evidence could easily be noticed whenever there was a thermodynamic exchange and part of the "living force" contained in mechanical power (or heat, light, or electromagnetism) was transformed into another energy form. Without this endless chain of conversions, neither nature nor human life would hold up. Since the different physical processes associated with the transformation of energy were perceived as mottled versions of the same thing, in a realistic interpretation of physics, there was a sound basis for aggregating and summing up the different qualities of energy.

As it turned out, aggregation resulted from improved specifications in measurement combined with the dematerialization of that under measurement. While the useful energy delivered by an animal, a windmill or a worker captured most attention in the preindustrial era, the fossil fuel economy of the nineteenth century shifted the focus toward the cycle of energy transformations based on the yardstick of chemical energy. Only following the order of changes and only measuring how much was conserved at each stage could aggregation produce meaningful results. By this means, nineteenth-century discoveries laid the ground for input-cycle assessment: whether of primary, secondary, or final energy. What was missing in this measurement scheme, largely drawing upon the First Law of the Conservation of Energy, was recognition of the differences in the qualities of energy, more specifically that the ability to do useful work varied according to the power sources applied.

At the same time, energy accounting overlooked almost everything that fell short of the measuring rod of money. In official estimates, primary energy registered only commercial carriers, and particularly coal and oil, brushing aside all the useful work of thousands of windmills, watermills, men, and animals. What makes this neglect particularly biased is the fact that in around 1880–1890, humanity reached the peak in the level of muscular force ever deployed by manpower, largely owing to the unprecedented extension of working times (Schor 1991; Costa 2000), the number of working horses, employed both on farms and in cities, also reached its all-time peak in mankind's history (Thompson 1983; McShane and Tarr 2007; Mom 2009) and a large number of US plants persisted in their use of waterwheels and water turbines despite the growing relative cost disadvantage

(Hollerith 1883; Atack 1979). Swift technological development, however, turned everything that was not modern and tradable at a long distance, irrelevant.

Even though the theme of electrical unit systems is beyond the scope of this chapter, the main lesson imparted by the measurement of resistance, current, and electromotive force illustrates the role of standards set by the ingenuity of individuals and spread spontaneously throughout market interactions or personal scientific networks. A completely different arrangement surfaced at the close of the nineteenth century with the installation of state sponsored national laboratories encharged with the production of standards through systematic scientific research (the German Imperial Institute was the first in 1887), of wide reaching international meetings (the 1893 Chicago Congress was the first to set out a coherent system for electromagnetical units defined by material standards) and, finally, by supranational entities (the International Electrical Commission, following the St. Louis Congress, 1904). It was also according to the framework established by the electrical unit institutions that the measurement units for thermal, kinetic, chemical, and radiant energy came to be reviewed, reformulated, and further standardized.

References

Atack J (1979) Fact in fiction? The relative costs of steam and water power: a simulation approach. Explor Econ Hist 16(10):409–437

Atwater OW (1887a) The chemistry of food and nutrition: the composition of our bodies and our food. Century Illustrated Mag 34:59–74

Atwater OW (1887b) The potential energy of food iii: the chemistry and economy of food. Century Mag 34:397–405

Atwater OW (1894) Foods: nutritive value and cost. Farmers' Bull 23:1–30

Ayres RU (1998) Eco-thermodynamics: economics and the second law. Ecol Econ 26(2):189–209

Ayres RU, Ayres LW, Warr B (2003) Energy, power and work in the US economy 1990–1998. Energy 28(3):219–273

Ayres RU, Warr B (2005) Accounting for growth: the role of physical work. Struct Change Econ Dyn 16(2):181–209

Bijal T (2009) The calorie delusion. New Sci 203(2717):30–33

Carpenter KJ (1994) The life and times of W. O. Atwater (1844–1907). J Nutr 124:1707–1714

Carpenter KJ (2003) A short history of nutritional science: part 2 (1885–1912). J Nutr 133:975–984

Cullather N (2007) The foreign policy of the calorie. Am Hist Rev 122(2):337–364

Costa DL (2000) The wage and the length of the work day: from the 1890s to 1991. J Labor Econ 18(1):156–181

Desrosières A (1998) The politics of large numbers. Harvard University Press, Cambridge

Falk G (1985) Entropy a resurrection of the caloric—a look at the history of thermodynamics. Eur J Phys 6(2):108–115

Fox R (1970) Watt's expansive principle in the work of Sadi Carnot and Nicolas Clement. Notes Rec R Soc London 24(2):233–253

Giampietro M, Sorman AH (2012) Are energy statistics useful for making energy scenarios? Energy 37(1):5–17

Grigg D (1981) The historiography of hunger: changing views on the world food problem 1945–1980. Trans Inst Br Geogr 6(3):279–292

Hargrove JL (2006) History of the calorie in nutrition. Nutr J 136:2957–2961
Hargrove JL (2007) Does the history of food energy units suggest a solution to "calorie confusion"? Nutr J 6(44):1–11
Hartman PM (1982) Energy, force and matter. The conceptual development of nineteenth-century physics. Cambridge University Press, Cambridge
Hills RL, Pacey AJ (1972) The measurement of power in early steam-driven textile mills. Technol Cult 13(1):25–43
Hills RL (1993) Power from steam. A history of the stationary steam engine, 2nd edn. Cambridge University Press, Cambridge
Henry R (1902) History of the growth of the steam engine. D. Appleton and Company, New York
Hollerith H (1883) Statistics of power used in manufactures. U.S. Census—tenth census of the United States—Government printing office, vol 2. Washington
Joule JP (1847a) (1892) On matter, living force and heat. In: Reynolds O (ed) Memoir of James Prescott Joule. Manchester Literary and Philosophical Society, Manchester, pp 1–32
Joule JP (1847b) (1884) On the mechanical equivalent of heat, as determined by the heat evolved by the friction of fluids. In: Scoresby W, Playfair L, Kelvin WT (eds) The scientific papers of James Prescott Joule (1884). Physical Society of London, London, p 277–281
Kanefsky J, Robey J (1980) Steam-engines in 18th century Britain: a quantitative assessment. Technol Cult 21(2):161–186
Kerker M (1960) Sadi Carnot and the steam engine engineers. Isis 51(3):257–270
Kula W (1986) Measures and men. Princeton University Press, Princeton
Lightfoot HD (2007) Understand the three different scales for measuring primary energy and avoid errors. Energy 32(8):1478–1483
McShane C, Tarr J (2003) The decline of the urban horse in American cities. J Transp Hist 24(2):177–198
McShane C, Tarr J (2007) The horse in the city: living machines in the nineteenth century. Johns Hopkins University Press, Baltimore
Mom G (2009) Compétition et coexistence: la motorisation des transports terrestres et le lent processus de substitution de la traction équine. Le mouvement Soc 229(1):13–39
Musson AE, Robinson E (1959) The early growth of steam power. Econ Hist Rev 11(3):418–439
Pollard S (1983) Capitalism and rationality: a study of measurements in British coal industry 1750–1830. Explor Econ Hist 20(1):110–129
Schor JB (1991) The overworked American: the unexpected decline of leisure. Basic Books, New York
Sibum HO (1995) Reworking the mechanical value of heat: instruments of precision and gestures of accuracy in early Victorian England. Stud Hist Philos Sci 26(1):73–106
Sibum HO (1998) Les gestes de la mesure. Joule, les pratiques de la brasserie et la science. Histoire Sci Sociales 53(4):745–774
Smil V (1994) Energy in World history (essays in World history). Westview Press, Boulder-Colorado, Boulder
Tann J, Breckin MJ (1978) The international diffusion of the Watt engine 1775–1825. Econ Hist Rev 31(4):541–564
Tann J (1998) Fixed capital formation in steam-power: 1775–1825. In: Feinstein C, Pollard S (eds) Studies in capital formation in the United Kingdom 1750–1920. Oxford University Press, Oxford, pp 164–181
Thompson FML (ed) (1983) Horses in European economic history: a preliminary canter. British Agricultural History Society, Reading
Thurston H (1894) The animal as a machine and a prime motor. Wiley, New York
Uffink J (2007) Bluff your way in the second law of thermodynamics. Stud Hist Philos Mod Phys 32(3):305–394
Warr B et al (2010) Energy use and economic development: a comparative analysis of useful work supply in Austria, Japan, the United Kingdom and the US during 100 years of economic growth. Ecol Econ 69(10):1904–1917

Chapter 3
The Rebound Effect

Abstract The view that energy-efficiency improvements will actually serve to increase rather than reduce energy consumption was first proposed by the British economist, William Stanley Jevons in 1865. The effect he singled out was dubbed the "rebound effect" and its foundations became known as the "Jevons paradox." This chapter places the debate back in its historical context by drawing attention to the issue that triggered major theoretical developments: whether or not Victorian Britain faced a threat of coal depletion alongside all the consequences such as the ensuing slowdown in economic activity and decline in imperial might. We detail how many of the forecasts associated with the impending coal scarcity thesis were not borne out by reality. Furthermore, we also explain what went wrong with Jevons's view before the final pages discuss the role of energy efficiency in the contemporary world.

3.1 The Riddle of Saving and Spending

Put forward for the first time in 1865 by the British economist William Stanley Jevons, the rebound effect asserts that saving energy actually increases energy consumption. The riddle-like formulation of this concept has puzzled many minds, even while Jevons' argument has itself remained undisputed. The crucial point arises out of the impact of innovation on prices: the savings leveraged by improvements in technology tend to diminish the amount of fossil fuel necessary for a given output, thereby lowering the energy costs. By this means, what is saved through the adoption of more efficient devices subsequently becomes offset by the stimulus to enhanced diffusion and consumption. The wheel thereupon turns a full circle: energy savings increase consumption.

This basic outline is faithful to the original reasoning. A more contemporary and comprehensive rendering is found in Saunders's statement that rebound effects takes place in an economy with fixed real energy prices whenever energy-efficiency gains increase energy consumption above that which would be its level

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_3,

without such gains. Another way of expressing the same concept is to recognize that if the achievements in energy productivity are not matched by gains in the productivity of other inputs (lowering the energy cost relative to other inputs), there will be a rebound effect on energy consumption, offsetting the initial reduction in energy usage (Saunders 1992; Jaccard and Bataille 2000).

Consequently, the effects of engineering efficiency cannot be directly transposed onto the macroeconomy. According to the "rebound" view, the aggregate effect of a new technology is different to the sum of its engineering efficiencies and while textbook assumptions might be technically correct, they still prove economically wrong. The introduction of the price mechanism was precisely the fact that overturned established beliefs. Referring only the most commonly cited examples in the literature, a reduction in the level of coal consumption from 4 to 2.25 pounds per horsepower/h might result in more steam engines entering into operation in just the same way as an improvement in the on-road fuel economy of vehicles from 14 to 12 miles per gallon, for example, might actually result in more driving. In both cases, energy-efficiency savings are first canceled out and then overwhelmed by enhanced consumption of the same good or service. When this happens, the causal mechanism at stake is termed a direct rebound effect. However, even when consumers choose not to drive any further, or industrialists choose not to buy any more fuel-efficient steam engines, they might still wish to spend the money saved on other types of goods and services that also require energy to provide. Whenever energy-efficiency gains lead to increased consumption of goods or services other than those generating the actual savings, this causal mechanism is called an indirect rebound effect. One way or another, what gets saved through improved efficiency is taken back by the escalation in production.

Significantly, the theoretical arguments posed by Jevons still nourish contemporary research a 150 years on from his original essay. Discovering the extent of the rebound effect and how it varies with types of energy usage and with the different stages of development of particular technologies still remain crucial questions with overwhelming implications for policies aimed at reducing carbon dioxide (CO_2) emissions and counteracting global warming. It is indeed worth noting that should Jevons prove right then encouraging energy efficiency as a means of reducing carbon emissions will turn out to be a counter-productive policy (Sorrell 2009).

The following pages place the debate back in its historical context by drawing attention to the issue that triggered major theoretical developments: whether or not Victorian Britain faced a threat of coal depletion alongside all the consequences such as the ensuing slowdown in economic activity and decline in imperial might. Puzzling as it may seem, the fact remains that the adoption of mechanical machinery, railways, steamships, and long-distance communications was accompanied by growing concerns about the eventuality of running out of coal. Though this merely represented one possible outcome, it nonetheless shook the foundations of society. The scarcity of the energy source upon which most modern technologies depended foreshadowed a stormy future and an uncertain present. History granted canonical status to this debate and dubbed it "the coal question."

3.2 Antecedents of the Coal Question

3.2.1 *Tapping the Unobservable*

Up to the nineteenth century, detailed knowledge about coal reserves was mainly in the hands of mine owners, merchants, and colliery overmen sometimes augmented by information gleaned from some antiquarian writer. These were the right people to ask when trying to assess how much coal could still be recovered across a whole region. Their knowledge was based partly on direct or indirect experience from excavation in its most basic form, from outcrops, or more generally, from information derived from the removal of the overlying rock and soil (overburden) and the rate at which coal seams were thus exposed. One way or the other, they undoubtedly knew which were the shallowest and easiest seams to explore. Likewise, they could foresee how much of the mineral could be removed and how long it would take to exhaust the pits. The calculation itself was conveniently made by rule of thumb based upon the prevailing British assumption that a cubic yard of underground coal yielded a ton in weight. To reach a final figure, particular consideration was given to the thickness of the seams and the surface area beyond which the mining of coal was no longer judged profitable. Ultimately, sinking deeper shafts, with greater costs of excavation, haulage, and drainage set a ceiling for the recovery of underground stocks. Deeper mining was undertaken only when the thickness of lower seams or higher-quality coal justified the extra costs.

Hence, experience of mining in progress painted a good picture of the local potential. However, what about undiscovered coalfields? And the potential of more distant regions? In this respect, common knowledge up to the nineteenth century seems to have been based on only exposed areas of coalfields without considering the possibility that they might continue beneath the cover of later deposits. The disregard for new fields where "hidden coal" might be available was particularly significant in continental Europe and in Russia, reflecting a market environment in which the industry spread step by step through to the 1860s. Regions like the Upper Silesian-Moravian Basin (a zone held by Germany) were consequently locked into a geographical pattern that was established at the dawn of the nineteenth century with all shafts lying at a distance of no more than about a mile beyond the margin of the exposed coalfield. In spite of the mounting demand for coal for iron smelting, the pattern in the Silesian-Moravian Basin remained unchanged for at least another 60 years (Pounds 1958).

Indeed, by the early nineteenth century, coal mining was nothing but a minor activity outside Britain and Belgium, hindered by enclave economies, small-scale operations, and companies with demand restricted to the immediate vicinities. In these circumstances, an estimation of the stock of fossil fuels was built into the very process of getting the coalfields up and running. Reserves certainly mattered, and they mattered above all to mining people.

A new approach toward the potential of undiscovered mineral basins based on the commercial geological prospecting of coal resources began to surface in the

1820s. This move into the unknown was both offensive and defensive. In the USA, it was on the offensive in seeking out new and flourishing business opportunities. On the private side, the discovery of coal near Pittsburgh aroused interest among economic boosters and land speculators from the East who hired teams of geologists and civil engineers in the 1820s to compile scientific reports on George's Creek Valley (Buckley 1998). Geological reports were important in attracting the attention of interested investors, as well as providing information blueprints for decisions. In the public domain, state legislatures began taking the same steps a few years later, hiring geologists to identify rich deposits of mineral resources. In cases where it was thought large coal basins might be unveiled, for example in Pennsylvania and Illinois, the authorities went so far as to instruct geologists to map the coalfields. However, in spite of the state's engagement in sponsoring the costs of obtaining information, the results from this geological prospecting and investment front were short (Adams 2006).

While Americans attempted to push their control over the territory further and to seize gilded opportunities (gold was also an object of research), the British displayed a much more defensive stance since their primary goal was to find untapped natural resources able to replace those already exhausted. Although the recruitment of geologists to carry out regional surveys was a major innovation in the 1830s—an innovation shortly to prove inherent to the concept of "reserves"—the driving force behind British endeavors was geological pessimism. Surveys were therefore carried out in regions where the coalfields had long passed their maturity and were on the verge of exhaustion by the early nineteenth century (Church 1986: 8–9). These were worrying cases in which the inescapable decline in production was already threatening local prosperity and social equilibrium. Albeit exceptional and largely outweighed by the growing production of fresh, expanding regions, such as South Wales, Yorkshire, East Midlands, and Scotland, the coalfields on the verge of exhaustion became the focus for public policy. Sooner or later, the story would certainly be repeated in other regions. Hence, the disturbing question was: just how soon? Given the lack of certainty regarding the stock of British solid fossil fuels, isolated cases of depletion which statistically held little current relevance were perceived as on the verge of snowballing into a frightening future.

Geological knowledge broadened the prospects for coal availability from that existing to future supplies. Regional surveys based on the mapping of geological structures from coal basins were later undertaken in continental Europe: the geologists Jakob and von Dechen assessed the evenness of the Ruhr basin and estimated its proven reserves while Jules Gosselet studied the coal basin of Belgium and established the likelihood of finding coal in Northern France, along the Pas-de-Calais southern border (1863–1873) (Gillet 1973: 64–69; Baumont 1928: 38–39).

Overall, these regional surveys proved highly instrumental in efforts to attract investment and chart business opportunities. Naturally, the transposition of these experiences into a national plan entailed a much larger-scale investigation with the involvement of and financing by central states, the recruitment of teams of geologists, and the redefinition of survey methodologies and logistics. By the 1860s and 1870s, governments decided to advance, thus opening the way for the first national geological surveys.

3.2.2 The Geological Evidence

With the institutionalization of data searches over entire territories, the remnants of the amateur tradition began to fadeaway and with them the liberal practice of geology. Even in Britain, where the gentlemanly way of researching "earth sciences" had strong social roots, the combined effects of both professionalization, through the survey, and greater university access changed the profile of this scientific field (O'Connor and Meadows 1976; Porter 1978). No less important was the fact that the move toward the national survey occurred in a context where the local coal economy had already been replaced by regional market integration, rising demand and external trade, the interdependence of coal mining and railway companies, and concentration through merging and incorporation (Schurr and Netschert 1975; Clark 1990: 16–28). Industrial progress had enhanced the value of this asset, raising its importance when prioritizing the natural resources to be mapped.

In the United States Congress, supporters of the decision to undertake the US Geological Survey (1879) underscored the strategic role of coordinating the geological mapping of mineral resources in the vast trans-Mississippi region and, particularly, in the largely unknown region west of the Rocky Mountains. Moreover, specific instructions were issued regarding coal mines. The final results published in the 1880s presented a threefold view of the same area: one map entitled "Areal Geology," outlined the surface distribution of the various rock masses; another unveiled the "under geology" called "Structure Sections"; and a third disclosed the rock masses known to be of economic importance due to their yield of iron, coal, gold, or other metals. From this contextual perspective, "economic geology" could be associated to the composition, structure, properties, and history of the planet's physical material, thereby becoming a matter for professional expertise (Baker 1895).

Compared with other inventories of US natural resources, such as iron ore and gold, the coal survey was often admired as the most accurate assessment ever made. In 1909, it was said "that future discoveries of coal outside of the limits indicated for existing fields are not likely to exceed 1 % of the total known supply; and future investigations are likely to diminish rather than increase the estimates of the quantity available within these limits" (Gannett 1909: 96). This fact, associated with the large reserves found, drove nineteenth-century conservationist thinking, later entangled with political progressivism, to emphasize timber, forest products, and oil as the key resources to be preserved and used and, furthermore, to be fairly distributed preventing "monopolization in the hands of a few favored interests" (Garfield 1909: 183). To counteract the voracious depletion of natural resources and particularly the lack of forest protection, some like-minded men in

Chicago set up the first contemporary environmental organization, the American Forestry Association. Helped by hunting and fishing clubs, the Association gave voice to the idea that the common good of renewable natural resources was greater than their private commercial benefits. Besides taking direct action, such as gathering information and raising public awareness, the movement called for a more active federal position to encourage tree planting and curtail claims from the lumber industry. Furthermore, they also worked on plans to establish forestry reserves and parks. In policy terms, however, coal was off the hook (Hays 1959; Clark 1987).

Unlike the business-oriented and later citizen-oriented perspective held by the US federal government as well as the fledgling environmental movement, the British elites saw the survey as a central state concern or, to be more precise, as a key issue for the Empire's political economy. By the 1860s, apprehensions about depletion had already made themselves felt in the House of Commons and prompting a blistering debate. The debate had been triggered by the signature of the Cobden-Chevalier Treaty, which forbade both the prohibition of and duties on coal trading between France and Britain. The arguments as to whether favoring French coal imports could hit British interests and thereby contribute toward the exhaustion of its mines divided the assembly. Fundamentally, no one could be truly sure that the liberalization of trade would not jeopardize a commodity which had become the mainstay of the Empire, of industrial supremacy and economic competitiveness (The Treaty and the Coal Question 1860).

There were clear priorities for data collection, and this explains why the British Geological Survey of 1861 focused exclusively on coal formations. Run by the geologist Edward Hull, the account of natural resources not only ascertained the recoverable coal available through direct surveys in each mining region, but also advanced a forecast of the likelihood of finding undiscovered "hidden coal." Thus, systematic assessment added the concept of unknown resources to the estimation of those proven. Edward Hull resorted to mapping the geological data on each rock system across the country to assess the likelihood of finding valuable untapped seams. After considering other factors, the geologist came up with the figure of 932 square miles in England and Wales worth exploring for more coal. Though potential extraction was not fully quantified, this approach ensured an extra buffer against the menace of sudden and swift depletion (Hull 1861: 138).

In the end, the recoverable resources as assessed by the Geological Survey for England, Wales, and Scotland amounted to about 80,000 million tons of coal which, at the rate of production in the late 1850s, would last for 1,100 years. Moreover, the potential for further discoveries in untapped seams further reinforced the conclusion that "for many generations to come the mineral resources of England [were] capable of bearing any drain to which they [could] possibly be subjected either for home or for foreign consumption" (Hull 1861: 139).

A clear answer had been given. Coal was sufficiently abundant to remove any shadows which might loom over the forthcoming decades: the British could expect to hold onto its lead in manufacturing, in trade and in sea power. During this process, the urgency to restore confidence made geologists deploy robust methods and forceful means (with 57 geologists employed in the 1860s) to estimate natural resources; methods soon to be emulated in other coal producing nations (O'Connor and Meadows 1976: 80). Besides the pioneering nature of the 1861 Geological Survey, two points stand out as scientific benchmarks: first, the endeavor to gauge resources beyond what had previously been acknowledged through commercial mining operations. The survey was not just a sum of the resources existing in coal mines; it also sought to draw the map of probable geological occurrences. Secondly, due to Britain's head start, the accumulated technological and scientific expertise allowed geologists to set the threshold of economically recoverable reserves at a depth of 4,000 feet (1,220 m) provided the seams were two feet thick, which meant that despite the extra costs incurred by extracting coal at 4,000 feet, the coal would still likely sell at a competitive market price.

Even though this threshold was publicly criticized as conservative, mostly by opponents of the scarcity thesis, people in the trade knew very well that mining at 4,000 feet deep, at a profit, was an achievable assumption even if in the near future. The deepest mine in Britain at that time had reached 2,100 feet and reached down to 2,370 feet a decade later. In continental Europe, working coal had been carried out at a depth of 2,760 (in the Austro-Hungarian Empire) and 3,000 feet (in Belgium) (Hodsworth 1866; Report of the Commissioners 1871:81–82). Reflecting on the subject, a British business man remarked that "with our present experience and at anything like our present cost," "we are only justified in expecting to penetrate to a depth of some 2,700 feet" (The Coal Supplies of England 1872). All in all, it took about 40 years for these expectations to materialize. Getting nearer the threshold of 4,000 feet (the limit of 3,500 feet deep was breached in 1902 at a single colliery at Rams Mine, Pendleton) was only possible through a protracted process involving a full range of technologies. To obviate the difficulties created by increasing pressures and increasing temperatures, miners had to install new pipe systems for watering and winding and fan engines for ventilation, adopt the longwall method for extracting coal, introduce compressed air machines for hauling and coal-cutting, and enhance the maintenance of roadways and airways (Royal Commission on Coal Supplies 1903).

In light of this evidence, the minimum that can be said about the 1861 Geological Survey is that it acted as a harbinger of various technological advances so as to account for "reasonably assured reserves," rather than for the straightforward criteria of "proved reserves in place." Edward Hull's figures were, from this point of view, clearly optimistic as they tend to reflect a forward-looking perspective along with great confidence in British technological accomplishments. As of 1972, the 4,000-feet threshold and the thinness of one or two feet were still recognized as a valid yardstick for reserve estimation (NA. COAL-96 1972).

In any case, the adoption of systematic geological criteria made the concept of exploitable resources dependent on technical–economic costs, rather than on the direct physical measurement of coal seams. And the dynamic upshot of this development was that the amount of reasonably proven coal reserves was expected to increase in tandem with changes in best-practice economic organizational technology. Its stocks therefore came to be viewed as a drifting edge, turning the basic uncertainty surrounding depletion into a probabilistic classification of natural resources.

3.3 The Economic Theory of Resource Depletion

3.3.1 Startling Economic Factors

The Geological Survey brought the parliamentary season of open apprehension to a sudden halt. The awesome forecast of a supply guaranteed for 1,100 years had relegated the problem of depletion to a few pits marginally located in pioneering coalfields. However, just as everything seemed settled, anxiety was actually raised another notch with renewed arguments highlighting the fragility of Britain's position.

No one was questioning Edward Hull's work or the data assembled by the Geological Survey as this problem went beyond geological estimations. Just 2 years after the figure of 80,000 million tons of reserves had been released, a prominent industrialist, inventor and patron of science, William George Armstrong, again called for changes in public policy in a speech made on becoming President of the British Association for the Advancement of Science. He argued that depletion would not be the result of the scarcity or abundance of underground seams, but of a dramatic increase in consumption. This turnaround in the understanding of the "coal question" marked a shift from geologic to economic factors. Like his scientific peers, Armstrong accepted the account given by the Geological Survey but complemented it with new forecasts based on the premise made from extrapolating the average growth recorded over the previous period (1853–1861) to find that coal consumption would increase by 2.75 million tons per year. If this were true, coal would last only for 212 years and the British government ought to sound the alarm yet again. Moreover, the best coal for industrial applications, and the least expensive to extract, would have gone long before the exhaustion deadline loomed, something for which the phrase "commercial exhaustion" was later coined (Armstrong 1863).

The security interval of coal supply was consequently pulled forward and 212 years did not exactly mean 212 years of prosperity. Before depletion, scarcity would work its evils though the mechanism of rising costs and price rise increases. This, in turn, would cause Britain's position in world trade to be taken by the emerging power of the USA, which was endowed with vast untapped reserves of good bituminous coal that would be much more competitive than British coal. Perhaps the 212 geological years of reserves would ensure <100 years of Britain's economic supremacy. It is worth noting here that Armstrong, like other British writers, equated industrial ascendancy with coal (Lozé 1900: 851–856). Within this line of reasoning, the moment of British decline was moved from the distant future of physical geologic reserves to the more immediate midterm of technical–economic profitability. Though the end was not in sight, the beginning of the end was close—at least, close enough to revive the debate.

After Armstrong had lent his weight as a successful entrepreneur, leading scientist and mine owner to the cause of protecting natural resources, the patriotic call attracted growing support among politicians, geologists, and economically minded politicians. In the ensuing years, several spokesmen appealed for a more active government position toward spoilage and waste in manufacturing and households, while others suggested reversing free-trade policies and place constraints on coal exports (Lozé 1900: 853–856). Either way, the general feeling was that an age of scarcity could not be avoided.

3.3.2 Coal and the Rebound Effect

It was in 1865 that an almost unknown author entered the debate "uninvited" and with no previous experience in related matters. William Stanley Jevons was in effect far from the main arenas of political and academic dispute when he decided to enter the debate on the coal question. For much of his life, he had thrived on adversity, and it was the stubborn ambition of pursuing an academic career in England that led him to abandon a lucrative position as assistant assayer in Sydney, Australia, and come to London to complete undergraduate and Master's degrees. During his studies, Jevons set himself the goal of creating a new synthesis of philosophy, logic, political economy, and mathematics. However, his hopes of a top academic calling were soon dashed: his first essays got a cold reception and his appointment as tutor at Owens College, Manchester, fell far below his expectations. Consequently, it seems that Jevons turned to the popular subject of coal because his theoretical contributions to political economy had received little attention. Indeed, he openly declared his intentions in a personal letter: "I am convinced that it is necessary for the present at any rate to write on popular subjects" (Jevons quoted in Missemer 2012; White 1991).

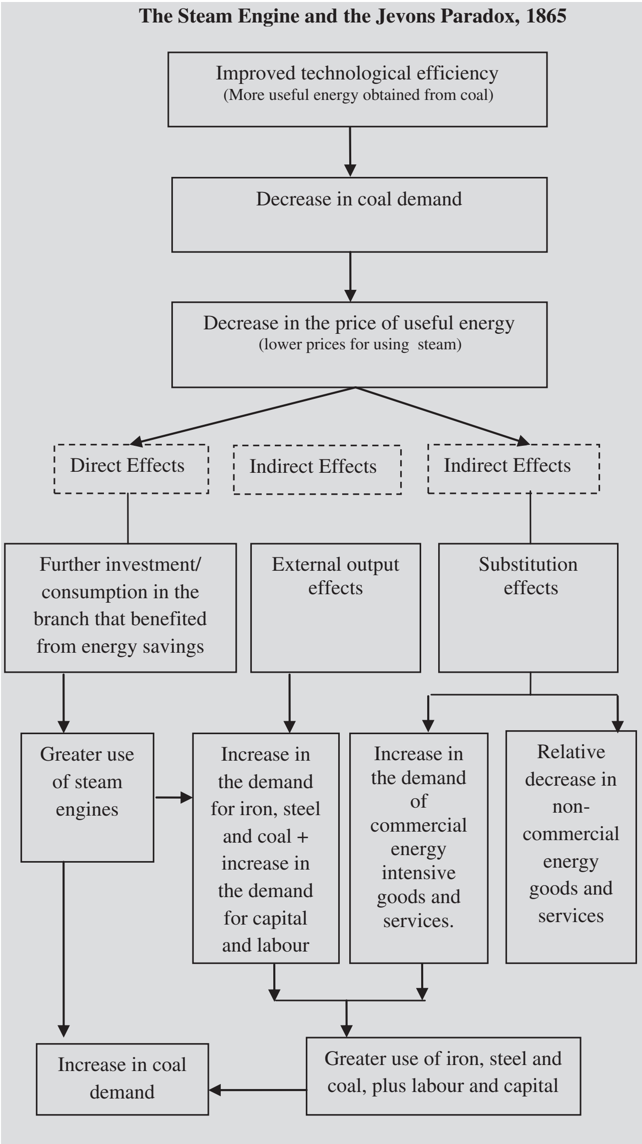
During the summer break of 1864, the Owens College tutor wrote 380 pages of a book entitled "The Coal Question: An Inquiry into the Progress of the Nation, and the Probable Exhaustion of our Coal Mines." The decision soon paid off. When the book was distributed the following year, it quickly captured the spotlight and pushed the author into the center of the public debate. The book gained such success partly because it backed the popular shortage thesis, but also because it provided a fresh economic explanation for impending coal exhaustion.

The Jevons point of view challenged established beliefs. Everyone was aware of the improvements in the thermodynamic efficiency of steam engines and how they had increased the amount of useful work obtained from any given quantity of coal. Available statistics demonstrated a tenfold rise in the "returns" provided by the best high-pressure steam engines of the 1860s in comparison with eighteenth-century atmospheric engines (Jevons 1865: 128). Insofar as this path of technological efficiency could yield continuing improvements, Britain would benefit from productive energy savings to counteract growing demand and the threat of depletion. Thus, the spirit of ingenuity which assured progress with ever less coal created a way of canceling out the effects of economic growth on natural resources, particularly if other factors also ran in the same direction, for instance, should mounting foreign production slowdown the volume of exports. Drawing on this line of reasoning, Edward Hull from the Geological Survey believed that British demand for coal would soon plateau, a view repeatedly echoed in the press. And with less consumption, the pace of depletion would slow correspondingly.

The Jevons perspective turned this argument upside down. He stated that "it is wholly a confusion of ideas to suppose that the economical use of fuel is equivalent to a diminished consumption. The very contrary is the truth…" (Jevons 1865: 123). He then reached the paradoxical conclusion that saving energy leads to more energy consumption. Owing to the peculiar formulation of this riddle, it became known as the "Jevons paradox."

The key to the argument is efficiency's effects on prices and therefore on profitability and demand. Greater efficiency works economically by lowering the price of energy with two direct consequences: an increase in the profits of trade, thereby attracting more capital into the industrial sector, and a reduction in the price of goods, thus fostering overall demand. Either way, through supply and demand, the net result is a rise in production, thus driving demand for energy: "If the quantity of coal used in a blast furnace for instance, be diminished in comparison with the yield … the greater numbers of furnaces will more than take up for the diminished consumption of each" (Jevons 1865: 125). Jevons called this mechanism a direct effect. It becomes "direct" as the energy savings in branch-specific and commodity-specific sectors prompt further investment in the same branch. As a result, part of the improvement in energy efficiency brought about by technological innovation is offset by the upward shift in production.

The author appears to have found the role of indirect effects even more remarkable. Defined as the "excitement" induced in other economic sectors by price decreases in a single industrial branch, the importance of these indirect effects tends to increase the level of business interdependency. Although Jevons was not very clear or explicit about the issue (Jevons 1865: 125–127, 135–136, 175), a reasonable interpretation is that he was primarily thinking about external output effects, for instance, producers making use of cost savings from energy-efficiency improvements to increase output, thereby increasing the consumption of capital, labor, and materials, which also require energy (text box bellow—External output effects). He was probably also considering indirect effects of a substitutive kind; in other words, both the energy-efficiency improvements and associated reductions in energy prices lessened the cost of energy-intensive goods in relation to the non-energy intensive, thereby shifting consumer demand toward the former (Substitution effects). Jevons believed that coal-based technologies would soon wipe out and actually replace the traditional non-fuel intensive sectors of windmills, animal, and water power, making Britain, and much of the developed world, ever more dependent on coal reserves.



Again, part of the coal savings induced by technological innovation (for example, generating steam power with less coal) is taken back by the demand-push conveyed through inter-dependent sectors (making more steam engines requires more iron, steel, workers, and capital which, in turn, requires extra coal). What appeared a paradox from a static point of view therefore becomes a consistent and logical explanation within the framework of a dynamic cycle, since the fall in coal consumption occurring at the moment of breakthrough *t* is formally separated from the rise in coal consumption over the period $t + n$. Due to the feedback loop in Jevons' argument, the "backfire" phenomenon was later called the "rebound effect."

An overall or economy-wide rebound effect represents the sum of the direct and indirect effects and is normally expressed as a percentage of the expected energy savings from an energy-efficiency improvement perspective. Hence, an economy-wide rebound effect of 10 % means that 10 % of the potential energy savings are "taken back" through one or more of the above-mentioned mechanisms.

The Jevons argument is grounded on the assumption that the economy-wide rebound effect is always >100 % so that the expected energy savings are entirely offset, leading to zero or negative net savings for the economy as a whole. In fact, rebound effects of more than 100 % are the backbone of the paradox. Jevons' conviction that this (worst case) scenario was the only admissible outcome in a continuous scale of rebound effects helps to explain why he was seldom seen as the harbinger of a gloomy future. It is worth recalling that the 100 % backfire assumption implies that irrespective of the path taken by technological innovation, greater technological efficiency savings would only contribute to actually stepping up the pace of fossil-fuel depletion (Fouquet 2008: 274–281; Sorrell 2009).

3.3.3 Inevitable Scarcity

What disturbed most contemporary readers of his 1865 book on the coal question was the puzzling idea that although technological progress was beneficial, it might hinder the conservation of coal reserves. In conjunction with the technological factor, Jevons factored in two more causes contributing toward ramping up the speed of solid fossil-fuel exhaustion: population growth and the effects of growing affluence and rising average income. All three—technology, population, and affluence—worked in the same direction, though technology was clearly the wildcard in the evolution of society.

In order to forecast the pace of British coal consumption, Jevons attempted to demonstrate that all three factors evolved in a uniform geometrical ratio—with uniform multiplication throughout uniform periods. Like an orderly yield of continuing progress, geometrical growth was not only as "necessary as a mathematical law" but a pattern of regularity characteristic of "man and all living creatures" (Jevons 1865: 170). Based on the sustainability of this progress, Jevons set the annual coal production growth rate at 3.5 %. As can be seen in Fig. 3.1, while the previous forecasts advanced by the industrialist William George Armstrong had already cut the depletion gap thanks to his method of fixed linear increases (a growth rate of 2.75 million tons per year), the time before exhaustion was again shortened under Jevons' geometric law—by approximately 100 years relative to Armstrong's estimate (from 212 years to about 100 years).

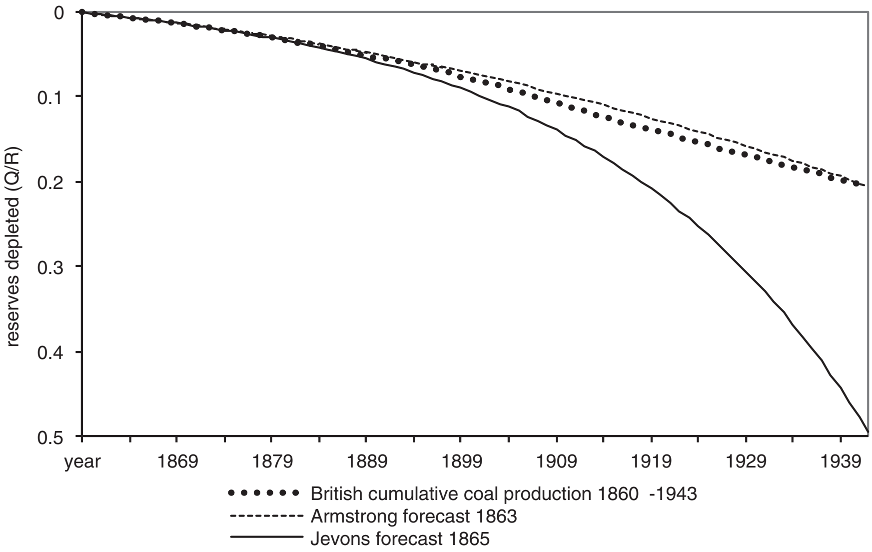


Fig. 3.1 British coal exhaustion according to nineteenth-century forecasts (1860–1943). (*Q/R*) cumulative production forecasts/recoverable coal reserves—UK 1861 geological survey; British cumulative coal production/recoverable coal reserves—UK 1861 geological survey

The disparity between the two forecasts is depicted in Fig. 3.1: while half of British recoverable coal would have been extracted by the 1940s according to the Jevons account, Armstrong predicted that only 20 % would have been removed by that time. The graph also charts the real-time series of coal production in Britain, revealing that Armstrong's simple linear method came closer to the truth than Jevons' geometric law of enhanced progress, despite the industrialist's tendency to underestimate growth potential.

Regardless of the discovery of the "rebound" or "take-back" effect, Jevons' vision of omnipresent coal made him greatly overestimate the growth trend. Why were Jevons' forecasts so far from reality? Leaving aside how easy it is to be wise after the event, two reasons stand out as possible sources of bias in the author's judgment: firstly, there was a clear overstatement of the opportunities for new industrial applications associated with the belief that the clear superiority of coal steam in terms of power capacity would swiftly replace the less powerful traditional energy carriers; secondly, Jevons was convinced that coal would prove the mainstay of contemporary civilization, and no other power source could ever be invented to replace it.

While there was still a broad margin for thermodynamic improvements in coal-fueled engines, other existing energy sources fell short of any secure future: electricity was encumbered by "fallacious notions," "miraculous effects," and all kinds of unattainable beliefs; wind power was uncertain and lacked capacity; the water wheel and turbine were not only rigidly set in fixed spots but were also dependent on winter streams and reservoirs; peat and turf were too costly; sun could perhaps be collected someday, though the consequence of such a discovery would "simply" be to "destroy British industrial supremacy." Finally, petroleum was "solely a new way of pushing the consumption of coal" and was "more likely to be an aggravation than a remedy" (Jevons 1865: 140–168).

To sum up, coal-based technology would rapidly seize the market share of traditional energy carriers and sustain long-term economic growth single-handedly, since no other substitute could ever match its performance and efficiency. Hence, the mounting demand created by the geometrical growth of population, income, and modern technology had to be supplied by a single resource. One may wonder whether Jevons' emphasis on the impossibility of substituting the existing power sources resulted from a theoretical framework borrowed from physics and based on the principle of conservation of energy: coal was irreplaceable because one could not conceive of getting energy "out of the vacant space" but only from natural sources that make use of oxygen to produce combustion, in such a way that the mechanical energy already resident in fossil fuels could be turned, or converted, into heat, light, chemical change, or mechanical motion (White 2004). Under this framework, the most feasible future source was coal rather than oil, which, at the time, was exclusively used as lighting fuel in lamps. Another possibility is that the exclusion of everything but coal was due to the need to heighten the contrast between ever-changing economic progresses on the one hand and the static and unalterable frame of natural conditions on the other. According to the logic of "The Coal Question," energy resources were seen as a fixed stock, a stock handed down in advance, and not dependent on human action. By locking the amount of physical resources into an immutable and non-substitutable aggregate, Jevons could define the role of coal as a constraint on British development, if not on overall human progress.

Working on the assumption that modern societies are torn between non-renewable resources and boundless needs and desires, Jevons drew the paradigm of economic science as the optimal allocation of externally given scarce factors of production. Along this line of reasoning, "scarcity" became a meaningful concept by playing down human inventiveness. Every element that could vary in the fixed stock of physical resources was thus downplayed: Jevons not only set aside the possibility of finding new coalfields, ignoring the probable reserves defined by the British Geological Survey (Jevons 1865: 19–20, 24, 47), but also expressed great doubts about the real chances of removing coal at depths of 4,000 feet since "nobody would be so foolish as to suppose we could go to that depth." (Jevons 1871). Moreover, Jevons clearly underestimated the chances of coming across new energy sources like hydroelectricity or solar energy which his contemporary, the industrialist William George Armstrong, had wisely considered and misjudged the recent take off of the petroleum industry on the other side of the Atlantic, as well as the feasibility of feeding reciprocating steam engines with crude oil and fuel oil.

In sum, we may certainly say that Jevons' analysis of the rebound effect and the price mechanism brought to light a fresh perception of the conditions that might foster coal depletion. However, his exclusive emphasis on a fixed stock of coal made him underestimate the effects of innovation and human ingenuity in changing a given set of initial conditions. By neglecting the potential for coal substitution, both through upgrading traditional technologies and the invention of new technologies, he made an upwardly biased forecast of coal consumption. Still more surprising, Jevons downplayed the key role of higher prices in driving supply and did not consider the spontaneous effect of resource adjustments through a "scarcity rent" embodied in the final price The omission of this sort of negative feedback effects of higher prices on slower resource depletion, though hardly understandable in hindsight, stems from the logic of proofing the scarcity thesis. At a later date, Jevons did seem to acknowledge this flaw (Jevons 1869). However, it was effectively too late to forgo on the depletion thesis that, in the interim, had become a brand mark of his intellectual stance.

3.4 The Triumph of Geology

The new economic explanation for the inevitability of coal depletion and the revised forecasts left a trail of uneasiness in British society. Like an unwelcome and suddenly revealed truth, Jevons' ideas gave form to collective fears concerning coal scarcity and, most of all, to the decline that would necessarily ensue throughout the Empire. This was partly due to the psychological effects created by the long persistence of unsettling doubts concerning the size of coal reserves. However, it was also partly due to the repercussions caused by the debate, cutting across all sorts of social and institutional networks and colored with distinct overtones: there was political discussion among parliamentary factions over the budget and the future of the nation; a business technical discussion carried on by local elites about the future of industry; and a diplomatic reflection within the imperial administration about British positioning in international competition.

The parliamentary divide was triggered when William Gladstone, Chancellor of the Exchequer, subscribed to the Jevons thesis, turning the coal question into a matter of political dissension by considering that inasmuch as fossil-fuel reserves were being consumed at an enormous rate, the British government ought to pay off, or at least reduce, the National Debt before the time of their exhaustion. Liberal and Tory realignment along the National Debt split attracted widespread public attention to the coal question.

A second channel of dissemination was set in motion courtesy of the concerns of local upper classes, businessmen, and scientists, who took the task of reexamining the forecasts for coal depletion in their districts into their own hands. To mention just the meetings reported in the press, geological debates dedicated to the theme of regional scarcity were held at scientific societies in Staffordshire, Worcestershire, Scotland, North Wales, Derbyshire, the Bristol region, Northumberland in general, and Newcastle upon Tyne in particular.

Finally, a third stream of inquiry was set into motion by the close watch on worldwide production and trade in coal. As aforementioned, the Foreign Office seized upon the available expertise of embassies and consular networks to grasp every scrap of data on coal mining, prices, exports, imports, and the future potential of producing nations. In so doing, much of the institutional framework of the Foreign Affairs was reshuffled to serve the needs and ends of these diplomatic surveys.

Caught up in this web of social and political interests, Jevons' book was thrown starkly into the spotlight. Maybe in the afterthoughts of such success, the British economist might well have had occasions to remind himself of the popular saying that warns to "be careful about what you wish for, because it might come true." Jevons certainly sensed the looming counter-attack and heard few voices rising in his defense. Within parliament, the prospects of industrial decline were increasingly targeted by critical members and by Gladstone's opponents. As indignation mounted, Gladstone was forced to comply with the appointment of a Royal Commission to investigate and to report on the quantity of coal available. Publicly, this amounted to the recognition that something might be profoundly wrong with the Jevons and Armstrong assessments. Quite naturally, the Commission arrayed some of the most-avowed enemies of the depletion thesis, namely the copper smelter and Deputy Lieutenant for Glamorgan, Hussey Vivian, and the colliery owner from Sunderland, George Elliot, future president of the North of England Institute of Mining Engineers. It also kept Jevons on the outer fringes of the inquiry (White 2010). In every respect, the stage seemed set for the rebuttal of the scarcity threat. And thus it proved. The reviewed estimates publicly released in 1871 (not without some prior leaks to the press) assured a comforting horizon of 360 years coal supplies, based on forecasts of lower population growth, lower consumption growth, larger workable physical reserves, and larger probable reserves (Report of the Commissioners 1871).

Notwithstanding the political bias behind the scenes, the 1866–1871 Coal Commission set new benchmarks in geological knowledge. For the first time, a methodology was tested out for quantitatively estimating the dimensions of undiscovered seams, turning the concept of "probable reserves" into an operational, comprehensive, and verifiable gauge. Departing from proven coalfields, the surveyors carried out systematic research to assess whether their geological boundaries were cut off from any other coalfields, or turned round to join, in a continuous underlying superincumbent stratum, other separate basins. By means of hypotheses formulated over the boundaries of geological seams and their probable thickness, the Commission arrived at a figure of 56.2 billion tons of "hidden" coal (56.2×10^9), which were added to the 90.2 billion tons of proven reserves. But it went still farther and advanced the estimate of 41 billion tons at depths >4,000 feet, some of which lay in the undiscovered coalfields of southeast England (Report of the Commissioners 1871, Vol. II: 413–526).

Taken as a whole, the economic thesis of scarcity boosted geological research to unprecedented heights, leading in just a decade to the formulation of fresh classificatory schemes that coped with the uncertainty of coal reserves. Each degree of uncertainty was placed in a separate category, with a level of certainty deriving from the conditions describing the geological event. Capriciously, however, another turnaround was about to happen. The results of the Coal Commission were released in 1871, right in the midst of an upsurge in coal prices and financial turbulence (Church 1986: 50–55). Once more, the evidence of experts was confronted by the counter-evidence of business. The "coal panic" unleashed the general fears of an age of commercial exhaustion, overshadowing much of the Report's content. Therefore, and in spite of the geological pledge that there was 360 years of secure supply, a new Select Committee "to inquire into the causes of the present dearness and scarcity of coal" was again appointed by the House of Commons (Report of the Select Committee on Coal 1873). Only afterward, in the 1880s, did the pessimistic view begin to recede, owing to the repeated restatement of larger geologic reserves and the downward trend in coal prices. Contemporaries also took note of the systematic divergences with Jevons' forecasts and concluded that reserves would certainly last longer than predicted.

Several factors strengthened the optimistic stance toward coal. At the close of the nineteenth century, a spurt of investment in mining unveiled the potential of new coalfields located outside the core nations and developed mostly in countries of low mining production (in Southern and Eastern Europe, Africa and Asia). Simultaneously, large geological formations were also confirmed in Canada, China, Germany, and Russia. Step by step, coal geology began to reveal the incidence of fossil fuels across the globe, filling in the blanks on the world map. At the same time as coal turned into a global resource, its substitution was eased by technological changes that took place in the shipping and railway sectors: the development of the first successful applications for pulverizing crude oil and blowing it into a furnace in spray form provided a successful substitution for solid fossil fuels in steam engines, allowing for the flexible switchover from coal to oil (Dunn 1916). The effect of technological substitutability, together with more widespread proven reserves, boosted the prospects for long-term coal usage.

More relaxed consumption and a recognized safety margin for the next generations meant that the debate shied away from economic and technological theories and returned to the materialistic ground of geological appraisal. The crux of the matter was again the amount of physical reserves, irrespective of price movements or consumption trends. In the belief that the time was ripe for worldwide recognition of geological knowledge, the organizing committee of the Twelfth International Geological Congress, scheduled for Toronto in 1913, challenged each participating nation to collect and submit evidence on the important topic of the world's supply of coal (Middleton 2007; Dominian 1915).

The globalization of geological estimates at the 1913 Toronto Congress confirmed prior suggestions that Britain's undisputed primacy in world exports might not correspond to its position as the best endowed nation. As far as coal was concerned, the future belonged to the largest territorial countries like the USA, Canada, and China. Russia was mistakenly categorized among the secondary ranks of energy powers. The more the debate about depletion moved away from the British perspective, the more the outlook for coal supply improved.

3.5 What to Expect from Energy Efficiency

The contention that energy efficiency may increase energy consumption above what it otherwise would be without those improvements does not seem very alluring. Besides all the gloomy predictions of impending collapse caused by the scarcity of coal, the "rebound effect" also impaired belief in the net advantages of technical progress and pushed energy efficiency into the realm of factors intrinsically bad for the environment. As ever, those most concerned with the fate of this fossil-fuel-based civilization were the richest industrialized nations as they had the most to lose should the depletion path prove unavoidable. However, just how correct was this view? How much could society gain and lose from making recourse to enhanced technologies? What have we actually learned 150 years after the discovery of the rebound effect?

In seeking to understand these questions, the first dimension to pin down is that backfire effects do not occur out of the blue but are rather constrained by the prevailing economic, historical, and technological conditions:

- *Economic conditions*. If the long-run own-price elasticity of energy demand for a particular energy service is inelastic, consumer behavior is not sensitive to the decrease in the energy cost of energy services brought about by greater energy efficiency. In these circumstances, lower real prices do not prompt increased consumption of the good or service at stake.
- *Historical conditions*. To the extent that the demand for a particular energy service approaches saturation, future improvements may be associated with smaller rebound effects. In the starkest terms, when historical developments spearhead a drive for higher levels of welfare and well-being, the marginal utility of further well-being declines. Thus, as the maximum level of personal automotive transport time, thermal comfort, or in-door lighting approaches saturation, the size of the direct rebound effect may be expected to shrink.
- *Technological conditions*. To the extent that energy-efficiency improvements take place at the maturity stage in the development and diffusion of a particular technology, and this technology has branch-specific applications, zero externalities, and smaller effects on overall productivity and economic growth, the size of the direct rebound effect may be expected to be limited in scope (Sorrell and Dimitropoulos 2007; Sorrell 2009).

In more concrete terms, these theoretical guidelines imply that direct rebound effects should attain a lower level of impact in periods with falling prices than in periods with rising prices; in the USA rather than in China; in affluent classes rather than in low-income groups; in telecommunications rather than in improved thermal insulation.

Drawing on a wide sample of econometric estimates undertaken over recent years, experts have arrived at best guesses for the size of the long-run direct rebound effect in OECD member states. The results point to somewhere in the region of between 20 and 30 % of the savings achieved in personal automotive transport being retaken by more driving; 10–30 % of space heating gains similarly wiped out, while <25 % of the gains in space cooling are applied to further cooling (Graham and Glaister 2002; Sorrell et al. 2009). Even accounting for indirect rebound effects, in which consumers/producers deploy part of their remaining savings to purchase other goods and increase output, triggering the looping effect of greater energy demand (Dimitropoulos 2007), the conclusion nevertheless remains: energy efficiency still pays off. It pays a great deal for the companies and individuals fostering innovation; it pays a little less for environmental conservation and the abatement of CO_2 emissions.

References

Archival Sources

National Archives (NA), London. NA COAL-96 (1972) A.M.C., The potential of British coal reserves Aug 1972

Bibliography

Adams SP (2006) Promotion, competition, captivity: the political economy of coal. J Policy Hist 18(1):74–95

Armstrong WG (1863) Report of the 33rd meeting of the British association for the advancement of science. BAAS, Newcastle-upon-Tyne

Baker M (1895) The U.S. geological survey. Geogr J 6(3):252–260

Baumont MM (1928) La grosse industrie allemande et le charbon. Gaston Doin, Paris

Buckley GL (1998) The environmental transformation of an Appalachian Valley, 1850–1906. Geogr Rev 88(2):175–198

Church R (1986) The history of the British coal industry. Clarendon Press, Oxford

Clark JG (1987) Energy and the Federal government: fossil fuel policies, 1900–1960. University of Illinois Press, Chicago

Dimitropoulos J (2007) Energy productivity improvements and the rebound effect: an overview of the state of knowledge. Energy Policy 35(12):6354–6363

Dominian D (1915) The coal resources of the world: a summary. Bull Am Geogr Soc 47:761–766

Dunn FB (1916) Industrial uses of fuel oil. Technical Publishing Company, S. Francisco

Fouquet R (2008) Heat, power and light: revolutions in energy services. Edward Elgar Publications, Cheltenham

Gannett H (1909) Report of the national conservation commission, Feb 1909. National Conservation Commission, Washington

Hodsworth J (1866) On the extension of the English coal-fields beneath the secondary formations. R. Middleton, London

Hull E (1861) The coal-fields of Great Britain, their history, structure, and duration: with notices of the coal-fields of other parts of the world. Edward Stanford, London

Clark JG (1990) The political economy of world energy. Harvester Wheatsheaf, New York

Garfield JR (1909) Address. In: Blanchard NC (ed) Proceedings of a conference of governors in the White House Washington D.C. Government Printing Office, Washington, p 179–183

Gillet M (1973) Les Charbonnages du Nord de la France au XIXe siècle. Mouton, Paris

Graham DJ, Glaister S (2002) The demand for automobile fuel: a survey of elasticities. J Transp Econ Policy 36(1):1–26

Hays SP (1959) Conservation and the gospel of efficiency. Harvard University Press, Cambridge, Massachusetts
Jaccard M, Bataille C (2000) Estimating future elasticities of substitution for the rebound debate. Energy Policy 28(6):451–455
Jevons WS (1865) The coal question: an inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines. Macmillan and Co, London
Jevons WS (1869) On the probable exhaustion of our coal mines. In: AAVV (ed) Notices of the proceedings at the meetings of the members of the royal institution of Great Britain, with abstracts of the discourses delivered at the evening meetings, vol 5. William Clowes and Sons, London, pp 328–334
Jevons WS (1871) The probable duration of our coal seams. Glasgow Herald, 6 Sept 1871, Issue 9885
Lozé E (1900) Charbons Britanniques et leur épuisement. Libraire Polytechnique Ch.Bérangeur éditeur, Paris
Middleton GV (2007) The 12th international geological congress, Toronto, 1913, Episodes 30. pp 290–301
Missemer A (2012) William Stanley Jevons' the coal question (1865), beyond the rebound effect. Ecol Econ 82:97–103
O'Connor JG, Meadows AJ (1976) Specialization and professionalization in British geology. Soc Stud Sci 6(1):77–89
Porter R (1978) Gentlemen and geology: the emergence of a scientific career, 1660–1920. Hist J 21(4):809–836
Pounds NJG (1958) The spread of mining in the coal basin of upper Silesia and Northern Moravia. Ann Assoc Am Geogr 48(2):149–163
Report of the Commissioners Appointed to Inquire into the Several Matters Relating to Coal in the United Kingdom (1871) House of Commons, parliamentary papers, vol 1. George Edward Eyre and William Spottiswoode, London
Report of the Select Committee on Coal (1873) House of Commons, parliamentary papers. House of Commons, London
Royal Commission on Coal Supplies, First Report of the Royal Commission on Coal Supplies (1903) Minutes of evidence, and appendix, House of Commons, parliamentary papers, vol 2. Wyman & Sons, London
Saunders HD (1992) The Khazzoom-Brookes postulate and neoclassical growth. Energy J 13(4):131–148
Schurr SH, Netschert BC (1975) Energy in the American economy, 1850–1975. Greenwood Press, Westport-Connecticut
Sorrell S, Dimitropoulos J (2007) UKERC review of evidence for the rebound effect: technical report 5—energy productivity and economic growth studies. UK Energy Research Centre, London. Available via www.ukerc.ac.uk/support/tiki-index.php?page=ReboundEffect. Accessed 18 Jan 2013
Sorrell S (2009) Jevons' Paradox revisited: the evidence for backfire from improved energy efficiency. Energy Policy 37(4):1456–1469
Sorrell S, Dimitropoulos J, Sommerville M (2009) Empirical estimates of the direct rebound effect: a review. Energy Policy 37(4):1356–1371
The Coal Supplies of England (1872) The Derby Mercury, 17 January 1872, Issue 8218
The Treaty and the Coal Question (1860) The Caledonian Mercury, 28 February 1860, Issue 21973
White MV (1991) A biographical puzzle: why did Jevons write the coal question? J Hist Econ Thought 13(2):222–242
White MV (2004) In the lobby of the energy hotel: Jevons's formulation of the postclassical economic problem. Hist Polit Econ 36(2):227–271
White MV (2010) A revised bibliography of publications by W. Stanley Jevons. Hist Econ Rev 51:106–128

Chapter 4
Technological Hybridization

Abstract This chapter draws on the two major technological hybridizations that have occurred in the field of energy: the current hybrid electrical vehicle that combines the internal combustion engine with an electric battery and the oil–steam hybrid in which fuel oil was adapted to feed steam engines originally designed as coal-fueled. The analysis highlights some of the problems inherent to hybridization processes, specifically investment in new network infrastructures; the consolidation of a technology with a dominant design; and the operational asymmetry between high specific energy and high specific power. Our study of the oil–steam combine embraces its diffusion across leading producing nations such as Russia and the United States, its diffusion into industrial and transport activities in South America, and its spread throughout European navies. We show how this process of hybridization entailed the transformation of oil into a geostrategic good and triggered an international scramble to seize sources of this natural resource.

4.1 Hybridization and Operational Asymmetry

The coordination of multiple energy sources within a single power device represents the key feature to technological hybridizations. This coordination may involve different engines or different fuels. Whatever the case, the system always needs regulating through recourse to specific hardware for energy exchanges. The presence of power converters, power flow controllers, switchers, or boiler adapters accordingly provides a physical indicator of the hybridization process. Moreover, history has shown how technological hybridization requires the prior consolidation of a dominant design technology (see also Chap. 9.3.1), investment in a new infrastructural network, and a sequence of incremental innovations. Owing to their long periods of maturation, hybrid technologies are initially channeled toward the supply of market niches.

Another distinguishing feature stems from operational asymmetry. Though hybrids bring multiple energy sources into play, there is always one technology or fuel that typically constitutes the dominant system, while the other fulfills the role

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_4,

of a secondary or support system. From this point of view, hybridization seems to imply a trade-off between high specific energy and high specific power, between normal modes of operation and intensive modes of operation (force, acceleration). System optimization in such a way that the unique advantages of each energy source are fully utilized represents the golden key to enhancing overall efficiency (Chau and Wong 2001). Toyota, the producer of the first large-scale hybrid vehicle to be launched globally, developed a system called *The Hybrid Synergy Drive System* that "automatically switches between the electric motor and the internal combustion engine depending on the power needed to move the vehicle. The intuition is to rely less on the petrol fuel engine at low speeds and use its full capacity when more power is needed" (Zapata and Nieuwenhuis 2010).

This coordination type is characteristic of the standard "hybrid-electric vehicle" that deploys an internal combustion engine as the primary system, with an electric motor adapted to power the vehicle for short distances or in support of the main motor. However, other alternatives are also possible. The "plug-in hybrid-electric vehicle," for instance, relies on a battery recharged from plug-in electricity as the primary system and resorts to a small on-board internal combustion engine to boost battery recharging. By this means, the up-and-down motion created by the fuel burnt in the internal combustion engine does not have to be conveyed to a crankshaft and converted into the turning motion rotating the wheels but is instead used to directly feeding the battery with mechanical energy. Unlike the two energy storage systems in the first example, plug-in hybrid-electric vehicles are equipped with but a single energy storage system.

One may therefore conclude that operational asymmetry exists irrespective of differences in design and conception. The same applies to the requirements for a dual infrastructural network of stations for electricity recharging and gas filling. Reflecting on the framing of high-tech paths among the automobile industry, Zapata and Nieuwenhuis (2010: 16) have stressed the strategic interests of car manufacturers over their choices of particular technologies from the perspective of protecting their investments and market share: Leading producers of hybrid-electric vehicles such as Toyota and Honda did not have to "abandon their existing investments in internal combustion engine manufacturing technology. They have merely to add another element to the established internal combustion powertrain, thus safeguarding their sunk investments. In this sense, from the point of view of a car manufacturer, it is more akin to adapting an existing Internal Combustion engine to run on alternative fuels, rather than replacing an existing powertrain manufacturing system with a different system altogether."

Aside from this factor, improvements in operational asymmetry also hinge upon the evolution of fixed and variable costs, and on the expectations of relative fuel prices. Since hybrids are prepared for the switchover between alternative energy sources, the industry is quite sensitive to price alterations that may favor one pattern of consumption/equipment over other. Likewise, they also know at which relative fuel prices hybrid equipment becomes most competitive. Above all, hybrid's manufacturers understand quite well the customary quandary faced by their customers: to pay more for one-off fuel-efficient vehicle or to pay more for higher life cycle fuel costs? Save money later or save money now?

4.2 Fuel Oil: A New Energy Source

4.2.1 Coal and Oil

This chapter addresses the first large-scale phenomenon of energy hybridization in the contemporary world: the adaptation of steam machines, originally devised to be coal powered to instead burn the "new" energy source of fuel oil. This change led to the flexible switchover between energy carriers adding a further power source to feed steam engines. Logically, steam power from coal constituted the dominant technology, while oil-firing evolved as an emerging market niche. Equally, the major drawbacks to the innovation spreading lay in the absence of a network of bunker stations with fuel oil stocks similar to the coal bunkers already in effect. The network issue was particularly important in the case of steamships because resupplying bunkers with oil proved hard to manage in zones with scarce levels of petroleum discovery and negligible production outputs such as Asia, Africa, and the Middle East (until the 1930s).

Chronologically, the gap between steam and internal combustion technologies covers the period from the 1880s to 1930s. The complementarities between coal and oil during this transition were so overwhelming that they set the stage for the commoditization of oil, transforming a tradable commodity into a chief geostrategic good for international competition.

The next subchapter traces the evolution of the oil industry in the United States and Russia and portraying how the creation of a global market for kerosene stimulated the emergence of regional markets for other petroleum by-products, in particular fuel oil. As it was not possible to produce illuminants without also obtaining large quantities of fuel oil, international economies in lighting supply had to mesh with regional economies in power generation. The ensuing section highlights the invention and consolidation process of the market for these "secondary" and poorly valued by-products obtained by distilling crude, a process marked by the hybridization of existing steam technologies. In the final section, we emphasize the fact that "in-between technologies" may generate bold macroeconomic consequences as well as explaining the circumstances in which combined oil–steam hybridization led some countries into taking distinctively different energy consumption paths.

4.2.2 US Oil Industry Markets

The discovery that seepages of "rock oil" flowing from below the ground could unveil astonishing reservoirs of petroleum prompted a rush for deep drilling in Pennsylvania, the United States and, later in Azerbaijan, Russia. From 1859 and 1870 onward, the contemporary oil industry began to take hold and the market swiftly organized the downstream areas of transport, storage, packing, refining and distribution. One of the most distinctive traits of this "build and grow" phase was the existence of a dual economy in the markets for products obtained by refining: While some products came to the fore as global goods, e.g., kerosene for illumination and lubricating oils for greasing machines that were soon integrated in worldwide distribution networks, the remaining by-products (naphtha, gas oil, fuel oil, paraffin) lagged behind as regional-tradable goods which meant they could only be sold near their production source. From a technical standpoint, the bulk of these poorly tradable by-products came from the "bottom fraction" of crude oil—the last fractions with heavier molecules removed from the stills or leftover as residuum.

In the late nineteenth century, all distillation—the heart of refining operations—was performed in "straight-run" stills by gradually raising the temperature so as to drive vapor from the boiling oil along a pipe. Once in the pipe, the steam was normally condensed to a liquid by means of a condenser box with many pipes containing running water. When the specific gravity of the distillate reached a certain level, thus becoming too heavy, it was separated or "cut" by running the distillate into another tank. Due to the application of more intensive heat, the next component of the crude was then removed by means of steam and cooling devices again. Since each successive still with its higher temperature was placed below the preceding one, gravity allowed the oil to flow steadily through the entire batch. The technical jargon—"batch-operation" captures precisely the system's basic feature. In the end, the different fractions were separated by boiling edges ranging from the lighter by-products of naphtha and kerosene to the medium boiling range of gas oil, and to the "bottom" of heavier distillates: paraffin wax, lubricating oil, and fuel oil. The "batch-still" process was accompanied by other methods that consisted of simply skimming or topping, i.e., distillation of the naphtha and kerosene with steam and fire followed by the extraction of the fuel oil.

Significantly, most of the technological improvements that appeared in the United States sought to improve the separation of the different "cuts" and, whenever possible, to convert the fractions of little market values to those of greater appeal (Williamson and Daum 1959: 253–308). At a time when the demand for private lighting was reaching extraordinary heights (Fouquet and Pearson 2006; Fouquet 2008), the kerosene "fraction" naturally became the most valued product and the mainstay of the oil economy. Moreover, its use in lamps had a profound effect on everyday life from Europe to the Pacific. The efficiency of kerosene was similar in candle-hours per unit of energy input to the existing alternatives of "portable" light–tallow candles, colza oil, and whale oil, but it was cleaner, burned without unpleasant smells, required only half the storage space, and, above all, was much cheaper. Estimates of the prices of the light flux (dollars per lumen hours) with the available technology of nineteenth-century lamps revealed that kerosene was 14–20 times cheaper than other substitutes of vegetable and animal oils (Nordhaus 1998). On the basis of the high demand for this new substitute good, some of the first contemporary multinational corporations soon concentrated trade along five major routes: North America to Western Europe; North America to South America and the Pacific; Rumania and Poland (Galicia) to Central and

Western Europe; Dutch East Indies to East Asia; and Burma to South Asia. As the integration process and competition in these markets began to take hold, the sales of the other by-products from crude oil receded to the backstage as regional-tradable, or even non-marketable goods: In the USA, some fractions below the boiling range of kerosene, e.g., the volatile naphtha fraction, were generally thrown away with ecological impacts on lands and lakes; at best, the heavier fractions of gas oil and fuel oil were used by the oil enterprises for their own consumption in boilers and, at worst, "run into lakes of liquid petroleum which were set on fire to get rid of them" or alternatively "carried by pipes into the sea" (Donkin 1894: 266). Whatever the case, fuel oil had little or no market value (Gerretson 1955; Williamson and Daum 1959).

There are two explanations for why oil was wasted in this manner: First the light density of the Pennsylvania crude made it suitable for refining into illuminants allowing yields of 65–75 % of kerosene until security and technical regulations were set. The potential for developing economies of scale for others by-products was substantially reduced as only small amounts of residuum were left over. Secondly, petroleum was discovered in the vicinity of major coal producing centers, thus hampering its acceptance as a substitute fuel. Overall, the shortcomings that affected the bottom fractions of oil enabled the reinforcement of a highly competitive and globalized sector based on the "upper" fractions of illuminants.

4.2.3 Innovation in Russia

The situation in Russia was quite the reverse. A particular combination of circumstances led the petroleum industry around the city of Baku, Azerbaijan, to concentrate on the production of fuel oil, with kerosene as a by-product. The chemical properties of Russian petroleum, its high density, and the fact that the bottom fraction represented an average of 70 % of the oil distilled in the stills (vis-à-vis 13–18 % in Pennsylvania) explain why the residuum was the chief commercial commodity from the outset. Though this refuse could not be volatilized by the application of heat, it could be broken up or divided into spray and used by injecting air or steam into it. Atomization allowed the oil to be burned in boilers just like coal, thus providing a technological bridge for its use in steam engines.

However, it was not only Russia's "comparative advantage" in terms of natural resources that was at work. Lack of capital and the dispersion of entrepreneurial initiative locked the 140 small refineries that appeared across the region into very primitive refining techniques resulting in high levels of waste, appalling environmental conditions, and poor-quality finished products. At the same time, the geographical location of the Baku oilfields near the Caspian Sea did not make export to European markets easy; on the contrary, it made the region inward-looking, strengthening local supplies in the perimeter of the Caspian and in southern parts of inner Russia via Astrakhan and the lower Volga (Leeuw 2000; Gerretson 1955: 212–217). Finally, and again in contrast to the US situation, not only was the oil province of Azerbaijan completely depleted of wood but it was also far from the Donetsk Basin mines, which were Russia's main center for coal production (Elliot 1974). This scarcity of energy sources enhanced the potential for petroleum's new usages.

To sum up, the specific quality of Azerbaijan's heavier crudes, combined with poor capital and technology, high transport costs, and depletion of natural resources, pushed the Russian industry toward the regionally tradable market for fuel oil and made illuminants a secondary (by-) product. Given these constraints, it would seem to follow that the Russian industry should evolve along the track of slow capital accumulation, conceivably stepped up by some imperial reform. However, development proceeded at breakneck pace: Azerbaijan was turned in one fell swoop into the experimental laboratory of the world industry by the early arrival of foreign capital, foreign entrepreneurship, and technical ingenuity. Most observers described the oilfields as "deserts literally caked with petroleum which solidified into asphalt on which no vegetation could grow" with "valuable by-products burned off or passed into the Caspian," thereby transformed into a second Black Sea (Hewins 1958: 24), thousands and thousands of workmen living in damp, unlit, dark, dirty barracks where three sleep together in a small uncovered cot" (Bey 1931: 18). However, on their arrival by chance in the region, the Nobel brothers foresaw a land of gilded opportunities, where most men discovered the materialization of hell on earth. Robert and Ludwig Nobel brought to Baku the experience and capital of one of the most distinguished European industrialist families, their personal engineering expertise, and contacts with Swedish and British enterprises at the forefront of technological innovation, close acquaintances in Moscow governmental circles, and the capacity to influence Russian policy (Fursenko 1990: 69–75). Nevertheless, it was the cascade of innovations that they introduced in a short time span that made the arrival of this special breed of foreigners astonishing: The purchase of the giant Balakhany oilfield near Baku in 1874 set the stage for the subsequent building of a modern refinery (1875); the ground-breaking installation in Russia of a pipeline system powered by a 27-horsepower steam pump (1877); the design, assembly, and launch of the world's first tank steamer, the Zoroaster, with revolutionary innovations such as the use of Bessemer steel and an improved system for boiling oil as a substitute for coal (1878); the successful experimentation of the new technology of continuous multistill distillation for refining oil (1881); the application of chemical purifiers to improve the color and flash point of the kerosene obtained in distillation (1881); the adaptation of the American prototype of railway tank cars (1881); and the innovative recruitment of a geologist to a permanent position in a corporate petroleum undertaking (1885) (Owen 1975; Tolf 1976; Ratcliffe 1985; Fursenko 1990).

Thereafter, the largest Russian companies benefited from Ludwig Nobel's policy of making no secret of his achievements and soon began to emulate Nobel's best practices. The spectacular growth in output from Baku fields was due both to increased productivity in refining and transport and also to new discoveries in the relatively unscathed reservoir: From 1874 onward, it was possible to strike a well and unleash a "petroleum fountain" or gusher that delivered 6,500–43,000 barrels every 24 h.

On top of these exceptional historic conditions, the transportation barrier that had largely confined the petroleum industry to the internal market was also removed. In fact, this barrier had a name: the Ottoman Empire. As long as the Ottomans were able to rule on the Black Sea and maintain a tight control over the Bosporus strait, inner Russia was cut off from direct links with the outer European–Mediterranean routes. The opportunity to redress this geostrategic equilibrium arose when a wave of rebellions broke out in the Balkans in 1875 and 1876. Calling upon its Pan-Slavic mission, Russia declared war on the Ottoman Empire and dealt a severe blow to the caliphate's ambitions. Despite having to split the spoils of war with other European powers at a special international peace conference in Berlin in 1878, Russia was able to uphold its conquests in the Black Sea (Cleveland 1994). After that, the "Russification" of the new territories proceeded as fast as possible, and within 5 years, the railway connection between the oilfields of inner Baku and the free-trade harbor of Batumi on the Black Sea Coast was completed. The ensuing oil boom heightened Russia's advance toward the Mediterranean route.

Thus, by the mid-1880s, the industry was in fact moving toward a mixed production system based on internationally and regionally tradable goods. The spurt of technological innovation combined with the access to new commercial routes laid the ground for sales of good-quality kerosene obtained with a better yield in refining, as well as of the fuel oil residuum, called mazout in Russian. The complementarities between these by-products can be understood as a self-reinforcing process supported by economies of scale and economies of scope: The cost of production per unit falls as the output from refining increases (scale), but the average total cost also falls as the number of different goods produced is augmented (scope) (Chandler 1994). Such double-edged dynamics meant that the larger the sales of Russian kerosene in world markets (Europe, but also the entire region east of the Suez Canal after 1891), the more industrial applications had to be found to substitute fuel oil for coal (Henriques 1960: 27–34).

One thing could not advance without the other: The seizure of competitive international markets pushed the growth of homeland demand through innovation in energy carriers. As a consequence, at the end of the nineteenth century, Russia moved toward a path of primary energy consumption that was distinct from all other nations (Etemad and Luciani 1991; Mitchell 1992).

4.2.4 Fuel Oil's Market Share

In the meantime, the oil industry in the United States also underwent change. Following the discovery of new oilfields in Lima, on the border of Ohio and Indiana, in the mid-1880s, in California in the 1890s, and in Texas and Oklahoma in 1901, as well as the first imports of Mexican crude, heavier oil flooded the markets, some of which was unsuitable for refining into kerosene (Yergin 1991). Thereafter, the conversion from coal into oil proceeded swiftly, albeit later than

Table 4.1 The production of oil substitutes for coal in the US and Russia, 1909–1935, in thousands of US oil barrels (42 gallons)

Country/date	Crude oil production (1)	Runs to stills (2)	Fuel oil* marketed (3)	Crude oil used as fuel (4)	Potential substitution of coal (%) (3) + (4)/(1)
USA 1909	183,171	120,465	40,475	50,720	49.8
USA 1925	763,743	740,500	370,990	90,145	60.4
USA 1935	996,596	965,310	364,890	24,400	39.1
Russia 1910	62,187	52,875	19,915	8,380**	45.5
Soviet Union 1925	52,535	37,770	19,815	11,925	60.4
Soviet Union 1935	184,931	157,645	62,780	18,815	44.1

(Russians likely to exceed 1927; Gerschenkron and Nimitz 1952; Hassmann 1953; Schurr and Netschert 1975)

in Russia, and attained particular scope among the sectors of industrial power, railways, steamships, electricity production, and space heating. Small portions of the residuum also started to be used for road-building bitumen, particularly after 1918. The switch in the United States was fostered not only by the availability of ample stocks of heavier crude in the west and southwest but, more particularly, by a very sharp downward trend in prices after the staggering discoveries in Texas and California: Oil made its mark in comparison with coal, thanks to crude barrels as low as 30 % of the reference coal's price (Pennsylvanian crude) (Williamson and Daum 1959: 39).

Despite taking different paths, Russia and the United States had both established robust branches in internationally and regionally tradable products of the oil industry by the beginning of the twentieth century. Moreover, Table 4.1 shows that the evolution of the regionally tradable branch was quite similar in the two nations. Columns 3 and 4 of the table present data on the two industrial products used for coal replacement: first the above-mentioned residuum of fuel oil and secondly the consumption of non-refined crude for general fueling purposes. Together, their relative share of oil production amounted to almost 50 % by 1910 and increased to 60 % in 1925. This means that the by-products that could compete with coal attained the largest share of oil usages both in Russia and the United States. However, the relative importance of crude oil for fuel steadily declined over the years as the industry matured. Since crude provided approximately the same thermal energy as fuel oil at only a fraction of its costs, it was a very cheap alternative. But the experts soon realized that "by reason of its searching and corrosive effects, crude oil had a greater tendency than refined oil to attack the steams and tubes of modern boilers" (Melville 1904: 431). Whenever forced-draft situations were at stake, this was a reckless option: In normal conditions, the low (variable) costs of crude entailed high costs of maintenance, cleaning, and repair of capital goods. Coal substitution was therefore based increasingly on fuel oil, and this trend was hastened by the tide of technical regulations and standardization procedures for hybrid steam engines that were introduced before the First World War and in the 1920s (Dunn 1916; McAuliffe 1927).

Judging by the preeminence attained by coal substitutes, it is possible to see how the emergence of the new world market for illuminants drove the leading producing nations, Russia and the United States, along the distinct path of adapting oil to the steam age. But this was a national/regional phenomenon from the outset. Fuel oil exports only attained any significance in intercontinental trade after 1911, with the conversion of specific vessels to carry “dirty cargoes.” Even then, fuel oil as a tradable commodity was largely limited to coal-poor enclaves: This was the case in the Caribbean and South America, where North American enterprises established controlled branch marketing stations to extend the kerosene sales to other petroleum by-products: in the Mediterranean and the Arabian Peninsula, where the British Anglo-Iranian Company set up a chain of bunkering stations along the shore; and in India, where the British government contracted regular fuel oil supplies for the Indian railways (BPA. Doc. 136390 Historical Notes 1916; Gibb and Knowlton 1956; Ferrier 1982).

4.3 The Hybridization of Steam Engines

4.3.1 Oil-Firing Burners

The initial contrast between the US and Russian industries meant that Russia was under greater pressure to discover efficient technologies, because of its excess residuum throughout the nineteenth century. Then, at the start of the new century, the mix of products refined in the two countries converged, so that the incentives to use oil as a source of power became more evenly spread. The following pages describe the history of this process. Switching steam engines into hybrids fueled by oil was a cheap, easy, and reversible process, simply involving the installation of new boilers, burners, and tubes without any interference in the engine itself. Since only the infrastructure for the furnace and storage rooms was affected, the conversion was a technological add-on, which could be removed in the years ahead if circumstances changed significantly. It is worth remembering that, irrespective of the pace of discoveries and estimated reserves, fears of shortages in oil supply were an onerous issue for private and public decision-makers (White 1920; Hassmann 1953: 35–42; Dennis 1985). Perhaps the most striking fact about this technological add-on is how a “light” and reversible technical device had such an enormous effect on oil consumption patterns and transformed the primary energy balance of producing countries. Indeed, it was a case of minor and decentralized changes cumulatively paving the way for an overall shift in the macroeconomy.

From the outset, Russia was the hub around which most of the innovation revolved. Drawing upon the first successful applications to pulverize raw oil and blow it into a furnace in the form of spray, Thomas Urquhart, the superintendent of the Griazi–Tsaritzin Railway, converted some coal-burning locomotives to oil burners in 1882. As this experiment obtained good technical and economic results, all circulating material running on the 423 miles of the railway line was gradually switched to oil. Given the scale of the enterprise, the Griazi–Tsaritzin Railway set a technical standard that was soon to be studied and emulated by American, Dutch, English, and other foreign engineers.

Urquhart's ideas stemmed from the principle of conducting the oil through a central supply pipe onto a diaphragm from where it was driven into the furnace by a separate steam spray (an invention tested by the Russian Spakovsky in merchant ships in 1870) (Snyder 2001), petroleum together with upgrading a conical head with spiral grooves that gave the flame a rolling motion on entry (an invention tested by Ludwig Nobel in his tank ship in 1878) (Tolf 1976: 70–71). Building on these principles, the Urquhart burner attempted to feed the furnace evenly and obtain a more uniformly distributed spray. After several tests conducted in experimental settings, Urquhart devised a burner where the steam tube and the oil mingled at the mouth of the nozzle and were injected as a fine spray into a fire box. At this junction, there was an opening to the atmosphere through which the air was drawn by suction to the nozzle. The air, steam, and oil together triggered a mingled blast that broke the oil up into a very fine spray. Moreover, the overall process was facilitated by the design of a flare-shaped slotted opening that ensured that the jet spray was distributed in a fan-tailed effect (Donkin 1894: 275).

The reports written by Urquhart underline an overall 43 % decline in power costs, as well as savings on engine repair owing to the absence of sulfur in the oil. This was despite restrictions imposed by the Russian government that prohibited the railway company from adopting fuel oil unless they could prove that the fire box could be changed at a moment's notice to burn coal. Caution was the keyword, and the reversibility clause was to be fully implemented, with imperial support, when fuel oil prices soared in 1907 (reaching 43 kopecs per pud, double that of 2 years earlier) and a temporary return to coal was the order of the day (Leeuw 2000: 73). Likewise, caution was the tenet of the American naval experts who, at the height of the oil boom in 1904, recommended that "no design of fuel oil installation should be permitted for marine purposes which would not permit the renewal of all grate and bearing bars within 24 h, so that a return to coal could be accomplished within a reasonable time in case of failure in the oil supply" (Melville 1904: 430). The commercial flexibility of private enterprises added to the geostrategic flexibility of energy switches as they installed coal and oil burners side by side to take advantage of price differentials in bunker stations for merchant ships, particularly between Suez and the Gulf of Mexico and the Pacific (Growth of world's bunkering 1924).

Despite reservations, there was a general feeling that the fuel oil age was at hand. In the early 1900s, a spurt of entrepreneurial invention extended the catalog of burning devices to every possible application, from portable burners operated by a single man to burners and furnaces for home heating, sugar and rubber plants, and metallurgical and shop furnaces. On the technical front, the range of options also grew, with two new possibilities that evolved alongside improvements in the Urquhart-type burner for steam atomization: The first was spraying induced by compressed air, and the second was spraying induced by mechanical atomizing. One company from Hannover, with a branch in Pennsylvania, achieved great success with their mechanical system based on the joint action of pump pressure on the oil and the centrifugal rotary action of screw guide blades. Named the Korting system after its inventors, this type of burner was particularly suitable for ships because it avoided both the load of fresh water for steam atomizers and the risks of compressed atomizer malfunction at sea. Moreover, the finer mist of oil droplets delivered by the whirling motion had the benefit of a lower smoke emission; this drew the attention of the most powerful navies in the world, pledged as they were to upholding the element of surprise in sea combat (Snyder 2001: 133–144). However, in terms of efficient combustion, the Korting mechanical burner did not differ much from its predecessor, the Urquhart steam burner, as both were able to evaporate a similar amount of water by unit weight of fuel oil (around 14 pounds of water per pound of fuel oil) (Donkin 1894: 275–277; Melville 1904: 320–335). The comparison with the coal ratios of 7–8 pounds of water per pound of good-quality bituminous coal is therefore overwhelmingly significant.

4.3.2 Margins of Decision

From the steam engineer's viewpoint, oil had two thermal advantages: A similar quantity of steam power could be produced with a smaller load in stockpiles and transport and a larger amount of steam power, and therefore energy, could be produced with the same load in stockpiles and transport. The first idea suggests indirect savings in costs, while the second implies power intensity and speed. Either way, oil renewed the possibilities of steam technology, whose technical development appeared to have slowed almost to a standstill by the eve of the First World War. Entrepreneurs, politicians, and military men were able to alter their previous set of options by upgrading the existing engines with a higher grade of fuel using low-cost burners. The new margins in decision-making included the preference for alternative costs of fuel, the preference to save space, time, and work hours and the preference for increased power and speed. From this perspective, it is no accident that the buzz phrase "collateral advantages of oil" gained strength within the industrial milieu. This expression means that, even accounting for lower prices of coal per unit of energy delivered, the "collateral advantages" could well tip the balance in favor of oil (Kewley 1922). If the technological options are run through several margins of decision-making in which price is just one of several factors, it is worth taking a more detailed look at each aspect that might foster the hybridization of steam: relative prices, indirect savings, and power intensity.

Taking relative prices, and considering the evolution of the real technical cost of producing steam by coal and fuel oil in US cents per pound of water evaporated, we may notice that in the run up to the First World War, the price gap was very limited, although those burning coal retained a slight advantage. In other words, despite coal's lower thermal efficiency, it still remained a good economic option in terms of the final cost of the energy produced (Enos 1962: 292–293; Schurr and Netschert 1975). This context led to the consolidation of a geography of prices around the major producing regions, with both fuels holding sway over their hinterland and losing competitive ground whenever burdened down with excessive transport costs to more distant areas.

In addition to the effects of the war, a major change took place in the relative costs of steam: Not only did the gulf widen in favor of cheap fuel oil, but also this shift was further enhanced by problems in adjusting coal supplies to peace times. Massive strikes in 1919 and 1922, disruption of traditional markets, excessive hoarding, speculation, and failure of cooperative regulation all ravaged the competitiveness of solid fossil fuels (Hawley 1968; Schurr and Netschert 1975: 62–78). Globally, this imbalance confirms the idea that fuel oil had a new competitive edge in the 1920s. To the delight of oil burner enterprises, inter-fuel substitution advanced swiftly and receded only at the beginning of the 1930s (see below).

As stated above, indirect savings were a second margin of decision-making. Fundamentally, the savings result from the physical and chemical properties of oil. This happens in part because it can be moved simply by pumps and pressure devices, thus reducing the number of workers assigned to filling the tanks and feeding the boilers. Moreover, due to its high caloric content per unit of weight, oil enables a cut in both the storage space and its weight. Most contemporaries heralded these achievements but overlooked the fact that they followed in the footsteps of a long history of improvements. In fact, the adoption of the compound steam engine, the triple expansion engine, and the steam turbine had already almost halved the average coal consumption of engines in use between 1870 and 1914 (Ville 1990; Mohammed and Williamson 2004). However, changes in engines were discrete and continuous, as opposed to the abrupt upheaval caused by burners—something that explains the enthusiasm for the novelty. Another difference is that while the development of engines prompted the substitution of energy by capital, the introduction of oil burners led to the substitution of labor by energy. The switch to higher-grade fuels had an immediate impact on the full range of services required by an energy carrier with minimum lower capital costs. It was precisely this change in the quality of energy services that Sam Schurr highlighted as the key driver to broader economic productivity (Schurr 1984). For instance, the savings created by the conversion of a state of the art, 9,000-ton seagoing steamer amounted to 1,030 m^3 of storage space previously reserved for coal, of which 700 m^3 of net area could be reserved as new cargo space. Furthermore, the operation of the boilers meant that the workforce in the fire room could be halved and the time taken to refuel the tanks cut from a minimum of 30 h to just 5 h (Dunn 1916: 158; Growth of world's bunkering 1924; Hardy 1931). For the economies of railway and ship transport, these kinds of saving in space and manpower were particularly appealing, provided that fuel was available in different places at competitive prices.

In land installations that were based mainly on working furnaces, the switch to oil yielded similar results, particularly as hard muscle power was replaced by the supervision of supply and storage. The fact that “the work of firing requires no physical exertion” and that “a clear eye and common sense is all that is required” was fundamental here (Dunn 1916: 16). Moreover, a side effect of transforming uneven operations into flow processes was that the regularity and control of the combustion enhanced the overall thermal efficiency and reduced the wear and tear on equipment. The third and final point is that the conversion to oil could also raise power intensity thresholds. Although few people seemed interested in pushing this option too far at the end of the nineteenth century, it was noticed by the engineers and high ranks of European, American, and Japanese navies, who sought every inch of technological progress that might tilt the balance of power. Raising speed a notch higher by adapting oil burners in navy ships meant new prospects for tactical and strategic mobility and hastened the ongoing transformation of the fleets. In effect, speed became a key issue amid a feverish armaments race framed by burgeoning theories of naval warfare. Operational asymmetry had now turned into a trump card.

Conversion to oil could give a 20 % head start in speed. From 1895 to 1907, naval strategists interpreted this advantage as particularly suited to non-capital ships that could act separately from the main battle fleet but were in close contact with the command. The recent introduction of wireless telegraphy, combined with the offensive and defensive capabilities of torpedoes equipped with guidance mechanisms, had the dual effect of a more widely distributed flotilla and the increased value of specialized ships. The Italian navy was the first to advance along the “speed” path, in 1895–1896 launching two armored cruisers equipped with oil- and coal-fired engines. Although these cruisers had mainly scouting missions, the Italian concentration on the Mediterranean enabled the military to focus on high speed at the expense of seaworthiness, with subsequent reinforcements of exclusively oil-fired destroyers and torpedo boats (Sullivan 2001). By the end of the century, the French were committed to defense mobiles and guerre de course, a warfare strategy also based on speed, while the United States, Japan, and the British navy envisioned an approach based on compromise, in which fuel oil usage was limited to coastal defensive ships. Germany, with no access to secure oil supplies, refrained from these innovations, opting in favor of “big gun battleships” (Halpern 2001; Evans 2001; Snyder 2001). In any case, the cumbersome issue of strategic logistics and tactical refueling heightened military planning, with a twofold supply of energy carriers to the fleets and especially to the lighter steamers of cruisers, destroyers, and torpedo boats. Throughout this process, the institutionalization of special Boards, engineering staff, experimental stations, storage facilities, and private interests established socioeconomic infrastructures for the cause of “fuel oil adoption” (Lyon 1977; Shulman 2003).

After 1907, naval rivalry underwent significant change. The trade-off between speed on one side and endurance and offensive power on the other began to tilt in favor of offensive power. With the launch of a new type of heavy battleship, the dreadnoughts, equipped with centralized fire-control systems with twice the hitting power and twice the effective fighting range of the latest battleships, the advantage of speed was whittled away (Lautenschläger 1983; Lambert 1995).

In addition, the new political realignment of the British, French, and Russian alliance against Germany set the stage for an armaments race in which new battleships and new submarines became the keystones of naval strategy. In the context of big, integrated, and powerful "blue water gun platforms," conversion to fuel oil ceased to be a priority for most countries, though not for the United States, whose Navy Department took a step forward and decided to commission two battleships fueled exclusively by oil. In 1913, the US Navy adopted the policy of building oil-burning vessels only (Hamilton 1933); similarly, after a prolonged debate and a series of vacillations in fleet conception, in 1914, the British government resolved to form a fast division out of the new dreadnoughts. Once the large ships were converted to oil, the smaller vessels naturally followed suit. What began as a quest for tactical speed in specialized navy steamers ended up as a bold change in the fueling of the most important world navies.

The overall process depicts the pattern of a self-reinforcing sequence of events, where previous commitments had a cumulative effect on the decision at hand. Yet, despite continued knowledge, human resources, and institutions, there was a shift over time in the meaning of the decision to switch to fuel oil and to speed. Thanks to the technological innovation that came along with the dreadnought age, "speed" was reallocated to serve the tactical mobility of the main fleet in its defensive and offensive extensions. Most of the advantages of fuel conversion could therefore be summed up in the establishment of the same logistical and tactical standard for the fleet's mobility along with inherent flexibility and savings: In the words of the Assistant Secretary of the US Navy, "fuel oil for the Navy gives increased speed and cruising radius, control of smoke-screens, reduces fire-room forces by 55 %, increases the efficiency of refueling at sea by 23 %, gives ability to sustain maximum speed for long periods of time without clogging the furnaces, flexibility in speed and finally greater safety from submarines" (Barron 1917). The problem was that such a step was only available to military forces that enjoyed secure access to homeland oilfields, storage facilities located on the seashore, and a network of bunker stations. This was the case for the United States but not for Britain. As some British politicians realized all too well, dragging the entire Royal Navy into dependence on overseas oil while abandoning the immense supply of top-quality steam coal located in Britain was, at best, a very risky move.

Time proved how bad things could get: Oil dependence not only provoked chronic problems in the supply of British forces during the First World War, bringing the military effort to the brink of collapse by 1917, but, more importantly, it also rendered any British imperial strategy in the Pacific impracticable after the peace settlement. On the other hand, there was little operational utility for the speed factor in the modern naval warfare against Germany. When reviewing the reasoning behind the above-mentioned decision, history has naturally turned toward the internal logic of events and circumstances in British policy and made a more comprehensive assessment of the core vision of idiosyncratic decision-makers—in this case, the First Lord of the Admiralty, Winston Churchill, and the former First Sea Lord, John Fisher (Stoll 1992; Neilson 2000; Babij 2000; Maurer 2003).

The most remarkable consequence was that this event ushered in a new era of a geopolitical clash for strategic energy sources. After the First World War, the disputes over influence zones were no longer in the hands of multinational enterprises but turned into a bitter global competition between nations. Again, it was the British decision to switch navy fueling to oil and the subsequent acquisition by the government of a majority shareholding in a private oil company (the Anglo-Persian concessionary of Iranian oilfields) that disclosed the political relation between international oil sources and national security. This unusual interference in market affairs gave the British state an aura of cold imperialist foresight. And, since the key developed nations lacked this natural resource (contrary to the position with coal), they tried to emulate the British and seize available supplies by political means: Oil became a geostrategic commodity along with and thanks to the technology of steam. Henceforth, economic diplomacy took hold of international relations and tightened the grip over economic resources in Mexico, Russia, Romania, Iran, Iraq-Mesopotamia, the Dutch East Indies, and Venezuela.

4.4 Hybridization and the World Economy

4.4.1 The Role of Oil-Producing Countries

When fuel oil reached the heyday of its development, it still lagged far behind coal in most countries. Although some European countries were aware of the strategic implications of this energy carrier, its overall consumption was very limited. After the First World War, the only significant development that came to the fore was the increasing number of merchant steamers fitted to burn fuel oil, a trend that fostered the conversion of commercial fleets particularly in the United States, Norway, Italy, Great Britain, France, and the Netherlands (Lloyd's Register 1920–1935). Nonetheless, this localized effect was not enough to outweigh the overall dependence on coal, so that the more intensive nature of energy consumption in developed Europe continued to be associated with coal–steam rather than oil–steam technology (Fig. 4.1). In contrast, the substitution effect in the United States was fully extended into the domestic, industrial, and transportation sectors, benefiting from both comparative advantages in relative prices and opportunities from indirect savings. The relative share of railroads and bunker oil for merchant and navy ships progressed steadily, reaching a plateau of 49 % of fuel oil consumption by 1930. Other uses related to burning furnaces saw a steady decline in the applications for mining and manufacturing, manufacture of gas, and electric power generation, somewhat counterbalanced by the increase in demand for residential and home heating, particularly after 1930. At its peak, fuel oil represented about 32 % of the total power sources used for electricity generation and 8 % of those used in city gas (Swanson 1931; Coumbe 1931). Overall, it accounted for 11 % of US inputs in primary commercial energy (see Fig. 4.1).

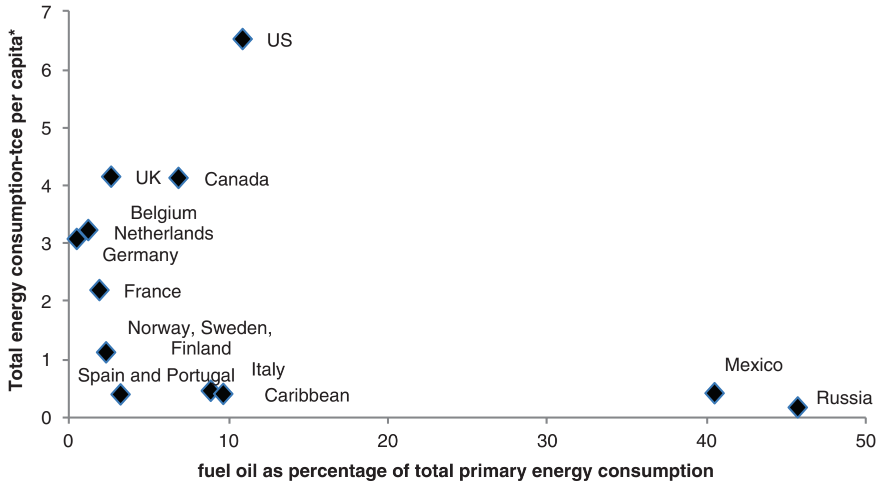


Fig. 4.1 Fuel oil consumption and total commercial primary energy consumption in selected countries, 1928–1929 (Darmstadter 1972; NICB 1930). * tce per capita—tons of coal equivalent per capita

Although it is beyond the scope of this chapter to enumerate the specific traits of the US energy market, from an international perspective, it is nonetheless interesting to note that this pattern of entrenchment of oil in diversified productive activities rippled through the Caribbean and other islands, under the political and military dominance of the United States: By the mid-1920s, Hawaii was absorbing about three million oil barrels per year in its sugar and fruit-canning industries, public utilities, railways, and bunker facilities; Cuba took eight million barrels for the sugar industry, railways, and bunker facilities, and another million went to Puerto Rico's transport and utilities (NICB 1930). Coupled with the strong presence of US entrepreneurs in these countries, the opening of the Panama Canal, in 1914, fostered extensive trade between the east and west coast ports, with large supplies of fuel oil coming to Cuba, Hawaii, and Puerto Rico. As a result, imports of fuel oil largely replaced coal imports in these islands. Further south, Chile was the only non-oil-producing nation in which a large fuel sector evolved, based on the supply of nitrate mines, railways, and ships. As a rule, the American pattern of diversified usages was only to be found in the foremost oil-producing nations: Mexico, Russia, Romania, and the Dutch East Indies.

Hence, it can be concluded that the use of oil in the steam age had a very asymmetrical effect on the world economy: Whereas in most countries it touched only part of the merchant and navy fleets, in oil-producing countries it provided a real alternative to coal, with an ensuing impact on the energy balance. Poor rural economies bearing high transport costs for lignite, anthracite, and coke with labor-intensive industries and very low levels of energy consumption benefited most from the outset. Stepping into the new fuel oil age assured these economies a basis for raising the competitiveness of industry and transport in ways otherwise unattainable.

4.4.2 The 1920s Boom

There was a sharp turnaround in the sociotechnological situation when the hybridization of steam reached its peak in the 1920s. The causes are twofold: On the one hand, there was awareness that oil was an exhaustible resource that could be depleted in about 20 years; on the other hand, there was the diffusion of the comparatively more efficient diesel combustion engine. Saving mineral resources became the priority of the day, and this concern was best expressed by the comparison between the amount of potential energy contained in the fuel and the amount that is actually converted into power. Against this standard, oil-fired steam engines loomed largely as wasteful equipment because they were now contrasted with the performance of the diesel engine.

Rudolph Diesel set for himself the goal of developing a prototype for an ideal engine in which all heat added in could be converted to work on the piston. Its main focus was consequently to avoid all types of loss, while converting energy sources into mechanical power. From a practical perspective, diesel devised the solution of a combustion chamber to which the fuel is directly driven at the time of maximum compression, becoming pulverized or atomized so that each particle of fuel can find a particle of oxygen with which to combine. Compression-induced spontaneous combustion did away with the need for sparking devices and burning furnaces (Cummins 1993; Smil 2007).

Diesel's position in the markets was consolidated with enhanced sales of oil-fired engines. Based on the promise of fuel economy and a range of improvements introduced by German, French, Swiss, and Scandinavian enterprises, the engine finally won a niche within the stationary power market of the 1920s and was readily adopted for the production of locomotives, buses, and trucks. However, public policy was not neutral and tried to sway the markets by means of a very interventionist stance: In the 1920s, the authorities both in the United States and in Soviet Union regarded the oil–steam combine as a technological aberration. This "aberration" needed to be wiped out because it did not compel society to save energy and did not meet "ideal" conservationist practices. The new regulatory body in the United States, the Federal Oil Conservation Board (established in 1924), identified gasoline as the superior application, whereas fuel oil was scaled down as an inferior application, giving priority to inter-fuel substitution (Nordhauser 1979). Moreover, the new technology of thermal cracking allowed the production of gasoline from the fuel oil residuum. In the Soviet Union, oil was increasingly restricted to uses in internal combustion engines in trucks and tractors, serving the needs of massive agricultural industrialization. After the First 5-Year Plan (1928–1932), the demand for boiler and furnace fuel oil declined sharply, and with the Second Plan (1933–1937), all increases in production were absorbed by internal demand, so that Soviet Russia abandoned its position as an exporter of cheap petroleum by-products to the European market (Campbell 1968; Goldman 1980). The picture of the 1920s is thus one of centralized policy attempts to contain the expansion of oil-fired steam engines and a decentralized boom in fuel oil usage, followed by a progressive decline.

4.5 Final Remarks

Between 1880 and 1930, oil became a large international business, an important energy carrier, and a key element in geostrategic world disputes. In terms of by-products, the timeframe is largely dominated by the production of fuel oil, an ideal substitute for coal. Initially, this commodity was simply discarded or "spoiled" in crude and basic industrial activities, but engineers from the leading producing nations soon discovered a way to adapt the fuel to existing steam engines. It was in Russia, where the heavier fractions represented a larger share of the total oil distilled in the stills, that the major breakthroughs took place. After 1901, the conversion of fuel oil into power reached an all-time high in the United States, as new seepages of heavy oil were discovered and production skyrocketed. Unlike Russian oil, which entered a crisis phase of political turmoil and technical exhaustion, the US industry was in its maturing stage.

The innovation and improvement of three types of burner to pulverize oil (steam, compressed air, and mechanical atomization), and to blow it into a furnace in spray form, assured a market for petroleum in the steam age. Through hybridization, it was possible to overcome the "dynamic inertia" that characterizes the maturity phase of many technological systems, with choices of the time dependent on the prior decision and investment paths. Steam engines became more flexible. By the end of the nineteenth century, they could switch between energy carriers according to circumstances, since the reversibility of decisions was assured by the add-on of burners at a low fixed investment. Yet, despite the relative simplicity of technological solutions, the hybridization of steam had a significant macroeconomic impact, with fuel oil consumption accounting for between 11 % (USA) and 46 % (Russia) of total commercial primary energy in 1928–1929. However, this was not a worldwide phenomenon. The economies of hybridization were above all a local opportunity, proving most effective where regional access to alternative coal supplies was difficult. It was therefore in oil-producing countries with little coal (such as Mexico, Romania, Venezuela, and the Dutch East Indies), when coal supply was disrupted, or at times of underinvestment or high costs (such as in the US and Russia) that combined oil–steam reached its peak. Conversely, jumping on the bandwagon implied large risks for outsiders. Great Britain is the best example of a non-producing country that tried to hybridize its fleet for military reasons. Suffice it to say that this decision triggered a chain of events that made oil the cause of world disputes.

While the economies of hybridization are mostly locally based, this stems from the need to deploy an infrastructural network, built from scratch, to ensure regular provisions for the non-dominant hybrid technology. Building such infrastructures entails the classic problem of the propensity for the undersupply of public goods. Fuel oil bunker stations for twentieth-century steamships were like recharging stations for twenty-first-century electric vehicles and from this point of view constitute a bottleneck that might cripple or, at least, retard, the diffusion of hybrids. Should this reasoning hold, with the current competitive price of hybrid-electric vehicles and the great development potential for battery-powered electric engines, one may expect burgeoning local economies of hybridization in towns that make significant strides toward the creation of a network of battery recharging stations: the hinterland of Amsterdam, Rotterdam, and five other Dutch cities; the US state of Hawaii along with the cities of Portland, Los Angeles, and Texas; Kanagawa municipality in Japan; and in China's three largest cities.

References

Archival Sources

British Petroleum Archives (BPA), Warwick
BPA Doc. 136390 (1916) Historical Notes, articles, Kerosene price policy. The Burmah Oil Company, Feb 1916

Bibliography

Anonymous (1924) Growth of world's bunkering ports. Oil Gas J 12
Anonymous (1927) Russians likely to exceed record. Oil Gas J 31
Babij O (2000) The Royal Navy and the interwar plans for war against Japan: the problem of oil supply. In: Kennedy G (ed) The merchant marine in international affairs, 1850–1950. Frank Cass Publishers, London, pp 84–107
Barron CW (1917) Oil expansion of peace to outstrip war demands. Wall Str J 3
Bey E (1931) Blood and oil in the Orient. Nash & Grayson Limited, London
Campbell RW (1968) The economics of Soviet oil and gas. John Hopkins University, Baltimore
Chandler AD Jr (1994) Scale and scope: the dynamics of industrial capitalism. Harvard University Press, Cambridge
Chau KT, Wong YS (2001) Hybridization of energy sources in electric vehicles. Energy Convers Manag 42(9):1059–1069
Cleveland W (1994) A history of the modern middle East. Westview, Boulder
Coumbe AT, Hedfield AR, Swanson EB (1931) National survey of fuel oil distribution 1930. American Petroleum Institute, Washington
Cummins CL (1993) Diesel's engine. Carnot Press, Wilsonville
Darmstadter J (1972) Energy in the world economy: a statistical review of trends in output, trade, and consumption since 1925. Johns Hopkins University Press, Baltimore
Dennis MA (1985) Drilling for dollars: the making of US petroleum reserves estimates, 1921–25. Soc Stud Sci 15(2):241–265
Donkin B (1894) A text-book on gas, oil and air engines. Charles Griffin and Company, London
Dunn FB (1916) Industrial uses of fuel oil. Technical Publishing Company, San Francisco
Elliot IF (1974) The soviet energy balance: natural gas, other fossil fuels, and alternative power sources. Praeger Publishers, New York
Enos JE (1962) Petroleum progress and profits: a history of process innovation. MIT Press, Cambridge
Etemad B, Luciani J (1991) World energy production, 1800–1985. Droz-Université de Genéve, Genéve
Evans DC (2001) Japanese naval construction, 1878–1918. In: O'Brien P (ed) Technology and naval combat in the twentieth century and beyond. Frank Cass Publishers, London, pp 22–35
Ferrier RW (1982) The history of the British Petroleum Company: the developing years 1901–1932. Cambridge University Press, Cambridge

Fouquet R, Pearson PJ (2006) Seven centuries of energy services: the price and use of light in the United Kingdom (1300–2000). Energ J 27(1):191–216
Fouquet R (2008) Heat, power and light. Revolutions in energy services. Edward Elgar, Cheltenham
Fursenko AA (1990) The battle for oil The economics and politics of international corporate conflict over petroleum 1860–1930. JAI Press, Greenwich (Connecticu, trans.Gregory L.Freeze)
Gerretson FC (1955) History of the royal Dutch shell. E. J Brill, Leiden
Gerschenkron A, Nimitz N (1952) A dollar index of Soviet petroleum output, working paper, Project Rand Research Memorandum, US Air Force
Gibb GS, Knowlton EH (1956) History of standard oil company: the resurgent years, 1911–1927. Harper & Brothers, New York
Goldman MI (1980) The enigma of Soviet petroleum. George Allen & Unwin, London
Halpern P (2001) The French navy, 1880–1914. In: O'Brien P (ed) Technology and naval combat in the twentieth century and beyond. Frank Cass Publishers, London, pp 36–51
Hamilton JE (1933) A short history of the naval use of fuel oil. J Am Soc Naval Eng 45(3):270–292
Hardy C (1931) Oil ships and sea transport. George Routledge and Son, London
Hassmann H (1953) Oil in the Soviet Union. Princeton University Press, Princeton
Hawley EW (1968) Secretary Hoover and the bituminous coal problem, 1921–1928. Bus Hist Rev 42(3):247–270
Henriques R (1960) Bearsted: a bibliography of Marcus Samuel. Viking Press, New York
Hewins R (1958) Mr five per cent. The story of Calouste Gulbenkian. Rinehart and Company, New York
Kewley J (1922) The petroleum and allied industries. Ballie`re, Tindall and Cox, London
Lambert NA (1995) British naval policy, 1913–1914: financial limitation and strategic revolution. J Mod Hist 67(3):595–626
Lautenschläger K (1983) Technology and the evolution of naval warfare. Int Secur 8(2):3–51
Leeuw C (2000) Oil and gas in the Caucasus & Caspian: a history. St Martin Press, New York
Lloyd's Register of Shipping (1920) Lloyd's, London, 1920–1935
Lyon H (1977) The relations between the Admiralty and private industry in the development of warships. In: Ranft B (ed) Technical change and British naval policy, 1860–1939. Hodder and Stoughton, London, pp 37–64
Maurer JH (ed) (2003) Churchill and strategic dilemmas before the world wars. Frank Cass Publishers, London
McAuliffe E (1927) Railway fuel: the coal problem in its relation to the transportation and use of coal and coal substitutes by steam railroads. Simmons Boardman, New York
Melville GW (1904) Report of the U.S. Naval liquid fuel board. Government Printing Office, Washington
Mitchell BR (1992) International historical statistics: Europe 1750–1988. Macmillan, London
Mohammed SS, Williamson JG (2004) Freight rates and productivity gains in British tramp shipping 1869–1950. Explor Econ Hist 41(2):172–203
Neilson K (2000) Reinforcements and supplies from overseas: British strategic sealift in the First World War. In: Kennedy G (ed) The merchant marine in international affairs, 1850–1950. Frank Cass Publishers, London, pp 31–57
NICB (1930) Oil conservation and fuel oil supply, National Industrial Conference Board, New York
Nordhauser NE (1979) The quest for stability: domestic oil regulation, 1917–1935. Garland Publishing, New York
Nordhaus WD (1998) Do real-output and real–wage measures capture reality? The history of lighting suggests not. In: Bresnahan T, Robert JG (eds) The economics of new goods. University of Chicago Press, Chicago, pp 29–66
Owen EW (1975) Trek of the oil finders: a history of exploration for petroleum. American Association of Petroleum Geologists, Tulsa

Ratcliffe M (1985) Liquid gold ships: a history of the tanker, 1859–1984. Lloyd's of London Press, London

Schurr SH, Netschert BC (1975) Energy in the American economy, 1850–1975. Greenwood Press, Westport

Schurr SH (1984) Productive efficiency and energy use: an historical perspective. Ann Oper Res 2(1):229–238

Shulman PA (2003) Science can never demobilize: the United States navy and petroleum geology, 1898–1924. Hist Technol 19(4):365–385

Smil V (2007) The two prime movers of globalization: history and impact of diesel engines and gas turbines. J Glob Hist 2(3):373–394

Snyder DA (2001) Petroleum and power: naval fuel technology and the Anglo-American struggle for core hegemony, 1889–1922. PhD thesis, Texas A&M University

Stoll RJ (1992) Steaming in the dark? Rules, rivals and the British navy, 1860–1913. J Conflict Resolut 36(2):262–283

Sullivan BR (2001) Italian warship construction and maritime strategy, 1873–1915. In: O'Brien P (ed) Technology and naval combat in the twentieth century and beyond. Frank Cass Publishers, London, pp 3–21

Swanson EB (1931) National survey of fuel oil distribution 1926. American Petroleum Institute, Washington

Tolf RW (1976) The Russian rockefellers: saga of the nobel family and the Russian oil industry. Hoover Institution Press, Stanford

Ville SP (1990) Transport and the development of the European economy, 1750–1918. St Martin's Press, New York

White D (1920) Petroleum resources of the world. Oil Gas J 5–6

Williamson HF, Daum AR (1959) The American petroleum industry. The age of illumination 1859–1899, vol 1. Northwestern University Press, Evanston

Yergin D (1991) The prize: the epic quest for oil, money, and power. Touchstone Books, New York

Zapata C, Nieuwenhuis P (2010) Exploring innovation in the automotive industry: new technologies for cleaner cars. J Clean Prod 18(1):14–20

Chapter 5
Last Gasp

Abstract This chapter examines the evidence on the vigorous and imaginative responses from older industries whenever their profit margins were threatened by competition from replacement technologies. The concept used to account for such processes in which threatened technologies display an extraordinary growth in efficiency just before they enter into abrupt and terminal decline is "the last gasp." The overview of the switch from wood–charcoal to coal–coke technologies in the sector preceding the large-scale iron and steel industry here provides a way to test the "last gasp" model. This reveals how, within the time frame for "transition," one might expect either the effective substitution of the older energy carrier or the incentive to its expansion. The same is to say that the course of action that leads to destruction through innovation and decreases in cost also unleashes creative market mechanisms.

5.1 The Resilience of Old Technologies

5.1.1 The Sailing Ships Effect

Some technologies have been observed showing surprising resilience when threatened by the arrival of more modern and cutting-edge technical innovations. Throughout a succession of improvements, growth in productivity, and supply-side adjustments, they fight back to hold onto their markets. Consequently, instead of a linear process of decreasing returns, competence–destruction, disinvestment, and inexorable substitution, what takes place is renewed and heightened competitiveness of supposedly outmoded technical capabilities. This happens partially because new inventions often create, rather than destroy, market opportunities for older technologies. According to the standard "last gasp" view, it is mostly during the initial stages of new technologies, characterized by slow performance improvements, that further incentives and opportunities are created

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_5,

for incumbent industries to survive. A final point, moreover, underscores the fact that the response from these incumbent industries is generally short lived and unable to avoid ending up in abrupt and irreversible decline. The "last gasp" concept pointedly captures the pattern of resolute rise and appalling fall of technologies for which technical innovations provide close substitutes.

Historically, this form of technological resilience was first identified in the replacement of sailing ships by steamships. Taking as his starting point the observation that the best sailing ships for ocean transport were produced in the second half of the nineteenth century when already displaced by steamships, the sociologist Seabury Gilfillan (1935) identified the importance of an array of innovations that actually served to maintain the competitiveness of the old technology by enabling the construction of faster, larger, more flexible, and more autonomous vessels. Furthermore, a significant part of the enhanced efficiency of sailings ships was "borrowed" from steamship-based breakthroughs, in particular, light composite hulls of iron encased in wood and, later, iron and steel structures, steam-driven winches, auxiliary steam engines, and the use of steam tugs for harbor towing. Recent econometric research on the total factor productivity of tramp ships found, for instance, that the decrease in port turnaround times on some major routes, largely brought about by the usage of specialized steam tugs and the better organization of port facilities and cargo handling, contributed almost as much to productivity growth as the direct effect of the increase in ship sizes (Mohammed and Williamson 2004). This conclusion underpins the complimentary role of new technologies in fostering efficiency gains in their older counterparts.

After Gilfillan's work, most findings related to resilient industries were subsumed under the concept of "sailing ships effect." For over forty years, the label remained stuck in place. However, several authors then noticed that the "sailing ships effect" was not exclusive to merchant trading activities but could also occur in economic sectors as diverse as typesetters, ice harvesting, alkali production, locomotive engines, vacuum tubes, or carburetors. To account for the mounting diversity, the term "last gasp" was introduced, reflecting the broad scope of situations whereby an old technology gets improved following the inroads of a new technology. However, since most theoretical generalizations evolved out of the accumulation of historical cases, in a typical "bottom-up" process of inductive reasoning, the relevance and applicability of the concept has yet to be determined.

The aim of this chapter is to single out the bearings of the "last gasp" in the field of energy. To this end, we examine the transition from the old technology of wood fueling to the new technology of the industrial revolution, coal fueling, in the sector forerunning the large-scale iron and steel industry. Besides its major impacts on expanding coal markets across Europe and North America, the iron and steel manufacturers supplied the ultimate structural material of the industrial revolution and experienced a technological change that was independent of, even though interconnected with, the technological improvements in steam engines (by the dawn of the nineteenth century, steam engines came to substitute water wheels for running blowing devices and blast engines). In other words, steam engines were a self-sufficient driver of nineteenth-century energy transition toward fossil fuels and the most important quantitative factor in Britain (Report of the Commissioners 1871). Should the last gasp concept prove right, the switch to coal-fueling technology should not immediately wipe out continuity in the traditional supply of wood: At least a significant proportion of the ironworks existing should still be able to continue to draw their resources from forests.

5.1.2 Creation and Destruction

Much of the sailing-ship/last gasp effect ran counter to the idea of a capitalist system driven by brutal waves of creation and destruction, within which radical technological innovations swamp traditional productive systems. The Austrian-American economist Joseph Schumpeter was chiefly responsible for this view in which progress does not come without costs. Writing in the inter-war period, Schumpeter noticed how entrepreneurship and technological change create dynamic imbalances in the course of development, from which results the creative destruction of the old economy and not its mere replacement. Inasmuch as technical change was broadly defined as a cluster of quality improvements to any given array of products, reflecting a "vertical innovation" process, its repercussions could only but be devastating. Aware of the role ascribed to multifunctional and multinational enterprises in modern capitalism, Schumpeter extolled the new type of competition that was then looming everywhere:

> the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest-scale unit of control for instance)—competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives (Schumpeter 1975: 85).

To an ever increasing extent, the classical market mechanisms of price efficiency and comparative advantage lost ground in favor of more modern forms of competition. In this brave new world, there was no alternative for traditional producers other than their immediate obsolescence and extinction. Industrial mutation triggered a selective evolution process in which any uncompetitive economic actors were doomed to failure (Grossman and Helpman 1991). Of furthermore significance is how the cases the Austrian economist made reference to as glaring examples of creative destruction were all interrelated with the history of coal-steam technologies, specifically the disappearance of iron and steel charcoal, of water-wheel motive power, mail coaches, and horse-drawn carriages (Schumpeter 1975: 84).

It was this single-sided relationship between innovation and destructive effects that the "last gasp" literature called into question. Aside from the obvious damages, there might also be certain benefits that industries supposedly due for replacement might draw succor from. Reflecting on the unforeseen consequences of technological innovation upon their traditional competitors, contemporary research has singled out two positive causal mechanisms:

- The first runs mainly through the transformation of the supply side of the market. According to authors like Snow and Tripsas, components from entrant technologies might "spill over" into incumbent technologies and thereby improve the incumbent. Another plausible outcome stems from markets split between entrant and incumbent spurring the application of each technology to those applications to which they were actually best suited and thus boosting efficiency by dint of specialization (Snow 2003; Snow and Adner 2010; Tripsas 1997, 2001). Ultimately, these complex and mutually reinforcing dynamics between old and new industries would prompt clear-cut market segmentation, technological imitation, and the transference of skills.
- The second causal mechanism runs throughout the transformation of the demand side, that is, through the expansion of the actual market. Alwyn Young envisioned a formal model for situations in which production is undertaken in cycles, stage by stage, such as in the iron and steel industries. In his model, the traditional producers are beset by a shift in producer demand toward new inputs and away from their old options. However, this tendency is largely offset by the gains accrued with new applications coming to the fore for the already existing inputs (Young 1993). Innovation is viewed as pushing through an overhaul in demand from which both traditional producers and ground-breaking entrepreneurs have much to benefit. This correspondingly means that the Schumpeterian destructive effect is matched by a creative or complementarity effect that extends to the market for "obsolete" technologies ("new applications for existing inputs") and thus deferring their "last gasp." However hard the competition, however hard the cutbacks in costs, further incentives for survival still remain.

The ensuing pages sketch out the battle between these destructive and creative forces throughout the transition of the nineteenth century, testing the adequacy of the aforementioned theoretical models to the larger iron-producing nations: Sweden, Britain, Belgium, France, Germany, the United States, Russia, and Canada.

5.2 Charcoal and Coke: An Overview

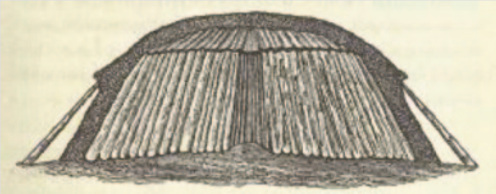
5.2.1 Objective and Subjective Assessments

The claim that coal represents a "higher-quality energy form" than wood is grounded on the capacity of solid fossil fuels to perform more useful work and on their larger contribution to the productivity of non-energy inputs. Basically, these two traits meant that coal has more chemical energy per unit of weight and permits greater efficiencies in storage, transportation, maintenance and repair, economies of scale, and ease of handling and usage. Whereas the first aspect stems from the physical properties of the raw material, all the others have involved some kind of purposeful human action throughout history. In truth, the very idea of civilizational progression from simple energy carriers toward more sophisticated ones, up what was termed the "energy ladder" (Hosier and Dowd 1987), tends to reflect a long history of successful achievements in entrepreneurship and science. It is also useful to keep in mind that before the turnaround brought about by the British industrial revolution, historical actors were utterly convinced that wood rather than coal was the superior fuel and especially so for feeding furnaces and forges.

This industrial wood was, in effect, charcoal, a secondary energy source. Charcoal was obtained from air-drying green wood for a period lasting from one to four months after which the dry wood is moved into a traditional hearth or pitsead in the form of dome-shaped stacks, generally located in the woods, where it is burned slowly and without oxygen for one to six weeks. When properly executed, wood drying removes as much as 15–30 % of the wood's original water with the remaining moisture being further reduced by means of the pyrolysis process (Svedelius 1875: 6–26, 200–201). In the end, the chemical transformation underwent by wood yielded a carbon-rich material, physically very similar to coal and with a similar calorific power. This means that despite the simplicity and low cost of this rural industrial transformation, pyrolysis in kilns returned a twofold increase in the thermal content of dry wood (from 14.4–17.4 to 29.5 MJ/kg), raising the potential heat per unit of weight. However, since charcoal was an extremely bulky fuel whereas coke is much denser, the former required two and a half times the storage space normally reserved for fossil fuels (175 cubic feet by metric ton as against 69 cubic feet occupied by coke) and about one and a half times the furnace cubic space, consequently a loss in the energy contained per unit volume of space (Space occupied by fuel 1882: 160–161; Bell 1884: 133). Moreover, the friability of charcoal capped the potential height of furnaces.

Equality of energy by weight, imbalance of energy by volume, and the relative fragility of the billets made the vegetable fuel more difficult to handle, to transport and to process. However, this apparent inferiority was largely offset by the smoothness of operations and the good results in terms of the average final product quality. Charcoal furnaces, bloomeries, and forges could easily control undesired chemical reactions, the contamination of heated metal with impurities from the fuel, and the volume of gases. These were precisely the kind of problems that appeared technically insurmountable whenever coal was experimentally used. For these reasons, European ironworkers of the modern period tended to view wood's "superiority" in terms of charcoal's chemical purity, namely its low sulfur and phosphorus content and the high porosity of the carbon-rich material. Defined in this way, the iron quality became an attribute of craftsmanship as the better the raw materials, the less the workforce effort required in processing it (for instance, high-quality charcoal produced less slag and demanded less hammering). Ultimately, this vision lent itself to moral analogies. As a nineteenth-century iron puddler put it: "Man's nature is like iron, never born in a pure state but always mixed with elements that weaken it" (Davis 1922: 31).

The process of making charcoal



A *meilier* is, in fine, a lot of wood piled up or raised up according to certain rules on a so-called hearth, a place which has been cleared and levelled and is either flat or gently sloping. The heap of wood is thatched with charcoal dust, if this is at hand, but if not, with dirt, sawdust, and the like, in order to prevent in this manner the admittance of air, for otherwise the whole would, when kindled, burn to ashes. When the *meilier* has been thatched it is ignited at some certain point, which varies in situation according to the construction of the *meilier*. When the fire is well established, the place of igniting is carefully covered, but if the thatching is everywhere perfectly tight the fire will soon be extinguished, while it will be nourished more or less lively by a more or less liberal access of fresh air. Sometimes the covering is so thin that the necessary air can pass through, but more commonly, especially in standing *meiliers*, it is necessary to make small openings in the covering for the air to pass through. Then the draught is so regulated that the oxygen entering may be just sufficient to burn what wood and charcoal is needed, in order that the coaling of the wood may progress gradually from the exterior to the interior; and in order that the development of gases may be kept steady by a well-regulated temperature.

(Svedelius 1875:26)

5.2.2 Fuel Efficiency in Ironworks

In terms of drawbacks, the intensive use of woodland resources constrained the scale of production, where not the sustainability, of charcoal manufacturers largely because it took at least 4–6 tons of wood to produce 1 ton of charcoal. Measured in energy units, this implies that 60–90 MJ/kg of primary energy were needed to obtain a secondary raw material which delivered 29.5 MJ/kg on entry into the furnaces or the forges. Under the pressure of wood shortages and compelling scientific approaches to and results from the pyrolysis process, European practices began to display some signs of convergence around best practice conversion ratios of 4 tons of wood to 1 ton of charcoal during the first decades of the nineteenth century (Svedelius 1875; Hammersley 1973; Benoit 1990; Blanchard, 2005). Further improvements in energy saving in this preliminary productive stage afterward came to a standstill, revealing that the meiler and kiln technology had possibly reached its technological limits. The adoption of a new system of wood distillation in apparatuses with retorts in the second half of the nineteenth century gave another new push to economic productivity and to the profitability of wood carbonization. With the diffusion of this technology, the conversion ratio dropped to 3.3 tons of wood for 1 ton of charcoal, while by-products such as tar, wood alcohol, and turpentine began to be recovered and sold in markets (Forsythe 1913: 82–85). Throughout this evolution, the most significant strides toward wood saving practices were attained by producers in the Urals, in Russia, whose brick–earth system equipped with iron retorts was able to deliver yields of 2.7:1 in birch–charcoal output (Blanchard 2005: 134–136).

Fashioned by the well-to-do who could afford the smokeless heat of charcoal, domestic consumption of this vegetable fuel had shot up in European capitals by the early nineteenth century (Boissiere 1990; Henriques 2009: 40–43). Aside from this domestic consumption located within major cities, charcoal fed the of malt beer and iron production sectors in modern Europe, in the USA and in Canada. Within the iron industry, fuel consumption weighted approximately the same in the first cycle of the "trade," the making of pig iron from ores in blast furnaces, and in the second phase, the refining of the metal and its shaping into bars. This was achieved by reheating the pig iron in a finery so as to reduce the carbon content in the molten iron from 4 to around 0.05 % and finally by forging and consolidating the throughput in a chafery (Tylecote 1991: 233–240). Once the iron was rendered malleable, it was accordingly sold to blacksmiths who transformed the bars into an array of final products—horseshoes, locks, nails, hinges, knives, scythes, and other agricultural implements. Here, charcoal was of lesser importance as the fabrication of hardware and tools could be performed with the aid of pit coal. In any case, each step in the production cycle added further fuel costs to the material produced. To single out the thermal efficiency attained in each respective productive phase of iron production, the ensuing pages present most data in terms of secondary energy consumption, that is, in terms of the calorific value of charcoal, regardless of the progress attained in the primary conversion of wood into charcoal. This methodology ensures a basis for comparison between nations and between periods of time.

5.2.3 British Leadership in Coke Fueling

Sweden with its bountiful forest areas and rich deposits of non-phosphoric ores emerged as the leading exporting nation in the seventeenth century, carving an unbeatable reputation for its combination of iron bar resistance, malleability, and chemical purity. Ranked above any possible European competitor, Swedish oregrounds iron and Swedish iron set a quality standard difficult to attain, seizing premium prices for forging purposes and for steelmaking. So overwhelming was its reputation that the many technical experiments pursued at length by British ironmasters in the domain of coal-heating methods were not intended to directly challenge Swedish competition, nor even Russian imports, which grew in importance from the 1760s onward. Rather, their best endeavors were directed toward making cheap iron for the supply of low-quality goods, in the interim securing the benefits of lower fuel costs and an abundant fuel supply (King 2005; Evans et al. 2002).

Recurrently attempted, the adaptation of mineral fuel to iron smelting was achieved in the context of substituting cast iron for the traditional brass and copper in the manufacture of small commodities, requiring the heating techniques deep-rooted in non-ferrous metallurgy be transposed to this metal (King 2011). Abraham Darby, a former iron and brass manufacturer, who had already taken out a patent for an improved method of casting pots, succeeded where many others had failed, thanks to the use of sand castings to make small cast iron items. Coke pig iron was probably used for the first time to make cannon balls at Coalbrookdale in the West Midlands region of England, near the Severn River, in one of the years in the 1690s. It was certainly used to make pots and other cast iron goods as from 1709 (King 2011). These pots and cannon balls were particularly brittle breaking when struck hard and weak in extension. However, cannon balls and pots were precisely the kind of goods that brushed aside the attributes of malleability and strength characteristic of charcoal wrought iron. In the wake of Abraham Darby's discovery, in 1709, coke smelting remained a minority pursuit. By 1750, only four furnaces out of the sixty-eight existing had adopted coal for iron smelting. Such a stalemate, extending over forty years, has been a puzzling problem to British economic history. While the classical explanation hinged upon the secrecy of the innovation, historians have recently placed much greater emphasis on the adversity of market conditions, specifically the growing competition from imports; the reduced efficiency of pioneering coke-ironworks; the low level of demand for cast iron and its limited usage; and the technological constraints of using too little blast, as well as the price disadvantage of coke-smelted iron (Ashton 1924: 24–59; Riden 1977; Hyde 1977; King 2005).

In conjunction, these factors hampered the diffusion of coal technology. Only in the second half of the century did some of the aforementioned factors (productivity, technology demand, and costs) begin to change heightening the incentives for a new wave of innovations. This period was marked by increasing pressures on charcoal ironwork profit margins, pinched between mounting costs and greater competition from imported iron staples (Hyde 1977: 34–35; Evans 2005: 19–21). Consequently, up to the 1770s, aggregate British output of charcoal pig iron declined by an amount that was exactly offset by the rise in the coke sector, with furnaces often only one being used one year in two or two in three or even one in three. In order to ensure their survival, traditional sectors responded by cutting back on fuel consumption, which accounted for the largest share of their overall cost structure (Hammersley 1973: 610), fighting hard to hold onto their markets and so counteract the new competition.

Following the erection of new coke furnaces in Shropshire, West Midlands, coke pig iron became a normal feedstock for finery forges. Coal usage henceforth became a self-reinforcing process that is grounded in direct smelting for the production of cast iron (augmenting the production of final goods) but also in feeding the fining sector of the forges with coke–pig iron and seizing at least one-fifth of the demand from charcoal fineries (Hyde 1974: 199). These developments come to characterize the second phase in coal diffusion and rested on the discovery of a new technique to remove the carbon from pig iron and convert it into malleable bar iron (decarburization), while overcoming the drawbacks from the relatively high share of silicon in coke-smelted pig iron plus the contamination of the heated metal by another undesirable impurity, the sulfur contained in the coke. Both the stamping and potting process, invented in the early 1760s by the Wood brothers and Henry's Cort's puddling and rolling technology patented in 1783–1784, drew upon the same solution: avoiding direct contact between the coke and the molten pig, thereby halting the transmission of chemical impurities. The widespread diffusion that ensued, particularly of the puddling and rolling technologies, led to the closure of most traditional charcoal forges, mostly by 1815.

The changeover to coal brought about an increase in the amount of energy per unit of output when compared to the traditional methods of charcoal fueling. By the end of the eighteenth century, the overall consumption of coke-ironworks amounted to 225–255 MJ/ton (8–9 tons of coke) per ton of bar iron produced. This figure comprised the fuel supplied at the furnaces and the subsequent feeding of the stamping and potting process or, alternatively, of the puddling process (Needham 1831, 27–28; Truran et al. 1865: 80–96, 205–240; de Beer et al. 1998; Hyde 1977). On the other hand, the corresponding charcoal yield was 127–163 MJ/ton (4.3–5.5 tons of charcoal) per ton of bar iron, which means that traditional ironmasters almost halved their own secondary fuel consumption throughout the eighteenth century (Hyde 1977; Hammersley 1973). Nevertheless, the accomplishments in charcoal fueling were not sufficient to reverse the tide or even to countervail the lower coke-iron prices. Conversion to coal proceeded apace spearheading the drive in society's demand for more energy in the early years of industrialization. Per capita secondary energy consumption grew relatively steadfastly because there was a sudden increase in the secondary energy intensity of final goods (MJ per Pound of bar iron or wrought iron); an upward shift in the demand of goods with high energy intensity (mounting consumption of cheaper iron in construction, naval ironware, and agricultural implements); and also a spurt in new manufactures that demanded higher-energy intensity equipment goods (iron plates, cast iron, larger steam engines for blast, and special machinery such as grooved rolls for puddling or steam-powered tilt hammers). All in all, it took about a hundred years for the English coke-iron industry to reach the level of heat efficiency at the entry of furnace and forge entrance, attained by the old charcoal industry, a threshold in any case only achieved by the most efficient plants (Allen 1979).

It would seem to follow from these considerations that the history of coal adoption by the British iron industry unveils a course of action in which the creation of innovative opportunities thrives alongside the destruction of the old business and substitution mechanisms rule out complementary mechanisms. In reality, destruction prevailed over creativity. The course of events came to be entirely dominated by price-switching effects in which not only did consumers of final goods shift their demand to coal–iron products but also producers of traditional final goods shifted over to the new intermediate inputs (charcoal forges began working on coke produced pig iron). Further invention stepped up these switching effects, creating a path-dependent trajectory in the substitution of the older technology. The striking point, however, is the absence of any countervailing mechanisms that might have led to an expansion in the charcoal-based iron goods market. Since overall market growth was insufficient to create new applications for older inputs, charcoal-fueled ironworks witnessed the substitution of their throughput as they were boxed into ever-narrower market niches and basically ousted beyond the production of high-quality metal for wire manufacturing (Hayman 2008). Decline proved irreversible and the fight back inconsequential. By the dawn of the nineteenth century, wood substitution had been fully accomplished with charcoal fueling representing less than 10 % of iron production.

Underpinned by its technical efficiency and lower raw material and transport prices, British iron industry entered upon its great era as the major supplier of iron and steel to the world market. For over seventy years (1800–1870), its cost–quality performance outflanked top-quality producers, such as Sweden and Russia, but also emerging industrial powers, such as Germany and the United States (Allen 1979).

5.3 The Fight Back from Traditional Industries

5.3.1 The First Fight Back

It was only after the end of the Napoleonic Wars, when trade resumed its course, that continental iron producers noticed how the competitive gap had in the meantime widened. Charcoal-based manufacturing hubs located near the ocean were the first to feel the destructive power of this new competition. Swedish and Russian charcoal iron exports were also hard hit. As events played out, there proved to be only one region in the world capable of sustaining the speed of innovation necessary to catch up with British technology. That region was the Walloon hinterland of Liège and Charleroi, in Belgium.

Drawing on a long tradition and renowned expertise in the processing of malleable iron for armaments and nails, Belgium entrepreneurs and British immigrants like William Cockerill erected the first puddling furnaces and coke blast furnaces there, in 1821–1823 and 1827–1829, respectively. Almost immediately, other coal-fired plants mushroomed across the region with ravaging consequences upon the traditional charcoal iron industry, which was plunged into irrelevance and basically restricted to a handful of charcoal hearths. By as early as 1835, the throughput of new technology had already surpassed the old (Fremdling 2005: 49–50). With this short time span, there also came a compression and intensification of the destructive effects driven by metallurgical innovation. So swift and successful was this changeover that Belgium stood out in the years ahead as the pivotal nation in spreading coal technology throughout Europe (Pollard 1981: 87–94). Some factors have been pointed out to explain this singular path which closely resembles the theoretical model of traditional industry instantaneous obsolescence developed by Schumpeter: To begin with, the Walloon area enjoyed an array of fortunate geographical conditions, with ore and coal in close proximity, low transportation costs, and an interface position between the British Isles and the continent. It is important to add that nowhere else did these common British-Belgium factors stand out as clearly. A second advantage lay in the significant transference of British capital and British expertise involving not only managerial and engineering skills but also "ground-floor" worker skills (Pollard 1981). Less commonly mentioned is the fact that Belgium had already depleted what represented one of the smallest forested areas of Europe, therefore hampering the development of its traditional charcoal industry (see Table 5.1). Given the opportunity costs for change, industrial restructuring quickly prompted an overall raw material substitution reaping in the meantime the benefits of "fast-second" innovation.

Unlike this near Schumpeterian pattern, France and Germany experienced a much more protracted process. The technological transference was intense in the final phase of pig iron fining with coal but less so in the current charcoal-based methods for ore smelting. David Landes was the first author to identify this development stating that it was as if continental Europe had reversed the "natural" order of British innovation because the adoption of puddling and rolling technology entailed lower capital costs, lower technical difficulties, and uncomplicated learning (Landes 1969: 175–176). Thus, the French and German entrepreneurial attitudes tend to reflect a risk-adverse adaptation sprinkled across those industrial structures that faced harder geographical conditions and a more backward level of technological development.

Recent research has, however, shown that this was not just a faulty emulation but an original pattern of specialization that endured for over forty years and successfully withstood competition. Dubbed the "champagne model," after its place of origin in northeastern France, the continental pattern efficaciously combined charcoal blast furnaces with the *méthode à l'Anglaise* of puddling and rolling (Fremdling 1991, 2005). The ensuing hybrid economy spluttered and struggled to take root and survive not only throughout central and northeastern France, but also in western Germany along the banks of the River Rhine, in the Ruhr Valley, Westphalia, and Saar in addition to the eastern band of Upper Silesia.

Figure 5.1 depicts an estimate of the energy transition dynamics. The wider gap between the lines represents charcoal usage in blast furnaces and in forges, in France, as compared to Germany, indicating that the split in the iron industry in a dual-sector economy was more sweeping than in the former nation. Additionally, it is clear that the "fight back" by traditional sectors in the terms defined by Snow-Tripsas hinged upon specialization in ore smelting and lasted only up to the 1850 s, dwindling afterward to residual market shares. In technological terms, this "fight back" entailed a new cycle of investment that hastened the pace of charcoal blast furnace along the tracks already opened up by their British and Belgian coke competitors. Particularly important was the swift adoption of Scottish hot-blast technology amid the charcoal milieu, which involved a redesign of furnace

Table 5.1 The impact of ironwork fuel consumption on forests. Selected countries, 1820; 1840; 1860; 1880 (Madureira 2012: Annex 1)

Belgium	Forest used to fuel the charcoal iron industry (thousand ha)	Forest equivalent to coal consumption in the iron industry (thousand ha)	Forest used to fuel the charcoal iron industry/total forest area in 1880 (%)	Forest equivalent to coal consumption/total forest area in 1880 (%)
1820–1822	253.8	0	52	0
1840	182.5	441.8	37	90
1860	106.5	1642.3	21	336
1880–1882	0	3824.4	0	782
France	**Forest used to fuel the charcoal iron industry (thousand ha)**	**Forest equivalent to coal consumption in the iron industry (thousand ha)**	**Forest used to fuel the charcoal iron industry/total forest area in 1882 (%)**	**Forest equivalent to coal consumption/total forest area in 1882**
1820–1822	848.0	97.8	9	1
1840	1272.2	892.7	14	10
1860	1120.9	3378.8	12	36
1880–1882	319.6	8189.0	3	88
Sweden	**Forest used to fuel the charcoal iron industry (thousand ha)**	**Forest equivalent to coal consumption in the iron industry (thousand ha)**	**Forest used to fuel the charcoal iron industry/total forest area in 1890 (%)**	**Forest equivalent to coal consumption/total forest area in 1890**
1820–1822	1942.3	0	11	0
1840	2541.8	0	14	0
1860	2764.7	0	15	0
1880	4144.7	0	23	0
USA	**Forest used to fuel the charcoal iron industry (thousand ha)**	**Forest equivalent to coal consumption in the iron industry (thousand ha)**	**Forest used to fuel the charcoal iron industry/total forest area in 1908 (%)**	**Forest equivalent to coal consumption/total forest area in 1908 (%)**
1820–1822	125.8	42.3	0.1	0.0
1840	2067.5	315.8	0.9	0.1

(continued)

Table 5.1 (continued)

USA	Forest used to fuel the charcoal iron industry (thousand ha)	Forest equivalent to coal consumption in the iron industry (thousand ha)	Forest used to fuel the charcoal iron industry/total forest area in 1908 (%)	Forest equivalent to coal consumption/total forest area in 1908 (%)
1860	1125.5	3862.2	0.5	1.8
1880–1882	1902.7	15961.0	0.9	7.2

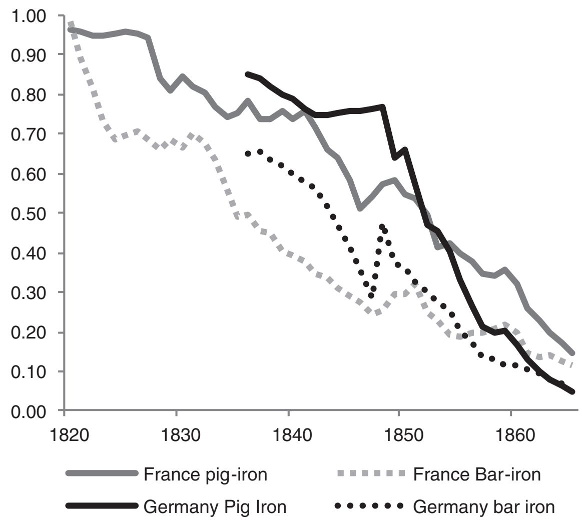


Fig. 5.1 Share of charcoal in the fuel consumption of the French and German iron industries (in percentages) 1820–1865 (Madureira 2012: Annex 1)

shapes and an increase in their height and internal cubic capacity (Benoit 1990; Fremdling 2005). This, in turn, led to the augmentation of average furnace temperatures with an immediate consequent decrease in the amount of charcoal required to smelt a ton of iron.

Altogether, the "Champagne model" sparked a full sequence of changes whose ultimate result was the maximization and valuation of forest resources. Wood was saved because hearths and forges were substituted by coal-fueled rolling mills and furthermore because the new hot-blast furnaces reduced energy consumption by between 12 and 27 % (Benoit 1990: 98–99). Part of the wood thereby released could henceforth be diverted to feed the growing economies of scale in blast furnaces, fostering the gains resulting from specialization. In fact, saving traditional energy sources so as to guarantee a regular fuel supply and end forced interruptions had been a key issue for industry in northeastern France, plagued with shortages and price hikes ever since the times of revolutionary period turbulence and destruction. The adoption of the hot blast was just one more step in long-standing efforts to improve in whichever way possible the consumption of biomass energy and reduce the pressure on forests. From this perspective, there is ample continuity between several eighteenth-century French innovations and nineteenth-century technological adaptations of foreign technology (Woronoff 1984: 245–250; Benoit 1990).

5.3.2 The Second Fight Back

There were, however, other regions where the charcoal industry proved more resilient. Judging from the historical evidence, it is hard not to notice that all these regions were endowed with large ore-mining resources and, most importantly, ore-mining resources embedded in outstanding areas of forests. Sweden, Canada, Russia, and the United States in this respect stood far ahead of any other nation, with their vast, unexplored, and bountiful timber reserves (Zon 1910). A brief overview of the path taken by the charcoal iron industry in these regions shows that the traditional ironworks witnessed a push for modernization strong enough to extend their competitiveness into the second half of the nineteenth century: displaying the criteria of absolute trend reversal, charcoal iron throughput continued to increase until it reached an all-time peak in the 1890s (Russia, United States) or in the 1900–1910s (Sweden, Canada). However, this second wave of innovation bore little resemblance to the continental European "fight back-last gasp" situation, which rested upon continuity of both ownership and investment, furthering traditional production through specialization, geographical inertia, and participation in local markets. On the contrary, the second wave involved development in leaps and bounds, industrial concentration, and the destruction of the old traditional industry, territorial displacement of the frontier of charcoal ironworks, specialized commercial markets, and a broader environmental span. Briefly, a brand new version of the traditional industry cropped up and detached from its historical moorings.

In spite of these differences, the rebirth of the charcoal iron industry proceeded, as ever, on the heels of coke and coal technologies. The first burst of successful imitation and adaption embraced the lesser branch of finery forges, whose obsolescence threatened the competitiveness of final products like bars, plates, rails, and nails. After several failed attempts, Swedish ironworks succeeded in carrying forward to the productive stage ideas arising out of the field of British techniques. In the 1840s, puddling and rolling methods were transposed to old charcoal forges through an adaptation of two separate hearths, the refining hearth and the welding furnace, along with a shingling hammer for shaping the blooms, a welding hammer, and a fine hammer for shaping the bars (Rydén 1998, 2005). A similar change took place concurrently in Russia by means of establishing an indigenous technological base for the production of malleable charcoal iron. This mostly affected the traditional charcoal iron industry of the Urals, which witnessed the installation of *svarochnik*-puddling hearths with lower fuel consumption but larger average capacity (Blanchard 2005). More in line with the last gasp specialization pattern, US charcoal producers hedged their backward development by raising the share of direct casting production from blast furnaces, thus avoiding the incorporation of fining costs. Accordingly, Swedish innovations in charcoal puddling were only very slowly adopted by American producers (Temin 1964: 25–29, 214–215; Gordon 1996: 129–132).

After the turn of the century, a second selective push was set in motion by the building of "monster–coal blast furnaces," a technological development that had begun in Cleveland, Britain, in around 1855–1860, and quickly sprang up in and across other producing nations. Aimed at increasing output by raising the height and at saving fuel by raising the blast temperature and pressure, the new style furnaces unleashed a new cycle of productivity gains coupled with industrial concentration. Its effects were far-reaching upon the industrial landscape: while some

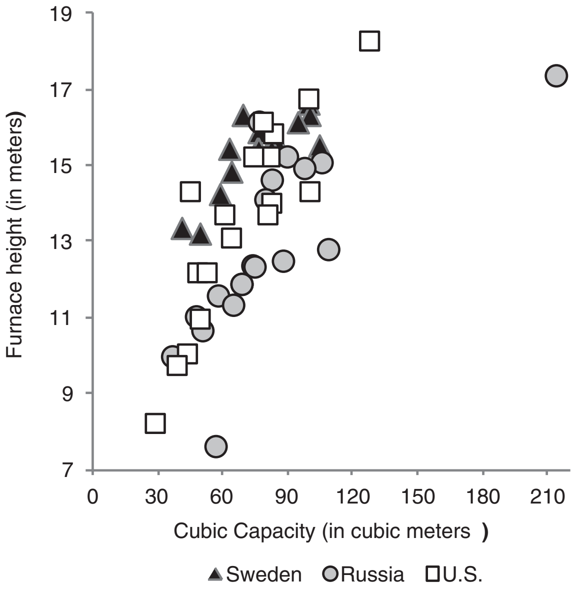


Fig. 5.2 Eighteenth topmost charcoal furnaces in Russia, Sweden, and the USA, 1880–1882. (Kirschhoff 1881; Swedish Blast Furnace Practice 1882; Narrative of the Third Annual Meeting 1882; Some Remarkable furnace work 1881; U.S. Census Bureau 1880)

coke pioneers in the early nineteenth century could make over 1,000 tons of pig iron in a full year, the largest blast furnaces could produce the same amount in a single week by the 1890s (Davies and Pollard 1988: 78–79; Temin 1964: 158–159). In what appears to have been a handy "tip for tap" reply, Swedish, Russian, US and, later, Canadian charcoal furnaces followed in the footsteps and increased height, erecting larger and taller units with powerful blast machinery. As Fig. 5.2 shows, this tendency was fully exploited in Sweden, which achieved regularity in height and design that contrasts with the wider-ranging dimension/capacity of US blast furnaces. At the bottom of scale, Russian plants displayed a smaller height profile situated around 9–13 m which is explainable by the chronic prevalence of cold-blast technology. In the light of this evidence, there is little wonder that the Swedish attained the greatest efficiencies in energy-saving methods with average charcoal consumption per ton of final manufactured rolled iron of 98.8 MJ/ton (3.3 tons in 1860), improved to 59 MJ/ton (an average of 2 tons of charcoal in 1900) and reaching the highest saving of 41.4 MJ/ton (1.4 tons of charcoal in 1900) through the utilization of a charcoal blast furnace with Bessemer decarburization and rolling equipment (Harpi 1953; metric unit conversion based on Jüptner, 1908: 198). This conveys how the newly resurgent Swedish charcoal iron industry was capable of outstripping its counterparts. Nevertheless, its tallest heights of 17–18 m were still dwarfed by the 22–27 m that could then be found in a few "monster" coke blast furnaces.

This point becomes even more striking if we account for the fact that the largest contribution to total factor productivity in the nineteenth century came from technological change in terms of fuel and metal input savings (Houpt 2007). Matters proved rather different among the competing coal–coke furnaces as in this case the trend toward increasing heights triggered a broader process of productivity growth based not only on fuel and metal input savings but also on other interrelated factors less relevant to charcoal ironworks, namely the diffusion of hard driving and improvements in ore selection (Temin 1964: 196–206; Allen 1977; Inwood 1985; Houpt 2007). Minor benefits obtained from minor economies of scale help explain, at least partially, why the charcoal iron industry floundered later in the twentieth century. Aside from tighter limits on productivity gains from economies of scale, two other reasons may explain the final demise of the charcoal industry: the destructive effects of "manufacturing iron without fuel" introduced by Bessemer and open-hearth steel upon charcoal iron products, particularly wrought iron plants and the new opportunity costs for industrial timber (wood chemicals obtained from the charcoaling process, wood pulp, and sawn timber).

Economies of scale became further interlinked with mounting industrial concentration and efficiency gains through mergers and takeovers. Judging by the empirical evidence, it seems the later the innovation occurred, the quicker and stronger came the push toward industrial concentration. The intensity of the movement toward integration and amalgamation into groups among the businesses of latecomer Canada supports this view (Donald 1915; Inwood 1986).

Having undergone a complete overhaul, the charcoal industry overturned a possible trend toward obsolescence. Like the mythical phoenix, ironworks fueled by biomass proved their ability to rise reborn from the ashes. And while these furnaces and hearths continued dependent upon their use of preindustrial energy inputs, their relationships with their surrounding forests nevertheless changed profoundly.

In modern times, the wood supply was assured by a 7–15 km radius of woodland area around the major plant (Sieferle 2001: 63; Harpi 1953: 12) and over much wider areas around populated towns (Sans 2004; Ortego et al. 2011). Subsequently, when average production reached the ceiling of 800–1,000 tons a year, a good yardstick for the first half of the nineteenth century was furnaces being compelled to haul their charcoal over distances of up to 17 or 20 km, something that was achieved "with great expense and vexation" (Warren 1973: 29). In this context, even when the geographic conditions allowed for locating plants near navigable watercourses, ironworks still had to maintain some distance from the water so as to favor the establishment near woodlands, in addition to having to be far enough apart from each other to ensure a ready fuel supply (Knowles and Healey 2006: 620).

It was only during the "Phoenix era," spearheaded by giant charcoal furnaces, that the very idea of drawing resources from the hinterlands became outdated. Henceforth, the natural milieu from which the industry could draw on raw materials was the geographic scope served by the railways. This was certainly an environmental change brought about by technological modernization. A whole new ecology of needs and wants began to surface with several devices invented for long haulage: special packaging procedures for lessening the waste from crushed charcoal; the construction of specialized railroad cars for conveying sacks or baskets with minimal damage; the adoption of mechanical equipment for loading and unloading; and the invention of special cage systems to facilitate transfers as well as improvements in the design of horse and sleigh (Risks in transporting 1883; Lilienberg 1884; Cambridge transportation of charcoal 1885). Still more importantly, the railroad not only provided a market for the new charcoal industry but it also enabled its displacement into new areas. These new areas were basically regions with plentiful ore, like Michigan, Alabama, and Wisconsin in the USA, Ontario and Nova Scotia in Canada, and the central region of Sweden closer to the northern forests. In the same vein as the coal trade, wood was disembodied from the nearby community of forest users and turned into a long-distance tradable resource.

5.4 Ironworks and Forests

Having sketched in very abridged terms the major trends and factors in the iron industry's energy transition, it is now the moment to test some of its consequences. Most generally, economic analysis resorts to the statistical criteria of the market share held by competing technologies to determine the scope of destructive and creative effects (for instance, see Fig. 5.1). However, this point of view does not fully account for the environmental impact upon forests of the switchover to coal-fueled iron technologies. From the perspective of resource endowment, what matters mostly are the ups and downs in the amount of woodland that had to be felled and reforested so as to satisfy charcoal production needs. This becomes especially the case when the consequences of technological change were not single-sided, but yielded contradictory results upon traditional energy sources.

To highlight the various historical paths, Table 5.1 depicts the amount of forested area that was used to feed the ironworks in four nation types: Belgium (charcoal's competitive destruction), France (charcoal's fight back), the USA (charcoal's rebirth with internal coal competition), and Sweden (charcoal's rebirth without internal coal competition). For the sake of comparison, for each benchmark period, the table also shows the virtual forested area that would have been necessary for producing the iron actually made with coal. The third column provides the percentage of forested area reserved for feeding the iron industry and the fourth and final column the proportion of forested area necessary for manufacturing the coal produced iron.

The figures from continental Europe reassert the view of the increasing release of forested areas from iron's dependence. The significant exception is the thirty-five-year period of charcoal "fight back" spearheaded by the French industry in the first half of the nineteenth century that swelled wood consumption. As previously mentioned, this drift took advantage of recent technological innovations in blast furnaces to specialize part of the charcoal industry in the capital-intensive branch of ore smelting. Looking at the data displayed in Table 5.1, it is possible to conclude that the French fight back prompted a 50 % increase in the forested area felled for the iron industry (1820–1840 and 1860). However, due to enhanced coke competition, by the second half of the nineteenth century, iron producers had turned their back on the forests, exploiting only a minor portion of the ongoing available resources. Comparatively, the energy transition in neighboring Belgium was significantly more straightforward and faster. In just a few decades, Belgium iron entrepreneurs, aided by their newly independent government, embarked on the complete transformation of their extraordinary dependence upon scarce forests. Furthermore, in no other country did the iron industry encroach upon half of the national area of woodland (Table 5.1; Belgium 1820). The strains provoked by this particular situation spanned entire regions and the speedy pace of depletion left most contemporaries alarmed (Alviella 1927). However, high dependency also meant that the nation could hardly rely on their internal forestry resources to boost the scale of the industry. This was all the more so as the transportation costs of foreign factors of input still made a difference in the allocation of resources. Under the Belgium banner, the switchover to coal was to shortly mean a release from the constraints of the biomass–solar energy system. Consequently, by the end of the century, by using coal, the iron industry was consuming eight times as much energy as could possibly have been obtained from wood in the country's forests (column "forest equivalent to coal demand/total forest area in 1880").

Unlike mainland Europe, Swedish and US manufacturers were able to continue drawing on their forest supplies throughout the nineteenth century. This long survival of charcoal energy sources should not obscure the booms and busts experienced by both countries as well as their respective aftermaths: the persistence of the traditional industry side by side with the "modernized" traditional industry. Moreover, while the industrial forested area grew steadily in the Scandinavian nation, there was stabilization in US demand, with the plateau reached in the 1840s. Several factors appear to have had contradictory effects. On the one hand, successful imitation of coal blast furnaces and puddling and rolling technologies pushed fuel consumption downward; the specialization in ore smelting with the progressive retreating of wood demanded by charcoal-based refineries in the final quarter of the nineteenth century also contributed to shrinking charcoal consumption; on the other hand, markets for the final goods produced pushed the industry in the opposite direction.

Turning premium-quality prices to good effect, charcoal producers were able to take advantage of the dislocation of the coal-fueled iron supply curve to the right, seizing market niches that became enlarged by the dynamics of lower prices and widespread usage. These comprised temporary niche market opportunities like iron rails and pig iron for Bessemer plants and long-standing quality goods like railroad wheels, plows, scythes, sickles, knives, nails, steam engine boilers and tubes, crank shafts, axles, gears, and even telegraph wire. Globally, the action of opposite factors resulted in increased throughput, more specialized goods, and less fuel consumption per unit of product. Price comparisons between the two competing technologies show that charcoal-fueled iron still retained a premium over coke-fueled iron up to the dawn of the twentieth century (Hammersley 1973: 354–355; Blanchard 2000: 112–113; Olsson 2007: 48–52).

One must nonetheless add that notwithstanding the rising trend for charcoal production in the United States, its development lagged far behind the outstanding boom in coal-fueled iron and steel production. Additionally, the forested area devoted to this type of manufacturing activity was completely irrelevant and occupying at most 1 % of the available surface area (Table 5.1 US. See also assessments made by Sargent 1884: 485–490 and Williams 1987: 16). Seemingly, the limitlessness of American riches offset the impact of charcoal growth throughout the nineteenth century. However, on another scale, the microscale of the region, the abundance of woodland had the perverse consequence of forest conservation mismanagement. Particularly after the Civil War, the negligence in preserving coppices allied to muddled intrusions by cattle breeders curtailed the chances of the tree cover regenerating over entire areas. Ultimately, this would lead to the exhaustion of woodlands around the furnaces and the squeezing of the industry (Williams 1987; Schallenberg 1975). For public opinion, the emerging conservationist movement, and the charcoal manufacturers, the blame for such environmental disasters was pinned on the reckless behavior of agricultural settlers with their wild agricultural clearing methods. The settler rather than the industrialist henceforth became the main enemy of environmental conservation. Precisely at this juncture, charcoal iron manufacturers felt that the moment was ripe to forthrightly fashion themselves as the true "protectors" and the true "restorers" of the forest (On the importance 1880).

5.5 The Last Gasp in Prospect

The iron industry was a powerful engine powering the switchover to coal that took effect in the nineteenth century. Political economists extolled the superiority of blast furnaces in terms of their technological advance, progress, and modernization, underlining the positive effects of the substitution of obsolete biomass-fueled ironworks. Similarly, the energy transition was portrayed as if some snowballing process in which traditional wood carriers got decimated whether immediately or gradually. Although the basic premises of coal ascendancy remained sound, some authors called into question the historical pattern of this supposedly linear evolution. According to Gilfillan, Tripsas, Snow, Young, and others, the mechanisms triggering the destruction of older technologies might also serve to promote their rise. Hence, one could expect to find a complex mix whereupon market creation and market destruction interact to jointly produce a new equilibrium that turns the transition into a period of complementary technological maturation rather than straightforward substitution.

Overall, three distinct patterns of energy transition emerged. A first case in point was the swift release of forested areas due to the iron industry switching to coal as happened in Belgium. This case quite closely resembled the Schumpeterian blueprint of creative destruction. A second situation encountered in France and Germany featured a lingering substitution process with the initial stimulus to exploring forestry resources embracing solely the branch of blast furnaces followed by competitive destruction. The fight back model chronicled by

Snow, Tripsas, and others captures the main traits of this evolution. A third pattern entailed market substitution but with concurrent incentives to the continuation of forestry exploration and the rebuilding of the traditional charcoal industry in accordance with the technical–organizational parameters set by their coal–coke competitors. Sweden, the USA, Canada, and, to a lesser extent, Russia experienced this type of phoenix-like rebirth. Judging from the historical evidence, two conditions stand out as requirements for such phoenix-entrepreneurship success: Firstly, the competing technology must be mature enough to amplify the overall market so as to create product segmentations of sizable scales. We would note that, unlike Young's model, what matters here is the market for final goods as much as the market for intermediate goods. In this respect, it is worth noticing that charcoal's rebirth took place simultaneous to coal-fueled iron and steel seizing the broader market for structural construction materials amidst a fully new technological approach to steel production, developed by Bessemer and Siemens-Martin, deeply changed the market configuration (Misa 1999: 45–48). This means the rebirth occurred under tough competitive pressures. Secondly, the resource endowment must have a positive effect upon prices and upon the choice of technology. As several authors have pointed out (Harpi 1953; Hammersley 1973; Inwood 1985), the vastness of woodlands with huge virgin forests enabled a large charcoal iron industry to persist in countries with moving frontiers long after it had disappeared from continental European nations.

A glance back at the history of the resistance of the old industry reveals not one but two distinct time patterns unfolding over the decades: On the one hand, there was the last gasp type of fight back, taking place in the infant stages of the competition with the new coal-fueling technology; on the other hand, there was the phoenix-like pattern occurring at a later stage of maturity, in which a whole new modernized industry resurrects from the ashes of the traditional charcoal manufactures ingrained from the outset into the highly competitive environment of market integration. Among other aspects, this second pattern of last gasp stood to further benefit from the demand-side incentives of the maturity phase, in particular, the creation of and the expansion of sizably scaled product segmentations.

The most important lessons to take away from the energy transition point to the fact that within the time frame for "transition," one might expect either the effective substitution of the older energy carrier or an incentive to its expansion. Thus, this correspondingly means that the course of action leading to destruction through innovation and cost decreases also unleashes creative market mechanisms. These basically act upon the supply side (technological imitation and specialization) and upon the demand side (new markets for the "obsolete" product) for those industries due for replacement.

As aforementioned, one must not forget that the supremacy of coal remained undisputed throughout the nineteenth century. In terms of the market share for coal-fueled iron, more than half of the world's production came from coal by as early as 1840. And, in the ensuing decades, this market share could not but expand. One must therefore recognize the prevalence of the destructive mechanisms over the creative in the macroeconomic domain and the fact that the new coal-fueling technology enjoyed undisputed leadership in price, output, investment, and scientific research. Likewise, countries that completed a relatively fast energy transition, wiping wood consumption off the map, gained a competitive edge over all others.

The general conclusion deduced from the foregoing analysis is that energy transitions are not just driven by the competitive edge attained by new fuel technologies over their incumbent rivals, or by mechanisms of destructive substitution. If the lessons from the past are to be taken seriously, energy transitions represent first and foremost critical leaps forward in secondary energy consumption, and significant changes in the map of applications, usages, and markets. As the new energy technology challenger reaches maturity, new opportunities for the rebirth or the last gasp of older industries are found to occur.

Giving the escalation of fossil fuel prices, the fading pace of reservoir discoveries, and the likely decrease in world supply of conventional oil, are we currently on the eve of a new last gasp? The answer is clearly no. And the reason is that despite the advances in renewables or biofuels, they will have a restricted reach in the forthcoming years and also a limited sway in the substitution of fossil fuels (Smil 2010). In all likelihood, fossil fuels will meet no real commercial competition from breakthrough technological innovations in the forthcoming decade. Only if steps were taken toward some fundamental alteration of the overall energy scheme, for instance, moving into a new hydrogen economy paradigm, involving decentrally operated virtual fuel cell power, the decarbonization of fossil fuels, and the development of commercial on-board storage systems, would the market position of hydrocarbons experience a radical shakeout. By all accounts, however, the progress toward the hydrogen economy has already proven much slower than expected (Lattin and Utgikar 2007; Winter 2005).

References

Allen RC (1977) The Peculiar productivity history of American blast furnaces, 1840–1913. J Econ Hist 37(3):605–633

Allen RC (1979) International competition in iron and steel, 1850–1913. J Econ Hist 39(4):911–937

Alviella G (1927) Histoire des bois et forêts de Belgique des origines à la fin du régime autrichien, vol I. Le Chevallier et Lamertin, Paris

Ashton TS (1924) Iron and steel in the industrial revolution. Manchester University Press, Manchester

Bell IL (1884) Principles of the manufacture of iron and steel. George Routledge and Sons, London

Benoit S (1990) La consommation de combustible végétal et l' évolution des systèmes techniques. In Woronoff D. (ed) Forges et forêts. Recherches sur la consommation proto-industrielle de bois. École des Hautes Études en Sciences Sociales, Paris, pp 87–150

Blanchard I (2000) Russian railway construction and the Urals charcoal iron and steel industry, 1851–1914. Econ Hist Rev 53(1):107–126

Blanchard I (2005) Nineteenth-century Russian and western Ferrous metallurgy: complementary or competitive technologies? In: Evans C, Rydén G (eds) The industrial revolution in iron. Ashgate, Aldershot, pp 129–150

Boissiere J (1990) La consommation parisienne de bois et les sidérurgies périphériques: essai de mise en parallèle (milieu XVè–milieu XIXè siècle). In Woronoff D (ed) Forges et forêts. Recherches sur la consommation proto-industrielle de bois. École des Hautes Études en Sciences Sociales, Paris, pp 29–54

Cambridge transportation of charcoal (1885) J US Assoc Charcoal Ironworkers 6:117–120

Davies RSW, Pollard S (1988) The iron industry, 1750–1850. In Feinstein CH and Pollard S (ed) Studies in capital formation in the United Kingdom. 1750–1920. Oxford University Press, Oxford, pp 73–104

Davis JJ (1922) The iron puddler: my life in the rolling mills and what came of it. Grosset and Dunlap, New York

de Beer J, Worell E, Blok K (1998) Future technologies for energy-efficient iron and steel making. Annu Rev Energy Env 23(1):123–205

Donald WJ (1915) The Canadian iron and steel industry. Houghton Mifflin Company, Boston

Evans C (2005) The industrial revolution in iron in the British Isles. In: Evans C, Rydén G (eds) The industrial revolution in iron. Ashgate, Aldershot, pp 15–27

Evans C, Jackson O, Rydén G (2002) Baltic iron and the British iron industry in the eighteenth century. Econ Hist Rev 55(4):642–665

Forsythe R (1913) The blast furnace and the manufacture of pig iron: an elementary treatise. D. Williams Company, New York

Fremdling R (1991) Foreign competition and technological change. In: Lee WR (ed) German industry and German industrialization: essays in German Economic and business history in the nineteenth and twentieth centuries. Routledge, London, pp 47–76

Fremdling R (2005) Foreign trade-transfer-adaptationritish iron making technology on the continent (Belgium and France). In: Evans C, Rydén G (eds) The Industrial revolution in iron. Ashgate, Aldershot, pp 29–54

Gilfillan SC (1935) Inventing the ship. Follett Publishing, Chicago

Gordon RB (1996) American iron, 1607–1900. Johns Hopkins University Press, Baltimore

Grossman GM, Helpman E (1991) Innovation and growth in the global economy. MIT Press, Cambridge

Hammersley G (1973) The charcoal iron industry and its fuel, 1540–1750. Econ Hist Rev 26(4):593–613

Harpi G (1953) The supply with charcoal of the Swedish iron industry from 1830 to 1950. Geogr Ann 35(1):11–27

Hayman R (2008) Charcoal iron making in nineteenth century Shropshire. Econ Hist Rev 61(1):80–98

Henriques ST (2009) Energy consumption in Portugal 1856–2006. Consiglio Nazionale delle Ricerche, Naples

Hosier RH and Dowd J (1987) Household fuel choice in Zimbabwean empirical test of the energy ladder hypothesis. Resour Energy 9(4):347–36

Houpt S (2007) Productivity and Transition in Swedish Iron and Steel, 1870–1940. Working papers in economic history, Universidad Carlos III de Madrid-WP 07–10, June 2007. Available via http://www.ekh.lu.se/ehes/paper/Houpt%20Productivity%20Swedish%20Steel.pdf. Accessed 12 Feb 2011

Hyde CK (1974) Technological change in the British wrought iron industry, 1750–1815: a reinterpretation. Econ Hist Rev 27(2):190–206

Hyde CK (1977) Technological change and the British Iron industry 1700–1870. Princeton University Press, Princeton

Inwood K (1985) Productivity growth in obsolescence: charcoal iron revisited. J Econ Hist 45(2):293–298

Inwood K (1986) Local control, resources and the Nova Scotia steel and coal company. Hist Pap/Commun historiques 21(1):254–282

Jüptner HV (1908) Heat energy and fuels: pyrometry, combustion, analysis of fuels and manufacture of charcoal, coke and fuel gases. McGraw Publishing Company, New York (Translated by Oskar Nagel)

King P (2005) The production and consumption of bar iron in early modern England and Wales. Econ Hist Rev 58(1):1–33
King P (2011) The choice of fuel in the eighteenth century iron industry: the Coalbrookdale accounts reconsidered. Econ Hist Rev 64(1):132–156
Kirschhoff J (1881) The manufacture of charcoal pig-iron in Russia. J US Assoc Charcoal Ironworkers 2:200–206
Knowles AK, Healey RG (2006) Geography, timing, and technology: a GIS-based analysis of Pennsylvania's iron industry, 1825–1875. J Econ Hist 66(3):608–634
Landes DS (1969) The unbound Prometheus: technological change and industrial development in Western Europe from 1750 to the Present. Cambridge University Press, Cambridge
Lattin WC, Utgikar VP (2007) Transition to hydrogen economy in the United States: A 2006 status report. Int J Hydrogen Energy 32(15):3230–3237
Lilienberg N (1884) Manufacture of charcoal iron in Sweden. J US Assoc Charcoal Ironworkers 5:253–272
Ortego JH et al (2011) Energy production, ecological footprint and socio-economic transformation of the territory in an organic economy. The case study of early modern Madrid. Universidad Autonoma de Madrid, working paper 03/2011
Madureira NL (2012) The iron industry energy transition. Energy Policy 50:24–34
Misa TJ (1999) A nation of steel. The making of modern America. The John Hopkins University Press, Baltimore
Mohammed SIS, Williamson JG (2004) Freight rates and productivity gains in British tramp shipping 1869–1950. Explor Econ Hist 41(2):172–203
Narrative of the Third Annual Meeting of the USACIW (1882) J US Assoc Charcoal Ironworkers 3(1):305–317 (18–21 Oct 1882)
Needham M (1831) The manufacture of iron. Baldwin and Cradock, London
Olsson F (2007) Järnhanteringens dynamic. Produktion, lokalisering och agglomerationer i Bergslagen och Mellansverige 1368–1910. Umeå studies in Economic History. Nr 35/2007. Available via http://www.umu.diva-portal.org/smash/get/diva2:140326/ATTACHMENT01, Accessed 21 May 2011
On the importance of giving timely attention to the growth of charcoal for metallurgical uses (1880). J US Assoc Charcoal Ironworkers 1(3):67–80
Pollard S (1981) Peaceful conquest. The industrialization of Europe 1760–1970. Oxford University Press, Oxford
Report of the Commissioners appointed to inquiry into the several matters related to coal (1871) House of commons, parliamentary papers, her majesty's stationery office, London
Riden P (1977) The output of the British iron industry before 1870. Econ Hist Rev 30(3):442–459
Risks in transporting charcoal by railroad (1883) J US Assoc Charcoal Ironworkers 4:282–288
Rydén G (1998) Skill and technical change in the Swedish iron industry, 1750–1860. Technol Cult 39(3):383–407
Rydén G (2005) Responses to coal technology without coal. Swedish iron making in the nineteenth–century. In: Evans C, Rydén G (eds) The industrial revolution in iron. Ashgate, Aldershot, pp 110–127
Sans JUB (2004) Combustible para Madrid en la Edad Moderna. El difícil equilibrio entre las necesidades urbanas y los recursos del territorio. In: AAVV (ed) Mélanges de L′École Françáise de Rome. Italie et Méditerranée. Ecole Française de Rome, Rome, pp 683–704
Sargent CA (1884) Report on the forests of North America. Government Printing Office, Washington
Schallenberg RH (1975) Evolution, adaptation and survival: the very slow death of the American charcoal iron industry. Ann Sci 32(4):341–358
Schumpeter JA (1975) [1942] Capitalism, socialism and democracy, 2nd edn. Harper, New York
Sieferle RP (2001) The subterranean forest: energy systems and the industrial revolution. White Horse Press, Cambridge
Smil V (2010) Energy transitions. History, requirements, prospects. Praeger, Santa Barbara

Snow D (2003) Extraordinary efficiency growth in threatened technologies: explaining the carburetor's "Last Gasp" in the 1980s. University of California at Berkeley, September 2003. Available via http://elsa.berkeley.edu/~bhhall/others/Snow03%20carbs.pdf. Accessed 22 Jan 2011

Snow D, Adner R (2010) Old technology responses to new technology threats: demand heterogeneity and technology retreats. Ind Corp Change 19(5):1655–1675

Some remarkable furnace work (1881) J US Assoc Charcoal Ironworkers 2(1):370–374

Space occupied by fuel (1882) J US Assoc Charcoal Ironworkers 3:160–161

Svedelius G (1875) Handbook for charcoal burners. Translated from the Swedish by Anderson RB. Wiley, New York

Swedish Blast Furnace Practice (1882) J US Assoc Charcoal Ironworkers 3(1):385–396

Temin P (1964) Iron and steel in nineteenth-century America: an economic inquiry. The M.I.T Press, Cambridge

Tripsas M (1997) Surviving radical technological change through dynamic capability: evidence from the typesetter industry. Ind Corp Change 6(2):341–377

Tripsas M (2001) Understanding the timing of technological transitions: The role of user preference trajectories, Harvard Business School, Working paper 02–028

Truran W, Phillips JA, Dorman WH (1865) The iron manufacture of great Britain, 2nd edn. Spon, London

Tylecote RF (1991) Iron in the industrial revolution. In: Tylecole F, Day J (eds) The industrial revolution in metals. The Institute of Metals, London, pp 200–260

Census Bureau US (1880) Census of Population and Housing. Washington, Census Bureau

Young A (1993) Substitution and complementarity in endogenous innovation. Q J Econ 108(3):775–807

Warren K (1973) The American steel industry 1850–1970. Oxford University Press, Oxford

Williams M (1987) Industrial impacts on the forests of the United States, 1860–1920. J For Hist 31(3):108–121

Winter C-J (2005) Into the hydrogen energy economy—milestones. Int J Hydrogen Energy 30(7):681–685

Woronoff D (1984) L'industrie siderurgique en France pendant la Revolution et l'Empire. Ecole des Hautes Études en Sciences Sociales, Paris

Zon R (1910) The forest. Resources of the world. US Government Printing Office, Washington

Chapter 6
Oil Reserves and Peak Oil

Abstract The history of estimating oil reserves has proven a history of long-lasting misunderstandings. The split between the environmental perception of depletion and the common man's view, further contrasted the differences between the standpoints of the scientific and business communities, led to the most controversial topic ever debated within the energy community. While some forewarned that a critical moment for humanity was approaching, others found that all evidence and experience pointed in the opposite direction. This chapter traces the history of this debate from the first oil survey undertaken in the USA in 1909 through to current days. We trace how the accuracy of already known fixed stocks (proven reserves) has dominated all public concerns up until the moment when the peak theory of geophysicist King Hubbert shifted the terms of the discussion and grounding it in the size of the unknown—estimates of resources still undiscovered. The final section provides a review of the state of the art in the oil depletion debate.

6.1 Classifying Oil Reserves

Estimating oil reserves has proven the most controversial topic ever debated within the energy community, surpassing even the clash over nuclear power. The controversy harkens back to the very beginning of the twentieth century, grounded in opposite views about the threat posed by oil depletion. This means that we are now into the third generation of contenders. As the dispute unfolded, different streams of thought became progressively entrenched in separate organizations and spheres of influence. What is more, each side narrowly pigeonholed the other: Advocates of free markets were quipped "cornucopians" by associating the efficiency of spontaneous market adjustments in the regular supply of oil with the myth of endless abundance promised by the "horn of plenty" a cornucopia-like symbol of ancient Greece; on the other hand, those warning against oil depletion risks were dubbed "doomsters," to buttress the parallelism between geoecological mindful conjectures and obscure insights into the omens of doom.

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_6,

From its very outset, the discussion was plagued by a profusion of classifications, definitions, and technical procedures and moreover impaired by the lack of binding international standards for the reporting of oil reserves. One important advance in this domain came from the common terminology established by the two most important professional organizations in the field, the Society of Petroleum Evaluation Engineers and the American Association of Petroleum Geologists (in 1997 and 2000, respectively). Beginning with the wider horizon of "petroleum initially in place," they propose the concept of "Ultimate Recoverable Resources" to account for the total amount of oil historically produced; the discovered reserves and the oil to be discovered in the future. Overall, the taxonomy encompasses several levels of forking from the existing mineral stock (resources), to the existing mineral stock which is both technically and economically exploitable (ultimate recoverable resources) over to the existing mineral stock technically and economically exploitable and that could be produced and marketed now (reserves). The point at which resources may be considered commercially exploitable sets the boundary for the new classification of "reserves." "Reserves," in turn, are further categorized according to three levels of confidence:

- Proven reserves (1P or P90) comprise the estimates for which there is a high degree of certainty of commercially recoverable oil given the prices and costs prevailing and under existing technical conditions. In probabilistic terms, P90 classification entails that after the accomplishment of detailed seismic analysis studies, core sampling, well-logging, and other tests, there should be at least a 90 % probability that the quantities actually recovered will equal or exceed the estimate.
- Probable reserves (2P or P50) comprise the estimates of more likely than not commercially recoverable reserves. P50 classification indicates that although unproven, geological, and engineering data suggest a probability equal to or above 50 % that the quantities actually recovered will equal or exceed the estimate. In technical terms, this is, for instance, the case of "reserves in formations that appear productive based on well log characteristics but lack core data or definitive tests and which are clearly not analogous to producing or proven reservoirs in the area" (SPE/WPC 2001: 135).
- Possible reserves (3P or P10) comprise the estimates of reserves less likely to be recovered than probable reserves. In probabilistic terms, P10 classification entails at least a 10 % chance that that the quantities actually recovered will equal or exceed the estimate. Generally speaking, the P10 classification is relevant whenever doubts about fluid, rock, and reservoir characteristics raise questions as regards commercial viability.

Until recently, much of the historical debate on oil depletion revolved around the amount of proven reserves (P1 or P90). In particular, this prevailed in conjectures in which uncertainties over the likelihood of finding untapped oil, favored the view that proven reserves might be nearer the total resource supply potential. With hindsight, these courses of action undersized the estimates in two different ways: Firstly, by disregarding undiscovered resources; and secondly, by downsizing the discovered part. A proportion of the historical controversy over forthcoming scarcity or even depletion did, in effect, stem from equivocal interpretations about what this "proven" category meant. Only latterly did experts realize that the oil reported within this "lot" has been systematically less than forecast expectations. The gap between the physical amount registered and the amount implicit in the main definition of "proven reserves" raised recurring misunderstandings regarding the factual situation faced by the industry. To make things worse, underreporting in field surveys, missing data in national oil surveys and politically adjusted data, amplified these statistical biases.

Field surveys are bottom-up inquiries that gather specific data about production in all the relevant wells/fields. Ultimately, their accuracy rests on figures and reports conveyed by oil firms and specialized information-scouting companies like IHS Energy (formerly Petroconsultants), which means that the dynamics of business interests leaves a most visible imprint on the final figures. From a microperspective, what matters is the proven oil that may represent an asset for accounting purposes and quite commonly equated as the quantity of recoverable discovered oil that might be extracted through recourse to the infrastructure in place. The clout of business-accounting criteria over geological criteria is furthermore reinforced by regulatory preferences for conservative reserve statements, with a view to reducing commercial fraud. Nowhere has this tendency been more evident than in the USA, with regulations imposed by the US Securities and Exchange Commission (SEC) on Wall Street listed oil companies. The Commission has set strict rules for "proven" reserves to be booked, emphasizing the net present value evaluated by discounted cash flow, the business investment committed, as well as respective government approval (Maugeri 2006: 224–225). Even though there is an indirect connection with the physical reality of recoverable oil, the regulatory meaning of "proven" falls rather short of the geological yardstick. With short-term accounting stocks replacing the broader geological principle of well/field life cycle, the problem of underreporting in the 1P category naturally accrued and particularly so in large fields.

At times, oil corporations also circumvent the need to update data by stating, in their annual reports, that the reserves have been more than replaced over the year in question whether by discoveries or by increases in the assessed recovery amounts in existing fields. Over the years, these practices stiffened and corporations moved into a reporting system in which reserves added through exploration just kept up with inventory maintenance. In so doing, they time and again postponed the revision of the real physical existing stock. However, it has been in the domain of national surveys that these missing data became an issue of major concern. One possible explanation is that public agencies and national oil companies skip over the filling in of annual data for reasons similar to private corporations, namely that they find pointless to issue a statement whenever discoveries plus revisions match, to a greater or lesser extent, the production for the year in question. Under such particular circumstances, the standard procedures for editors of benchmark reference publications such as the "Oil and Gas Journal," "World Oil," and "BP's Statistical Review of World Energy" are to reproduce the country's prior annual data. Owing to this, many authoritative sources publish proven reserve figures with no changes throughout consecutive years. Naturally, default reporting added further doubts as to the reliability of statistics, all the more so when half of oil-producing countries usually do not normally disclose changes in reserves (Bentley 2002).

Finally, we arrive at the political distortion of inquiries. Although politics has often interfered with the reporting of reserves, the most notorious case of statistical adjustment occurred in the Middle East oil states, during the 1980s, with proven reserves almost doubling overnight. Identified by the famous quip of "quota wars jumps," the boosting of 1P reserves somehow appeared an ingenious scheme for seizing higher production quotas while taking advantage of OPEC rules: Insofar as production shares were allocated on the basis of the size of reported proven reserves, the increase in declared amounts paved the way to grab a larger slice of the aggregate cartel output. Whether this was just a political expediency allowed by OPEC or also reflected something far more reaching, such as the end of the artificial measuring system established by the oil majors to restrict production, is still under discussion (Laherrere 2001; Maugeri 2006; Bentley et al. 2007). In either case, the unfolding of events does appear driven by political agendas.

Systematic flaws in reporting, omissions of data, and untrustworthy time series deepened the gap between technical classifications and economic–political constraints. In the end, factors other than geological probability and economic certainty combined to undermine the concept of "proven reserves." Regardless of the clear-cut framing of categories, formal definitions struck into the diversity of interests, practices, agents, and regulations. In this context, drawing upon so imperfect a construct to ascertain how much recoverable oil remained in the planet seemed of little help. Yet, experts and public opinion were utterly obsessed with the depletion of proven reserves in 1910, and all over again in the 1970s, sometimes taking the resource potential at face value. A very significant turnaround occurred only at the dawn of the twenty-first century, with the shift of the debate toward 2P—probable reserves.

To fully understand this changeover, it is necessary to realize that, with the passage of time, probable reserves (2P) are converted into proven reserves (1P). The sliding process results from the advance of field exploration and the inherent mastering of geological conditions, improvements in extraction technologies, market changes, and changes in reporting practices. Basically, better knowledge and better technology prompts an upward revision of the recoverable reserves, a process dubbed "reserves growth." In practical terms, the framing of the debate on oil depletion over the boundaries of 2P entailed the recognition of upcoming proven reserves growth as the closest gauge to assess the volumes of accessible oil in the ground. Strange as it may sound, "probable" turned out to be more accurate than "certain": not only did 2P reflect the empirical rule of thumb of "more likely than not," but it smoothly slipped the leash of interference by regulatory organizations, business interests, companies, and governments. Relatively insulated from the political and economic decision sphere, 2P data showed, moreover, high responsiveness to geological breakthroughs and the potential to provide the best estimate for each field's ultimately recoverable reserves (URR).

The final and obvious outcome of the switch toward the probable reserves category derives from the extension of the life span of stocks along with the view that the depletion path had no bearing on a deterministic time limit. The ensuing pages depict the historical contexts of the debate on depletion from the 1910s to the present-day, pinpointing arguments, methods, and outcomes.

6.2 Oil Conservation in the 1910s–1920s

6.2.1 The First Oil Survey

The completion of the first oil survey unleashed generalized fears of imminent depletion in the USA. Much in line with the precedent established by coal surveys in Britain, key advances in the knowledge on fossil fuel stocks spread alarm about the finiteness of non-renewable resources. It was as early as 1909 when Americans first confronted the realization that oil wells might rapidly dry up and all the more likely to do so when the boom in automobiles could surely only worsen the situation.

The first national estimate of oil resources was commissioned from the US Geological Survey (USGS) and integrated into the national inventory of mineral wealth conducted under Gifford Pinchot. The drive behind a systematic inventory of America's mineral wealth bore President Theodore Roosevelt's personal stamp and thus proved both a scientific project and a political venture. For the USGS, as an institution, this represented an enormous challenge as oil expertise was concentrated in regional field surveys, and there were no means available to undertake extensive and accurate estimations on a national scale. To overcome tight deadlines and surmount these practical difficulties, the geologist in charge, David T. Day, was compelled to resort to indirect forecasting methods that took advantage of the undisputed achievements of geological knowledge. One such undisputed area was the study of oil reservoirs as traps. By the time of this national survey, it had effectively become clear that petroleum needed a particular "trap configuration" to exist.

This trap was made up of:

- A source rock of shale or limestone where certain types and amounts of organic matter that give rise to petroleum could once have been deposited;
- A layer or formation of rock, generally sandstone or limestone, both porous and permeable to allow for the formation of a reservoir in the pore spaces, cavities, or fissures;
- A cap rock or "cover," commonly a stratum of shales, clays, or marls, located above the reservoir to retain the petroleum and block any possible surface outflow;

- A structural fault or anticline fold in the strata or a geological "unconformity" through which the reservoir was sealed, while in this trap "gravitational factors," related with different densities, would impel oil to rise above water and gas to rise above oil.

In spite of this common ground, experts did not completely agree on issues including the origin of petroleum (point 1) or the theory of oil accumulation (point 2). While some geologists, for instance, maintained that the oil and gas found in a porous reservoir originated somewhere in that reservoir, others claimed that most oil found its way there from beyond the reservoir's own extent (Johnson and Huntley 1916). Nevertheless, irrespective of these differences, the trap-configuration approach was solid enough to ensure the understanding that each reservoir enclosed a measurable space from which a commercially relevant amount of oil could be drawn. This enabled the oil reserves in place to be statistically inferred from the analysis of volumetric parameters.

David T. Day resorted to the volumetric method in regions where no other accurate information was available. In order to complete the survey's blanks, he made a simplified calculation based on the average porosity of oil-bearing sands and the amount of crude oil obtained per cubic foot of pay sands. He then extrapolated this cubic foot gauge into a rough forecast of entire oil pools and applied a value for the recovery factor, setting a ratio between the total oil found underground and the oil that could actually be raised to the surface under the technical conditions prevailing. Curiously, the volumetric method came to be tested on the regions David Day knew best from previous geological field work: Pennsylvania, New York, West Virginia, Kentucky, Tennessee, Ohio, and Indiana (Day 1909).

The process turned out rather differently when the geologist gained access to primary sources containing oil-well production statistics. In this case, the future recovery could be estimated from the past yield, since the record of the actual output of a well was interpreted as an index to the quantity of recoverable oil. Texas, Louisiana, Oklahoma, and California were the regions selected to apply this method of statistical forecasting. Actually, the method was very simple and dubbed the "percent decline curve." As its very name indicates, the percent graph depicted the decline in the production of a well, of a property, or an oilfield by expressing each year's production as a percentage of the first year of output. Under the assumption that nearby wells will return similar "curves," the amount of ultimately recovered oil could be estimated for entire oil pools.

Barely on the radar before the assessment undertaken by the USGS, the "percent decline curve" subsequently rose to prominence largely on account of its simplicity and utility, particularly for commercial and financial transactions. In one fell swoop, something that was designed with a view to guiding public policy became a decentralized instrument for private calculus applied to such diverse facets as assessing the rate of return on capital; evaluating the amount a company might afford to pay for the rights to a certain oil land; future property values; and the distribution of quotas or taxes among producers (Requa 1918; Beal 1919: 80–89).

Pressed by the political agenda and the deadlines set by the Conservation Commission, the USGS had to come up with a solution able to transpose the expertise held by regional oil-pool surveys into a national forecast. The mixed bag strategy of resorting in some circumstances to the volumetric approach and in others to the cutting-edge approach of production history statistics became the means of circumventing the practical difficulties.

The final figures portrayed bleak prospects for the future. David Day's assessment pointed to 10–24 billion barrels of oil left underground, with a general estimation of 15 billion recoverable barrels of oil (15×10^9). This would last for a minimum of 80 years should the 1909 consumption level be somehow stabilized or less than 25 years should the recent upward trend continue (Day 1909). In short, oil might be about to end actually quite suddenly.

Although the author recognized the conjectural and indirect nature of the assessment, this did not prevent him from proposing clear-cut policies. For David Day, it was the federal government's responsibility to divert oil from power stations, railways, automobile drivers, and exports and channel its usage to vital needs like lubrication and the military for which there were no substitutes. "Waste" was the produce of erroneous choices stirred by market prices, abundant resources, and unsuitable lifestyles (Day 1909). In this vein, depletion ceased to be a demand-side problem like that of uncontrolled wood-cutting in forests and became a moral and social question, a true challenge to citizenship impacting on the everyday life of Americans living off perceivably unsustainable resources. With depletion looming in the next quarter century, mechanisms other than price and individual choice in the allocation of resources seemed all the more justified. Moreover, the case of oil provided the ultimate confirmation of conservationist warnings as it demonstrated that those who had buttressed the moral, patriotic, democratic, and ecological character of natural resources had been right all along.

6.2.2 Forest Conservation and Oil Conservation

Much like a cause that becomes part of its effects, the conservationist administration of Theodore Roosevelt campaigned for an inventory of American mineral wealth, whose conclusions only underscored the need for conservation. Not only did the conclusions agree with the premises, but the conclusions were, to a certain extent, also part of the premises.

Identified by the catch-phrase "progressivism," Roosevelt's presidency (1901–1909) prompted a period of soul searching for an American identity deeply rooted in the President's personal fondness for the "true" American values and way of life such as the open country, breathtaking scenery, the wild and everything associated with the outdoor life: hunting, fishing, horse riding, and wood-chopping. The great and fair America envisioned by progressivism required an active stance at the federal level on environmental issues, sometimes stretching presidential powers well beyond the limits set by the constitution (Cooper 1990). Roosevelt's personal inclinations were further reshaped by increasingly closer contact with the Forest Bureau Director, Gifford Pinchot, appointed to advise Roosevelt on forestry matters. However, the mutual friendship and admiration that ensued brought Pinchot into the inner circle of the White House, and turning him into one of the most influential advisers on political questions with a say over all the critical points on the presidential agenda (Steen 2001: 133–141). Pinchot's political weight meant the conservation policy became attuned to a commercially oriented perspective based on competitive bidding, the regulation of big business, and a call for citizenship. The very language of foresters, with their emphasis on "repairing" forests destroyed by settlers or "restoring" virgin lands, had clear affinities with Roosevelt's message of "regeneration." Culturally, both stressed a return to America's roots. In accordance with key principles such as generational responsibility, social justice, and industrial liberty, the government claimed natural resources should be handed back to the average citizen. In Roosevelt's words, the function of "Government is to ensure to all its citizens, now and hereafter, their rights to life, liberty, and the pursuit of happiness. If we of this generation destroy the resources from which our children would otherwise derive their livelihood, we reduce the capacity of our land to support a population and degrade the standard of living" (Roosevelt 1909: 3). Emboldened by the idea of change and reform, the federal state pushed the institutions designed for the conservation of natural resources to the forefront of national policy and into the press headlines.

With the President's second-term drawing to an end and the survey's discovery that oil would run out in 25 years, the eyes of the administration fastened on petroleum with enhanced drama and urgency. Whereas forest conservation had been the hallmark of Theodore Roosevelt's time in office, oil conservation would prove its legacy. Continuity meant that much of the oil-related public policy was outlined by the agenda, the knowledge, and the legal instruments previously applied to forestry. Furthermore, this also meant that oil geologists, after the foresters, joined the rank and file in the protection of national resources in siding with the federal state. It is important to note that progressivism consolidated a group of highly trained and qualified civil servants imbued with a sense of mission and whose careers depended on the ability to wield preservation as a key political issue for public opinion. The very program of American regeneration was anchored on a powerful social network of university friendships and further cemented through highly personalized and faithful administrative bureaucracies (Schulman 2005). Because corporate actions were sometimes "illegitimate," "under cover," and lacking in "industrial democracy," they ought to be counteracted by equidistant public powers exercised by leading experts in transportation, agriculture, geology, utilities, and public health (Miller 2009). The best antidote to "the relentless exercise of unregulated control of the means of production" was joint action by government, scientific expertise, and citizenship. This triple alliance formed the core message of progressivism. Indeed, the Director of the US Geological Service had no problems in putting this down in black and white: "I have come to think of geology more as a phase of citizenship than as merely a branch of science" (Smith 1920).

Once the case of impending depletion spilled over into public opinion and the "oil fraternity" at large, two areas of federal and state intervention surfaced with paramount urgency: One was precautionary action so as to guarantee the fundamentals of national security; and the other involved rationalizing measures to tackle all sources of waste and find means to economize on oil consumption. On both fronts, petroleum geologists filled an important role by placing themselves as interpreters of the national interest. Unsurprisingly, the very first measures devised to face possible exhaustion mirrored the very same that had been applied in forest conservation. Security concerns prompted a strong position over oil-land ownership with rationalization efforts triggering a debate over tighter regulation. Globally, the 1909 oil survey proved instrumental in tilting the balance of power toward the "permanent public good" and against the "merely temporary private gain" (Roosevelt 1909: 4).

6.2.3 The Ecological Commonalities of Oilfields

The idea of ring-fencing strategic natural resources away from commercial usage by placing them under the management of the federal state was first enacted by the Amendment to the Land Revision Act of 1891. Though this, the US Senate recognized the President's authority to establish forest reserves by proclamation, leading to the constitution of a non-market sector set aside from the lands that would otherwise be commercially available. However, it was only during Theodore Roosevelt's presidency that the purchase of Western and even Eastern lands shot up and leading to the founding of a federal reserve totaling 150.8 million acres, spanning 159 forests, and extending over 27 % of the USA's forested extent (Hays 1959; Brown 1919: 3–7). Nevertheless, the greater the advance in Roosevelt's policy toward public ownership, the deeper the hostility of Congress toward such conservation in preferring to bow to local commercial pressures for quicker private sector exploitation. The clash came to a head in 1907 when Congress forbade the creation of more forestry reserves in Western states on the grounds that the creation of public property and environmental concerns served merely as the unjust means of expanding presidential authority (Penick 1968).

Under these circumstances, it became mightily difficult to transpose the recipe for public forest ownership onto the oil realm. All the more so when the geological warnings about forthcoming depletion were clearly offset by the intuitive evidence of prices, driven down to an average of $70 cents for a 42-gallon barrel (Williamson et al. 1968: 38–39). Inasmuch as the current evidence ran counter to the geological forecast, there was a kind of pathological split in public opinion. However, against all odds, a small but influential group of oil geologists, backed by the director of the US Geological Service, George Otis Smith, was able to circumvent opposition and carve a federal oil reserve through withdrawing public lands in California, and their subsequent conversion in order to supply the navy. Working behind the scenes and taking advantage of consolidated intra-governmental networks, geologists were able to seize the opportunity that came about in California, firstly by withdrawing prime oil land from the agricultural register and then by switching its usage to public property. As a result, in September 1909, roughly three million acres of oil-rich lands became the future Naval Petroleum

Reserves (Shulman 2003). This proactive course of events was particularly significant because when the breakthrough decision was made, the Navy remained undecided over which battle ship types should be fully converted to fuel oil. Equally, the usurpation of private ownership arrayed the opposition of the most significant sectors of the petroleum industry (Olien and Olien 1993: 49–50).

Aside from the shift in natural resource property rights, the first oil survey also set in motion unprecedented appeals for regulation, on behalf of saving the threatened liquid fossil fuel reserves. Geologists and public authorities repeatedly asserted that consumption should be constrained. This was particularly true in the righteous domain of automobile driving, which was deemed appropriate only to unavoidable work-related activities such as making deliveries, transporting doctors, transporting children, guaranteeing public order, and easing the life of isolated farmers. Beyond this array of utilitarian functions stood nothing less than the hedonistic usage of cars "for pleasure," a type of social behavior increasingly targeted not only by conservationist writings but also by articles published in specialist magazines such as "Motor," "The Motor Age," or the "Oil & Gas Journal." As the chief geologist of the US Geological Service put it in an interview that addressed the dangers of depletion: "the use of pleasure cars is growing beyond comprehension" (White 1919). More telling was the attempt to regulate the industry from the supply side, where sizable wastes were perceived as occurring. Scarcity implied the rationalization of exploitation. By the close of the nineteenth century, experts in scientific forestry had discovered the potential clash between the needs for the common management of natural resources and the economic system of separate ownership and private appropriation. The discovery of the problem of ecological commonalities (indivisibility, interdependency, sustainability) had the effect of pushing the state into previously excluded areas by asserting the importance of expert knowledge, norms, and regulations. In fact, it might be said that the ascension of professional scientific groups came with the territory.

6.2.4 Gas Pressure and Oil Recovery

Although the oil industry long knew about the tremendous waste involved in legendary oil rushes and town-lot developments like the Sindletop oil-field of Texas or the Breman oil-field of Ohio, the explanation for the harmful outcome boiled down to the reckless behavior of adventurers, real-estate speculators, and wildcatters. It was only by the close of the First World War that the issue of waste came to be perceived less in terms of economic greed and more in terms of the geological preservation of reservoir indivisibility, interdependency, and sustainability. What transformed the reservoir into an ecological commonality was the understanding of the bond between gas pressure and oil recovery: Since the pressure of gas forced oil out of the rocks into the wells, pressure turns into a matter of great economic interest. One of the most influential manuals on oil geology summed up petroleum extraction as a two-step process in which the bore becomes gradually filled with oil, accumulating gas below, until the pressure is sufficient to cause the oil to overflow. Then, as the oil flows to the casing head, pressure is relieved allowing the gas to expand suddenly and to rise up the column with force (Emmons 1921: 184). Doubts still remained as to whether any increase in temperature, in accordance with the deeper burial of organic sediments, would linearly accelerate the process of gas formation, leading to an overall increase in pressure. As a general principle, it was accepted that deeper reservoirs would return higher pressures; but geologists also pointed out, cautiously, that "rock pressure" was driven by an array of factors that acted over time: hydrostatic pressure; weight of superincumbent strata; rock movements; deep-seated thermal conditions; long-continued formation of natural gases; and the resistance to fluid in accordance with the principle that the level of maturation of the organic matter, through which lighter gas hydrocarbons are generated (methane, ethane, propane, butanes), increases exponentially with temperature and linearly over time.

In addition to natural conditions, human-made intervention plays no less a role in the reservoir pressure level. It was particularly the spacing of wells and the pace of oil extraction that captured the attention of geologists and public authorities. Industrial practices once tolerated henceforth came under close scrutiny (Requa 1918). The threat of depletion and the harbingered scarcity brought into the spotlight practices like allowing gas to issue freely from open wells or gas flaring, which both contributed to abnormal decreases in reservoir pressure, and was subsequently banned in several states. The legal framework for property rights known as the "rule of capture" was also criticized as a source of waste. Applied to oilfields, the rule of capture meant the owners of land atop a common pool could take as much oil as they wanted even when unduly draining the pool and reducing the output of nearby wells. This process meant everyone had to engage in a desperate struggle to extract just as much and as swiftly as they could. The ensuing haste and dense well spacing led to a steady drop in reservoir pressure with an inherent reduction in the volume of oil capable of being brought to the surface during the exploration life cycle. Contemporaries estimated that as much as half of the petroleum in US reservoirs remained underground after the fields ceased to return effective yields. This value was later corrected to between 75 and 65 % and finally settling on 60 % and interpreted as a regrettable dissipation of national resources. (Emmons 1921: 184; McLaughin 1939: 127; Schurr and Netschert 1977: 357–358). Ultimately, the amount of oil reserves could grow simply by improving the oil-recovery factor. However, to achieve this goal, production had to be more efficient, more science based, and more regulated.

Stirred by eager conservationist exposés, petroleum shortage forecasts made good copy in popular periodicals and provided appealing headlines. Soon, the basic principles of conservation made their way into Oklahoma State through the institution of prorationing among the wells, control over storage and transportation facilities and restrictions over the subsurface waste that caused pressure depletion (1913 and 1915, albeit with few practical consequences). Later on, Texas and Kansas followed in the footsteps of the Oklahoma regulations. With American participation in First World War, conservationism was temporarily diverted from its inward regulatory drive and oriented toward government–business cooperation.

The rapprochement stems from the idea that US shortages would have to be met by acquiring foreign oil lands and by taking a more aggressive stance in support of corporate interests abroad (Nordhauser 1979; Clark 1987).

To conclude this section, the first oil survey was framed by the necessity to present data on the conservation of natural resources so that the final figures released matched the pattern of a readily quantifiable total. Owing to the usage of expedient methodologies, tested on scattered oil pools, inferences from volumetric parameters and inferences from historical records of production were aggregated into national forecasts. In this manner, the amount determined, fifteen billion oil barrels left in "reserve" represented a menace to the future. Geologists looked at the glass as if half empty rather than half full. It suffices to point out that if the amount accrued by new discoveries (plus revision of the previously found and now recoverable petroleum), exceeded the amount of oil extracted from existing fields, the results would have shown a net increase in the volume of reserves. Indeed, as long as this situation lasted, the deadline for depletion would be extended rather than shortened. It was precisely this type of knowledge deriving from the business dynamic of discovery–exhaustion–new discovery that the 1909 geological survey utterly failed to take into account (Olien and Olien 1993).

Important as this logical viewpoint may be, the fact is the rate of discovery did slow considerably just after the survey's publication. The 1910s was a decade of "dry" wells, rising prices, and new discoveries falling short of replacement needs. National Petroleum News, the journal of the independent oilmen, reported in 1913 that "during the past years the prospector has gone over the country with the drill, selecting the most favorable locations, and has not in a single instance been rewarded with a barrel of commercial oil, outside of what is generally accepted as the proven area" (Dunham 1913). Because compensation for growing demand barely occurred, the balance between withdrawals and additions to petroleum reserves moved the countdown of the time elapsing until exhaustion from 17 years (1918–1919 surveys) (Clark 1987: 148). More than anything, such results reflected the normative acceptance of what is currently called the 1P reserves or the narrow classification of recoverable discovered oil that might be extracted resorting only to the infrastructure in place as an indicator of the overall risk of depletion. This course of events definitely leaned toward the interests and the views of conservationists and geologist.

6.3 The Technologies of Discovery

6.3.1 Finding Oil

The first concept of oil reserves closely embraced the amounts of oil available in tapped reservoirs. Forecasting techniques like the volumetric and historical–statistical methods required that some successful drilling had already been carried out. In this sense, reserves resulted in ex-post measurements, with geologists tracking a path previously opened by wildcatters and oil companies. Considering the epochal criteria, two omissions stand out as particularly relevant:

- The first is the failure to account for enhanced recovery practices implemented in pools that had long since passed their maturity, such as New York and Pennsylvania. In these regions, continued production was maintained chiefly by cleaning and deepening old wells or by obtaining oil from shallow sands, which had been thought too insignificant when the wells were first drilled (Bacon and Hamor 1916: 69). Thanks to these recovery methods, new oil from exhausted fields could be added to the reserves existing. While only small amounts were at stake in the 1910s, the importance of enhanced recovery methods (ERH) would attain new heights in the 1920s with the injection of gas, the injection of compressed air and flooding water into reservoirs on the verge of exhaustion (Miller an Lindsly 1934). Fostered by a string of technological improvements, recovery techniques rebounded again in the 1950s and 1960s with steam injection, the injection of water solutions with polymers, surfactants, or caustic chemicals, in situ combustion, and electric hydraulic shocks.
- The second was the omission of prospective and untapped reserves. To put the 1909 oil survey into perspective, it is worth recalling that for nearly 50 years geological coal surveys had followed the practice of ascertaining the recoverable coal left behind in pits plus assessment of seams with "hidden coal." Referred to as existent, probable and possible reserves, this assessment was quantitative in nature (Madureira 2012). Insofar as oil reserves were equated as fixed assets, the depletion of reservoirs could somehow be thought of as the depletion of a non-renewable forest. The geological survey thus became a contentious issue that carved a trench between the business view of a drifting amount determined by new discoveries and the official view of a fixed amount determined by the already confirmed oil reservoirs. Hence, the stage was set for a public confrontation between those who claimed "a petroleum famine is imminent" and those who countervailed with "there will always be enough petroleum to meet demand" (Garfias and Whetsel 1936: 213).

Henceforth, surveys were clouded by the suspicion that the conservative nature of the forecasts set the tone for those who argued in favor of government interference through regulation, prorationing, production controls, waste disposal, or even—the rumors persisted—partial nationalization. In an attempt to calm these troubled waters, in 1922 the USGS mobilized 10 geologists representing the American Association of Petroleum Geologists and six from the USGS for a comprehensive and accurate study aimed at once and for all stemming the controversies and bringing the debate back to indisputably geological grounds. For the first time, the distinction between known fields and undiscovered reservoirs was acknowledged. The oilman's view of exhaustion discovery cycles was translated into probabilistic categorizations that accounted for "prospective" and "possible" oil. The concluding estimate identified 5 billion (5×10^9) barrels of crude "in sight" and an additional 4 billion barrels as "prospective" and "possible." The former was judged "reasonably reliable" with the latter deemed absolutely "speculative and hazardous." In the end, neither the enhanced accuracy of petroleum in sight nor the acknowledgment of "speculative" discoveries reassured the industry. On the contrary, the enduring politicization of the geological survey opened the door to the institutionalization of competing reports on petroleum reserves sponsored by the government, by specialized reviews (Oil & Gas Journal, Oil Weekly), and by the American Petroleum Institute business association. From 1922 onwards, this pluralism of estimates became the rule: each vested interest, each major institution produced its own forecasts. Maybe the surprising issue in this evolution toward customized surveys is that there were hardly any discrepancies in the final figures of proven reserves, although that did not halt public and private bickering between institutions (Dennis 1985).

The crux of the matter was naturally the amount of oil still undiscovered. In this regard, the uncertainty could hardly be solved. There were bold stands on the subject, but little means to figure out a reasonable and acceptable forecast. As regards finding oil, geological knowledge had limited utility: It could forecast where oil was not supposed to be found (for instance in rocks dating from the Jurassic, Permian, and Silurian Eras), and it could provide some advice on defining areas worth exploring (areas of extensive limestone dolomitization, salt domes, or beds of porous sandstone lying within shales) (Johnson and Huntley 1916; Bacon and Hamor 1916). Nevertheless, the only way to be certain about oil reserves was to drill; as an experienced field-worker reported: "geologists have gone deeply into the matter and in a way seem to be able to select oil producing territory. But they are not infallible. A hole in the ground seems to be the only sure test" (Horlacher 1929: 24).

6.3.2 Surface Indicators and Seismic Methods

Up to the First World War, all geological knowledge was in fact exclusively based on surface indicators providing a vague clue as to the location of underground reservoirs. Throughout the USA, the most reliable signal for the oil prospector was the localization of natural eruptions like oil seepages or springs, natural gas springs, outcrops of sands impregnated with petroleum or bitumen, bituminous dikes, and bituminous lakes. These "eruptions" were the first feature to look for as they demonstrated that at least some oil existed in the vicinity and was able to migrate to the surface. Other sedimentary formations such as sands, sandstones, shales, and limestones were also potential though less certain clues. For field-working American geologists, this hint was nonetheless of limited relevance since the few unveiled seepages quickly got drilled by wildcatters. On the contrary, seepage search did prove very productive in countries such as Mexico and Azerbaijan—Russia where oil and gas leaked copiously from source rocks. So abundant was this type of primary surface indicators that the methodology for the second comprehensive Mexican oil survey relied chiefly on inventorying

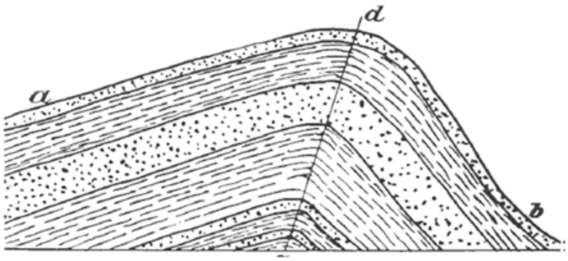


Fig. 6.1 Diagram of a non-symmetrical anticline (Decker 1920: 6)

"chapopoteras" (seepages) scattered all over the country and complemented by a geological description of the underlying sedimentary rock structure (Villarello 1908). Before the 1910 revolution, the country had consolidated a hub of national oil geology expertise centered in the "small but highly respected organization" of the "Instituto Geológico de Mexico," which kept in close contact with their North American colleagues (Owen 1975: 246–256). In Azerbaijan, on the other hand, far-reaching seepages made the tapping of oil from surface wells a remunerative business for local tribes and an ecological nightmare once every amateur, adventurer, and speculator began drilling at random during the oil rush of the 1880s. In truth, drilling appeared to be the single talent required to find oil.

Finally, in the absence of any such clear-cut indicators, geological advice could do no better than recommending searching for the usual landscape fold bed surfacing in an upwards convex form, with the oldest geological beds at its core. Unlike the former empirical guidelines, this particular suggestion was grounded on a theory of oil occurrence—in fact, the most accepted epochal theory within the scientific community: the anticlinal theory of oil accumulation (Arnold 1923; USGS 1934). This convex salience identified by the observer was likely to match a geological structure called an anticline. Anticlines are geological structures involving rock formations bent by a tectonic process into an upwards convex configuration and whose fold traps form an excellent reservoir for hydrocarbons, particularly when containing reservoir-like rocks at their core and impermeable seals on the outer layers (Fig. 6.1). The hypothesis that an extended "nose" at the surface could become an underground petroleum-bearing fold aroused interest in the systematic exploration of the American countryside, bringing topography back into the arms of geology. From the common perspective, this was summed up in the unwarranted idea that "all oil is found in folds" (Johnson and Huntley 1916: 50).

However, perhaps the most important contribution of the anticline theory to petroleum discovery lay in the technical innovations that accompanied it, especially the systematic observation of rock altitudes and the representation of anticlines by contour-line subsurface maps. Invented for a geological survey undertaken in Trenton, topographic contour lines represented lines in depth below sea level so that the highest points on the map were labeled with the lowest values. By disclosing the topographical relationship between the observable landscape and concealed petroleum reservoirs, the maps triggered debate about the whereabouts of gas and oil deposits. Above all, this new scientific "gadget" proved extremely useful to impress the value of geological prospecting on both the public and companies. As expected, geologists endeavored to play their trump card by every feasible means.

The anticline theory gained momentum as more oil was found in anticlines with oil traps than theoretically predicted. West Virginia and south-western Pennsylvania offered the best supportive evidence in this respect; conversely Ohio, Indiana, and Illinois cast serious reservations on the global validity of the theory. We know today that most of the world's oil was in effect discovered in anticline structures (Downey 2009: 98). However, this fact, per se, did not significantly raise the earlier probability of actually finding oil. Even when selecting anticlines as their main target, geologists of the 1920s could not single out precise location criteria. Surface indicators said little about whether or not anticlines might contain oil and gas, the amount of hydrocarbons in place, where the accumulation occurred, or the configuration of structural and stratigraphic traps. Ultimately, they could miss the spot simply because the oil was not at the top of a pronounced anticline or because the trap had an unexpected stratigraphic configuration. Furthermore, since oil was found in a great variety of structural positions, the basic anticline hypothesis underwent many vicissitudes (Hager 1923; Hubbert 1966).

The work with surface indicators required a sizable and labor-intensive organization. Nowhere as in the prospecting of foreign lands was this feature so remarkable. One may even say that an era of geologically inspired "invasions" began with the dawn of the twentieth century sometimes involving the overseas relocation of battalions of forty to two hundred men. This stream was fostered by planned investments made by the largest oil companies and reflected the pressure to find untapped sources of supply in the face of increasingly global competition. Mesopotamia (1904 and 1908) Trinidad and the British West Indies (1908), Argentina (1908), Ecuador (1909), Egypt (1911), Algeria (1914), and Venezuela (1917) were the most eminent cases of success in finding oil abroad. A geological expedition to China and Formosa (1914–1916) commissioned by the Standard Oil Company of New York also suggested that there was a likelihood of discovering good reservoirs, but the advance toward the production phase stalled for political reasons. In addition to the new production regions, multinational oil companies further reinforced their presence in Canada and in Peru, leading to a new cycle of discoveries, notably in Peru. So overwhelming was this trend that even firms long skeptical about geological endeavors ended up recruiting 10, 18, 26 geologists (Persia, Anglo-Persian, 1919–1924). Given the higher costs of oil prospecting in the international arena, the massification of discovery had to be spearheaded by some new institutional form of doing business: The multinational holding company was precisely the organizational structure able to finance a multiform presence in oilfields around the world.

After the First World War, the strategic commitment of these large corporations to get hold of secure supplies by constituting buffers of private reserves intensified the scrambling for oil and for leases. Soaring prices further increased the payoffs for each dollar invested in prospection. The more proactive geological stance prompted a phase of swift technological innovation with a bet on every technique that might disclose the sedimentary layers and structures lying beyond the anticline's surfaces. Between 1919 and 1929, the core of geophysical technologies, as we currently know them, was experimented with for the first time, improved, and put to good usage.

Gravity surveys, magnetic surveys, and seismic surveys derived from the idea that variations in rock density could be mapped by measuring the way they conveyed some signal. Hence, experiments with the torsion balance, a scientific instrument devised by the Hungarian Baron von Eoetvoes, relied on the assumption that the gravitational force exerted by low density ("light") rocks found close to the surface is less than those of very dense ("heavy") rocks. By the same token, the electrical current sent by a magnetometer depicted a different magnetic "anomaly" when encountering less magnetic sedimentary rocks and when coming across highly magnetic igneous rocks, thus enabling the identification of the former where oil was more likely to be found. Last of all, a concussive sound produced at the surface, in such a way that as much of its energy as possible was directed downwards, was then partially refracted backwards with greater or lesser velocity depending on the density or compactness of the geological formations encountered. In this echo-sounding technology, a picture could be formed by registering the way in which the velocity of vibrations changed with depth. The time taken for the sound wave to reach a seismic detector located on the surface was recorded on a strip of photographic paper. Owing to the fact that the speed of transmission was proportional to the density or compactness of the geological formation, the technique was firstly used to detect salt domes, which returned a high velocity of propagation. Later on, seismic refraction methods were improved and applied for the mapping of other rock strata (Forbes and O'Beirne 1957: 120–122).

Conceived for general scientific research in geodesy and geophysics (the gravitational method), for iron ore prospecting (the magnetic method), and for the location of enemy artillery firing positions (the seismic method), these technologies had to be further adapted to the particularities of oil surveying. As Bowker (1994: 22) pointed out, during the first phase of learning and adjustment, the data produced by prospecting instruments could be correlated with underground structures and those structures could sometimes be correlated with the presence of oil. Nevertheless, as of the 1920s, no link in this chain had been firmly established. It was only through further research and practical tests, financed by oil companies such as Amerada Petroleum Company, Royal Dutch Shell and Shell's affiliate Roxana, Gulf Oil and its subsidiaries, Louisiana Land & Exploration, Calcasieu Oil, Standard Oil of New York, Humble, Pure and Louisiana, Aguila and Burmah Oil, that fundamental improvements were brought about. Within a short period of time, these investments paid off and paid off handsomely. Successful discoveries of new reservoirs in southern Texas, in the USA, Mexico, and Hungary arose from the application of gravitational methods, while discoveries in states adjoining Texas, such as Louisiana, and Mexico derived from seismic refraction methods, while new finds in Texas, as well as Venezuela and Rumania, were brought about by innovative usage of magnetic surveys and electric logs (Williams 1928; USGS

1934; Forbes and O'Beirne 1957; Owen 1975; Bowker 1994; Robertson 2000; Petty n.d.).

Afterward, the effectiveness of these gravitational and magnetic methods became increasingly associated with reconnaissance surveys and efforts to measure sediment thickness. The seismic method additionally broadened its scope and seized the general-purpose geophysical exploration market outside of Texas, largely on account of its reliability, cost-benefit advantages, and enhanced opportunity "for securing preferred acreage over mapped structures" (Bignell 1934). The trend that turned seismic methods into the bedrock of core oil prospection activities was further reinforced by two international developments: First, the boom in offshore exploration that began in the late 1950s and was chiefly based on seismic marine surveys; in this respect, the production of waterproof microphones (hydrophones) deployed along a cable or a steamer proved to be, far and away, the cheapest and most efficient technology; second, the interface with computing power which led to 3-D seismic surveys and the revolution in "the process of exploration and production, since the early 1990s." Among other aspects, 3-D surveys had the advantage of easing the identification of the optimal drilling point (Downey 2009: 100–101).

6.3.3 The Oil Glut

A final piece in this puzzle may be called luck, coincidence, or the unexpected coincidence of different series of events: In 1926 and 1927, a series of discoveries in Oklahoma, Texas, and New Mexico hit some of the largest oil concentrations in the world, adding almost five thousand million (5×10^9) barrels overnight to the proven reserves of the USA. The frenzied oil boom that ensued flooded the markets and drove prices down, silencing the "famine," "shortage," and "exhaustion" theses for the next 50 years. Institutions that had been founded to deal with scarcity and to fight "waste" were subsequently reshuffled to enforce conservation through the self-regulation of the industry. Overall, the rise of geophysical exploration played a minor role in this spurt (circumscribed to part of East Texas) as most of the discoveries resulted from wildcat drilling practices and surface indicator insights. Hence, the urgency that had once turned geophysical exploration into a key science for the future of humanity became less momentous. Conservationist ideas were also hit. The oil being endlessly pumped out of the earth simply washed away the bleak predictions of the early 1910s.

The history of estimating oil reserves proved to be a history of long-lasting misunderstandings. Although American geologists combined volumetric and statistical methods specifically tailored to the realities of the petroleum industry, the final figures from the first oil survey, released in 1909, came to be interpreted by analogy with the forest conservation practices and policies. The discovery of 15 billion barrels of oil left in reservoirs was regarded as a sort of opening shot in a race against the clock of depletion. "Proven Reserves" were understood as a stock; a finite stock that had to be economized, held back, and set aside for future uses or contingencies. By adopting terminologies with familiar nontechnical meanings, and furthermore colored by the moving debate on presidential powers and federal forest "reserves," geologists ascribed the meaning of the concept to an observable fixed asset. Furthermore, given there was, after all, not so much of it left underground, they conveyed the idea that America was approaching its resource supply potential. What ensued was a sort of pathological split between the "scarcity" and the "overflowing" stances, a split sturdily entrenched in discourses, social networks, newspapers, journals, and institutions.

To counteract looming claims over the need for regulation, the API—American Petroleum Institute set up its own survey, based on preferential access to oilfields and business records, thereby building a spotless reputation for data gathering. Its aim was to replace geological uncertainties by a narrow but accurate appraisal of oil reserves. Between 1925 and 1935, all open possibilities were locked-into the concept of "proven reserves," grounded on technical and economic feasibility. Such closure ran against the grain of current technological improvements as it excluded probabilistic methods of oil finding by the geophysical sciences and neglected the still novel enhanced oil-recovery practices. This means that, in the end, economic–political factors superseded the evolution in technology and science.

6.4 Peak Oil

6.4.1 Hubbert's Peak

The time of harbingered catastrophes constituted a time of learning, discovery, and experimentation related to the awareness of a civilizational change. Societies were confronted with an epochal shift that represented the dislocation of the core engine of growth from reproducible and universal productive factors toward exhaustible, non-renewable, and unevenly distributed resources. With fossil fuel consumption mounting to unprecedented levels and no clear idea of the remaining stocks underground, there was a growing sense of alarm as to whether the resources might run out in the near future.

Ultimately, the pessimism associated with the first forecasts sprang from a determinist conception of human action. Determinist is used here to mean that given an array of initial conditions, no alternative outcomes were feasible. For depletion viewpoints, not only did all factors push toward the same outcome, but the unintended consequences of these factors also did so. Strange though it may seem, the first generation of conservationists did not foresee (or, at least, failed to mention) that, at a certain point, consumption would begin to fall if the price of natural resources began rising. The disregard for every factor that might counteract the lemming-like rush toward "depletion day" set the stage for the realization of a catastrophic outlook. With economic growth, technological innovation, population growth, and social affluence all driving the increase in fossil fuel consumption, the time would be reached when the last non-renewable resource was consumed and spent. The striking point in this conception is that peak production fully coincided with depletion: Fossil fuels would cease to exist precisely when humanity was consuming the most commercial energy. It is this assumption of rising consumption trends through to depletion that is largely responsible for the dramatic, teleological representation of a Doomsday scenario. Central to such a view is the formalization of reserves-to-production ratios(R/P), calculated by dividing the estimate of proven reserves by the current level of production. The result discloses the time remaining before oil becomes completely exhausted, assuming the current level of extraction remains constant. Bluntly interpreted, reserve-to-production ratios become the telltale signs of an impending, unavoidable, threatening deadline.

Notwithstanding the doubts and uncertainties that surrounded the depletion scenario, the engagement of governments, social interests, and scientists in a prolonged public debate contributed to ingraining the idea throughout societies that the valuation of the future entailed choices in the present. And the more extensive the appraisal, the more justified the change. Following the theoretical insights made by other social sciences, we make recourse to the "time discount" concept to assess how people ascribed a present value to the rewards to be received in the future (Horowitz and Carson 1990; Frederick et al. 2002; Elster 2007). Somehow, time discounting quite sharply reflects the environmental stand toward conservation and how much people are willing to save or sacrifice their current benefits for the sake of a more balanced future. According to this concept, a high time discount rate means that future rewards hold only a small present value so that a higher discount is placed on potential returns. Low discount rates mean the reverse: A higher appraisal of what is to be collected in the future. Moreover, it is assumed that people may either discount utility according to inter-temporal preferences or set their own discount preferences in line with political and moral values such as intergenerational solidarity or landscape preservation.

As Fig. 6.2 demonstrates, the environmental point of view held by conservationists is represented by a dashed line. They perceive access to natural resources as part and parcel of citizenship and that the market should discount future profits at the same rate as society would wish to discount the welfare of future generations. As these natural resources are not just economic goods but the national heritage of future generations, the closer society comes to "depletion day," the scarcer the asset becomes and the higher its present value should also become. Therefore, the future rewards of present resources tend to evolve in a hyperbolic slope.

The second perspective is that of political economists and is portrayed by the solid line. What distinguishes this strand of thought is its recognition that the current value of fossil fuel resources will tend to increase in line with the approaching date of exhaustion. The reason for the future escalation of prices is the "degradation of costs" provoked by extracting oil (or coal) from deeper, smaller, and thinner reservoirs. Although technological improvements could contribute to slowing this trend, their overall effect is offset by the greater costs inevitably incurred as extraction meets new and more difficult conditions.

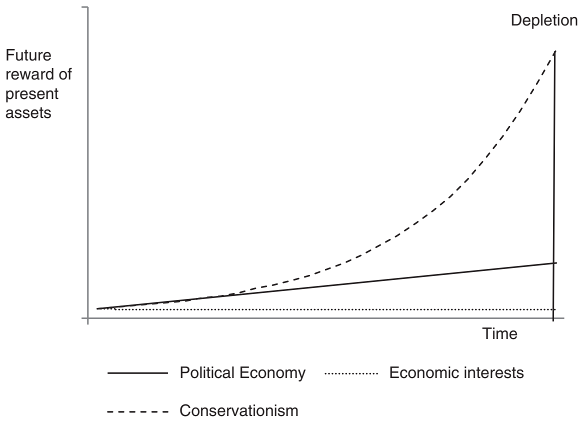


Fig. 6.2 Three perspectives of time discounting regarding fossil fuel depletion

Finally, there is the stance taken by oil explorers, companies, and others involved in the trade and described in the caption to Fig. 6.2 as "economic interests." Altogether, this group had to face the discourse of impending shortage and contend with the sheer inevitability of "D-day." However, it did seem acceptable to believe that future additions to known reserves might extend the depletion gap to unimaginable levels and behave as if "D-day" was completely out of both sight and mind. To underpin this position, they claimed that the present value of rewards should remain unchanged into the future.

The noteworthy point behind the different valuations of the future is that all streams of thought shared the same belief toward the forthcoming time in which the last available non-renewable resource would be spent at the very peak of production. Fossil fuels were doomed to disappear precisely at the moment when they were most needed: peak meant catastrophe.

It took some decades before the refutation of this thorny perspective emerged. Even nowadays, the idea that total exhaustion coincides with the moment of maximum production sometimes looms in public comments. However, in as early as 1956, an American geophysicist working at the Shell research laboratory, in Houston, made an important contribution that shattered these preconceptions. His name was King Hubbert. According to Hubbert, peak production does not signal the exhaustion of recoverable oil, but rather the reaching of the halfway point on the way to global exhaustion. On the historical time line, the peak indicates that humanity has extracted half of the oil that will ever be produced. Thanks to this viewpoint, the whole debate shifted its course from "how long will oil last?"—a question couched in reserves-to-production ratios, to "when does the peak of oil production occur?"—a question grounded in the estimate of the resources still undiscovered. Henceforth, the size of the unknown offset the accuracy of that already known.

The basic assumptions were disarmingly simple: "in the production of any resource of fixed magnitude, the production rate must begin at zero, and then after passing through one or several maxima, it must decline again to zero" (Hubbert

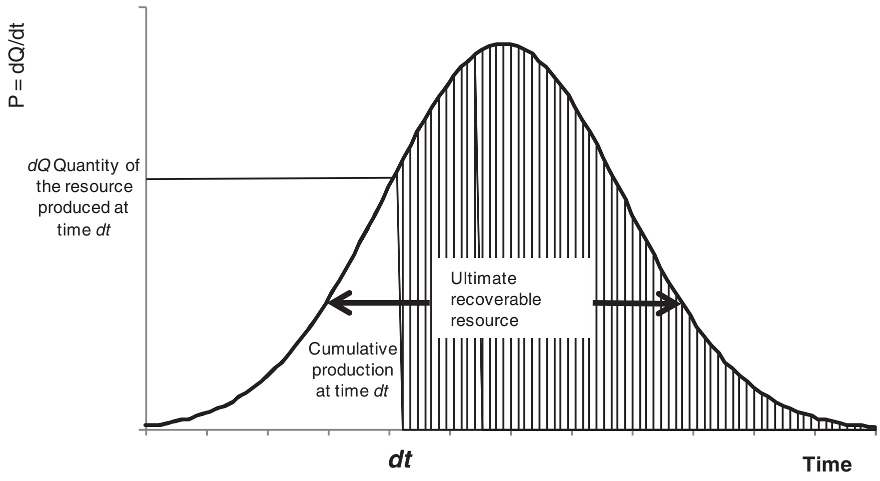


Fig. 6.3 Ultimate crude oil production according to Hubbert's logistic curve (Hubbert 1956: 22)

1956: 9). The challenge then becomes the representation of a growth curve connecting the two zeros such that the area under the curve might equal the estimate of the amount of total recoverable oil. Hubbert decided to take as his ordinate $P = \mathrm{d}Q/\mathrm{d}t$, in which d$Q$ is the quantity of the resource produced in time dt, plotted against time in the abscissa. With this solution, the area under the curve at any time dt comes to represent the cumulative production up to that point (Fig. 6.3). The preference for cumulative data eased the next step of fine-tuning with Hubbert, suggesting that production over time should follow a logistic curve, and thus, yearly production should also follow the first derivative of the logistic, the bell-shaped curve. As shown in Fig. 6.3, the area under the curve corresponds to the resource ultimately recoverable.

What began at the crossroad of intuition and mathematics ended up as a blueprint to model and forecast oil depletion. The fit of the logistic curve produced from cumulative data allowed two major inferences from the process of growth in resources:

- First, the midpoint in the curve stands simultaneously for maximum production and the depletion of one half of the oil recoverable, consequently justifying the post-peak expectation of a decline in oil extraction associated with the beginning of the downward slope toward depletion.
- Second, the principle of symmetry over time, in which the production rate tends to increase exponentially during the first phase of development and is mirrored by the exponential decrease after the peak. On the microlevel, the fitness of the symmetrical form relies on the trend found in individual production fields, whose rising production stems from the high-pressure differentials of the reservoir, and whose decline is equally prompted by exponential pressure loss frequently mingled with the breakthrough of water. On the aggregated level, the

> fitness of the symmetrical form relies on the high probability of hitting large oilfields and speeding up the pace of discovery during initial phases of exploration, as opposed to the decreasing likelihood of such endeavors as exploration advances and the size of the ultimate recoverable resource shrinks. Symmetry results therefore from the differential probabilities of discovering oil, which determine production levels with a constant time lag (10–12 years in the USA) (Doel 1989: 61).

Resorting to the functional form of the bell-shaped curve, Hubbert delivered some bad news and some mitigating evidence to his contemporaries: bad, because the single-peak, anticipated a turning point in the era of fossil fuel abundance, with a potential negative impact upon economic growth; mitigating, because the symmetry around the peak assured a relatively smooth transition to the period of scarcity and because the post-peak decline of fossil fuels could be overcome by a stronger commitment toward nuclear energy. Having recently discovered the magnitude of the energy that could be obtained from fission and the relative abundance of uranium and thorium reserves, Hubbert deemed it appropriate to complement the fossil fuel scenario with a convenient alternative (Doel 1989).

Ultimately, what came to the fore and was spotlighted by the press and energy experts was the central concept of peak along with the prediction that the US oil industry might shortly enter its downward slope. Unsurprisingly, the discovery got a very cold reception, to say the least and was criticized both by executives and by government officials. The turnaround took place in 1970, when the US production of crude oil and natural gas liquids started to fall, confirming one of Hubbert's peak forecasts. Although the controversy did not come to a halt, the geophysicist was at that juncture turned into a kind of conservationist folk hero and praised even by those who had previously opposed him. The oil industry had begun to side with the peak thesis and President Carter's election, in 1977, brought conservationists into the very heart of federal government (Bowden 1985; Deffeyes 2006; Hemmingsen 2010).

From its very inception, the breakthrough introduced by Hubbert stems from the representation of $P = \mathrm{d}Q/\mathrm{d}t$ in the graph ordinate. This idea enabled the mapping of the ratio of cumulative production to resource size through recourse to $\mathrm{d}Q$ expressing the quantity of the already produced resources. Hence, a key element in the overall approach turned out to be the estimation of the resource size or ultimate recoverable resource (hereafter URR). The point proved particularly sensitive since URR, or the area under the production curve, includes the oil-yet-to-be-found, precisely the unknown portion of the original endowment of conventional oil. This portion has to be estimated through geological techniques of inference or through the extrapolation of production or discovery trends, when exploration is well advanced. Given the contingencies present in such assessments, one can more easily understand what was previously stated about the prevalence of the "size of the unknown" over other aspects of analysis and the revolution it provoked in the depletion debate.

Based on a track record of outstanding geological and geophysical expertise, Hubbert proposed, in 1956, an original endowment of global conventional oil

(URR) of 1,250 billion barrels (1,250 $\times$ 10^9). Later, other geologists made an upward revision of this ceiling and by the 1970s several independent estimates clustered around the value of 2,000 billion barrels or 2,000 Gb (the abbreviation Gb—Gigabarrels henceforth designates billion barrels). Other things being equal, the greater the proposed size for the URR, the more the center of the logistic curve would be displaced toward the right, pushing further ahead the estimate of the turning point for global peak oil. The current situation can be summarized into two broad groups of forecasters: On the one hand, those forecasting an URR in the range of 1,900–2,800 Gb, implying peak oil occurring between 2009 and 2020; and on the other hand, those who estimate an URR clustered between 2,900 and 3,900 Gb that knocks the peak oil timeframe back to between 2021 and 2030. It is worth remarking that the latter group of more optimistic scenarios includes the major oil business institutions of reference, for example the International Energy Agency (IEA) of the OECD, the Energy Information Administration (EIA) of the US Department of Energy, and research departments at oil companies such as Shell, Exxon, and Statoil (Tsoskounoglou et al. 2008; Jakobsson et al. 2009; Sorrell et al. 2010). Another important conclusion is that should the lower "pessimistic scenario" estimates hold true, and then, the peak has already been reached, or is rapidly approaching. The question that logically surfaces is whether or not we have already reached the peak without noticing it? From a theoretical viewpoint, the potential answer is yes: yes, we could do. Production of conventional oil has recently gone flat, with a drop in 2009, and such a scenario matches the concept of a multiyear plateau in which production fluctuates by a few percentage points before definitively entering into decline. Whatever the case, the global peak oil phenomenon is only empirically susceptible to ex-post confirmation.

6.4.2 Peak Oil Critics

To recapitulate, world oil production is expected to peak between 2010 and 2030 with a corresponding prospective URR in the region of 2,250–3,400 GT. Thus far, some of these estimates have included conventional oil and oil from unconventional sources, a category that includes tar sands, extra-heavy oil, shale gas, and shale gas, gas-to-liquids, coal-to-liquids, chemical additives, and crude exploration in water depths of below 2,000 m. Owing to the take-off of these crude-like resources, the overall energy balance has become much more uncertain. Stirred by the comparative advantages brought about by soaring petroleum prices and by the pace of technological innovations that cut costs brusquely (particularly in shale oil extraction), unconventional oil attained its breakeven price in recent years, and is expected to remain competitive, provided that oil prices stay above the plateau of \$60 and \$100 US/bbl, respectively. Many of the alternatives are nonetheless expensive to produce, require huge new investments, present high environmental costs, and a low energy payback. In other words, besides the environmental air pollution, large water consumption, massive earth moving, and ecosystem disturbance they require large amounts of energy-inputs per unit of output generated. It is expected that unconventional oil might reach around 8.5 % of the world oil supply in 2030, meaning that its development, even in the best-case scenarios (crash programs), is not enough to mitigate the impeding "all-oil" peak (Gagnon 2008; Alekletta et al. 2010; Mohr and Evans 2010; Grushevenko and Grushevenko 2012).

Description of unconventional oil in 1916 (Bacon and Hamor 1916: 37). Outcrops of sands. Outcrops of sands impregnated with tar or bitumen are not common, but they exist in several parts of the world. Perhaps the best occurrences of this kind are the "tar sands" of the Athabasca and other rivers in northern Alberta in Canada, where the outcrop of the Dakota formation is impregnated with tar for scores of miles along the main rivers, leading to the supposition that oil and gas will be found in great quantity in those portions of the same sands, which are under cover and have the requisite geological structure.

Bituminous Dikes. As to dikes of asphalt and other bitumens, the relationship is not so apparent, since these substances are frequently solidified, and some varieties of them are as hard and compact as coal. This has been so deceiving to the public that in reference to one case at least, namely the Albert Mines in Albert County, New Brunswick, the Courts have decided that the material (albertite) shall be legally known as coal. In reality, however, it is an entirely different bitumen. In other localities, such as Mexico and California, oil is seen to ooze out of the ground, together with the asphalt, thus proving a close relationship.

Bituminous Lakes. Asphalt and related bitumens occur elsewhere at the surface in the form of lakes. In Mexico are a great number of seepages, some of small size, but many of them covering thousands of square feet, in which the asphalt has been seen emerging from the ground and taking the form of small lakes. The best-known example of a lake of asphalt is the Pitch Lake of Trinidad, hundreds of acres in area, frequently described in the literature.

Asphaltic deposits are direct evidence that oil does exist or has existed in the vicinity, for these substances appear to be the desiccated or oxygenated residues of heavy oils which have oozed out of the surface in past ages. In many cases, it is possible for a geologist to locate the field from which the asphalt has escaped. Consequently, it is important not to neglect the surface evidences, but in studying them, one must remember that certain structural relations hold true and consequently that oil pools will very seldom be found directly underneath the points of emergence of the substances.

From these considerations, it would seem to follow that the unproduced fraction of petroleum plus the bulk reserves of unconventional oil are the single determinant factor that might affect the course toward depletion. And herein lays the rub of much contention between the geological view of peak oil and the economic analysis of natural resource exhaustion. The assertion that oil production rates depends linearly on the fraction of the oil that remains to be produced and that every other factor like price or technological innovation, matters, but "it just doesn't matter very much" (Deffeyes 2006: 41), has left most economists uneasy. The most vocal opposition came from neoclassical economics for which peak oil is a non-issue, a speculative leitmotif raised out of the blue since there is nothing inevitable about any particular schedule of global oil exhaustion. Moreover, policies such as conservation, avoidance of waste, consumer restraint, and investment in new alternatives are already embedded in the normal functioning of decentralised markets, through price signals and reallocation of incentives, saving resources, and, above all, widening the gap between peak demand and depletion.

For this current of thought, price is the key market mechanism to offset depletion risks. Long before the first signs of scarcity were noticed a "scarcity rend" would burden the price of existing stocks, prompting lower consumption and an aging resource base would lead to a "degradation of extraction costs," raising fossil fuel prices again to new heights (Hotelling 1931; Solow and Wan 1976). As oil becomes dearer and the costs of production mount, the market will respond by using the resource more selectively and more proficiently, through enhancing the fuel efficiency of engines and other devices while simultaneously stepping up the substitution of traditional technologies by new "backstop technologies" that draw upon virtually inexhaustible resource bases (Nordhaus 1973) and by flexible switchovers to other fossil fuels. Once these alternatives reach the maturity stage, they can provide a ceiling for the price of oil and create a feedback loop between competing technologies. Recent examples of market response comprise the innovation in high compression ratio combustion engines and consumer changeover to diesel autos (efficiency); the innovation of hydrogen fuel cells and electrical batteries (backstop substitution); the boom in natural gas and shale gas (inter-fuel substitution). Hence, oil conservation is already embedded in the future value of present assets by means of a scarcity rent, cost degradation, technological innovation, backstop technology, or inter-fuel substitution. However, as the above examples suggest, none of the quoted technologies or resources is a perfect substitute for oil. In practice, any replacements are expensive, difficult to achieve, require very long and uncertain lead times, and incur negative externalities and network costs. A glance at history shows that human inventiveness and flexible competitive markets constitute the best antidote to the emerging problem of resource scarcity. Even so, the devil is in the details and, furthermore, in the timing of those details. A study by Hirsch et al. (2005) shows that given the state of the art in technology and costs, the switchover to new resources may take longer than expected.

If economists were asked about the mechanism that drives the peak of oil production, they would certainly pinpoint higher costs, higher prices, and the combined effect in destroying demand and replacing supply. Overall, this stream is also utterly critical to the forecasting presented by Hubbert's followers, stressing that URR assessments do not depict tangible fixed stocks but rather reflect advances in knowledge and technology. These advances expand the amount of physically “recoverable” resources. The rising trend in URR estimates over the last 60 years is therefore bound to continue adding new resources to the extreme right of the logistic curve. Likewise, they also expect a large margin of progression in Middle East reserves, which are underestimated owing to little greenfield exploration undertaken since governments took control of the oil industry and its resource base in the 1970s. More investment and improved technology are likely to disclose new findings and extra oil in the region (Maugeri 2004; Clarke 2007). For these reasons, not only is there no scarcity in sight but also any future scarcity could be efficiently overcome.

6.5 Epilogue

The tricky dimension to the last phrase is that it is half true and half false: Geological and mathematical analysis has provided robust evidence that oil scarcity is in sight—robust in the sense that uncertainty in the estimates does not affect the major conclusions; economic analysis, on the other hand, has shown theoretically and through empirical studies that scarcity could be overcome. The wise advice is therefore to turn to doomsters for diagnosis but call up the cornucopians for solutions.

Efficient as competitive markets may be, they do not provide the answer to all geophysical issues. Investment and research can cope with the invention of substituting technologies, with cheaper and cleaner recovery of unconventional oil, and with improved and enhanced oil recovery from maturing fields. What investment and research cannot accomplish, is the inversion in the downward trend in the discoveries of petroleum, the fall in the average size of the fields discovered, the beginning of a steep decline in maturing and aging oilfields, and the decrease in production in two-thirds of oil-producing countries. In all these indicators, the physical amount of the recoverable resource sets a first-order constraint on the productive activities. To bet all trump cards on second-order factors such as investment and technical ingenuity would be, at best, dogged voluntarism.

In the upcoming world, where the largest part of the original oil in place has already been discovered and recovered, some nations will fare better (Venezuela, Canada, Saudi Arabia, Brazil, Russia, Kazakhstan), while others will fare worse (India, Japan, China, France) (Lutz et al. 2012).

References

Arnold R (1923) Two decades of petroleum geology 1903–1922. Bull Am Assoc Pet Geol 7(6):603–624

Alekletta K et al (2010) The peak of the oil age—analyzing the world oil production reference scenario in world energy outlook 2008. Energy Policy 38(2010):1398–1414

Bacon WF, Hamor WA (1916) The American petroleum industry, vol 1. McGraw-Hill, New York

Beal CH (1919) The decline and ultimate production of oil wells, with notes on the valuation of oil properties. U. S. Bureau Mines—Government Printing office, Washington
Bentley RW, Mannan SA, Wheeler SJ (2007) Assessing the date of the global oil peak: the need to use 2P reserves. Energy Policy 35(12):6364–6382
Bentley RW (2002) Global oil & gas depletion: an overview. Energy Policy 30(3):189–205
Bignell LG (1934) Geophysical prospecting activity in all areas indicates need for additional oil reserves. Oil Gas J 32:52
Bowden G (1985) The social construction of validity in estimates of United States crude-oil reserves. Soc Stud Sci 15(2):207–240
Bowker GC (1994) Science on the run information management and industrial geophysics at Schlumberger, 1920–1940. The MIT Press, Cambridge
Brown NC (1919) Forest products and their manufacture and use. Wiley, New York
Clark JG (1987) Energy and the Federal Government. Fossil fuel policies 1900–1946. University of Illinois Press, Urbana
Clarke D (2007) The battle for barrels: peak oil myths and world oil futures. Profile Books, London
Cooper JM (1990) Pivotal decades: the United States 1900–1920. Norton, New York
Day DT (1909) The petroleum resources of the United States. In: USGG (ed) United States geological survey bulletin 394. Government Printing office, Washington, pp 30–50
Decker CE (1920) Studies in minor folds. University of Chicago Press, Chicago
Deffeyes KS (2006) Beyond Oil. The view from Hubbert's peak. Hill and Wang, New York
Dennis MA (1985) Drilling for dollars: the making of US petroleum reserve estimates, 1921–25. Soc Stud Sci 15(2):241–265
Doel RE. (1989) M. King Hubbert interview, Session VII, 27 Jan 1989. Marion King Hubbert: transcript of a recorded Interview, Bethesda, Maryland, 4 Jan–6 Feb 1989. Available via http://www.oilcrisis.com/hubbert/aip/aip_vii.htm. Accessed 12 Feb2013
Downey M (2009) Oil 101. Wooden Table Press, New York
Dunham WM (1913) Foresees famine in light crude oil. The Natl Petrol News 3:1–2
Elster J (2007) Explaining social behavior: more nuts and bolts for the social sciences. Cambridge University Press, Cambridge
Emmons HE (1921) Geology of petroleum. McGraw-Hill, New-York/London
Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: a critical review. J Econ Lit 40(2):351–401
Forbes RJ, O'Beirne DR (1957) The technical development of Royal Dutch/Shell 1890–1940. E.J. Brill, Leiden
Gagnon L (2008) Civilisation and energy payback. Energy Policy 36(9):3317–3322
Garfias VR, Whetsel RV (1936) Proven oil reserves. Trans Am Inst Mining Metall Eng 118:211–214
Grushevenko E, Grushevenko D (2012) Unconventional oil potential tends to change the world oil market. Energy Sci Technol 4(1):68–74
Hager D (1923) Hager on geology and the geologists. Oil Gas J 21:91–93
Hays SP (1959) Conservation and the gospel of efficiency: the progressive conservation movement, 1890–1920. Harvard University Press, Cambridge
Hemmingsen E (2010) At the base of Hubbert's Peak: grounding the debate on petroleum scarcity. Geoforum 41(4):531–540
Hirsch RL, Bezdec R, Wendling R (2005) Peaking of world oil production: impacts, mitigation and risk management. DOE report, Feb 2005. Available via http://www.netl.doe.gov/publications/others/pdf/Oil_Peaking_NETL.pdf, Accessed 2 April 2013
Hotelling H (1931) The economics of exhaustible resources. J Polit Econ 39:137–175
Horlacher JL (1929) A year in the oil fields. The Kentucky Kernel Press, Lexington
Horowitz JK, Carson RT (1990) Discounting statistical lives. J Risk Uncertainty 3(4):403–413
Hubbert MK (1956) Nuclear energy and the fossil fuels. Shell Development Company, Houston
Hubbert MK (1966) History of petroleum geology and its bearing upon present and future exploration. Am Assoc Pet Geol Bull 50(12):2504–2518

Jakobsson K, Soderbergh B, Hook M, Aleklet K (2009) How reasonable are oil production scenarios from public agencies? Energy Policy 37(11):4809–4818

Johnson RH, Huntley LG (1916) Principles of oil and gas production. Wiley, New York

Laherrere, J (2001) Estimates of oil reserves. In: Paper presented at the EMF/IEA/IEW meeting IIASA, Laxenburg, Austria—June 19, 2001. Available via http://www.oilcrisis.com/laherrere/. Accessed 8 July 2012

Lutz C, Lehr U, Wiebe KS (2012) Economic effects of peak oil. Energy Policy 48:829–834

Madureira NL (2012) The anxiety of abundance: Stanley Jevons and coal scarcity in the nineteenth century. Environ Hist 18(3):395–421

Maugeri L (2004) Oil: never cry wolf—why the petroleum age is far from over. Science 304(21):1114–1115

Maugeri L (2006) The age of oil: the mythology, history, and future of the world's most controversial resource. The Lyons Press, Guilford

McLaughin GL (1939) The economic significance of oil and gas. In: National Resources Committee (ed) Energy resources and national policy. Government Printing Office, Washington, p 123–236

Miller C (2009) The once and future forest service: land-management policies and politics in contemporary America. J Policy Hist 21(1):89–104

Miller HC, Lindsly BE (1934) Report on petroleum production and development. In: USGS (ed) Hearings before a subcommittee of the Committee on Interstate and Foreign Commerce. Government Printing Office, Washington, Part 2, p 1087–1222

Mohr SH, Evans GM (2010) Long term prediction of unconventional oil production. Energy Policy 38(1):265–276

Nordhaus W (1973) The allocation of energy resources. Brookings Pap Econ Act 4(3):529–576

Nordhauser N (1979) The quest for stability: domestic oil regulation 1917–1935. Garland Publishing Co., New York

Olien DD, Olien RM (1993) Running out of oil: discourse and public policy, 1909–1929. Bus Econ Hist 22(2):36–66

Owen EW (1975) Trek of the oil finders: a history of exploration for petroleum. American Association of Petroleum Geologists, Tulsa

Penick JL (1968) Progressive politics and conservation: the Ballinger-Pinchot affair. University of Chicago Press, Chicago

Requa ML (1918) Methods of Valuing Oil Lands. Bulletin of the American Institute of Mining Engineers (134): 409-428-

Petty OS (n.d.) Oil exploration. Handbook of Texas. Online published by the Texas State Historical Association. Available from http://www.tshaonline.org/handbook/online/articles/doo15. Accessed 12 Sept 2011

Robertson H (2000) A historic correspondence regarding the introduction of the torsion balance to the United States. Lead Edge 19(6):652–654

Roosevelt T (1909) Special message from the President of the United States transmitting a report of the National Conservation commission. In: Gannett H (ed) Report of the National Conservation Commission. Government Printing Office, Washington, p 1–9

Schulman BJ (2005) Governing nature, nurturing government: resource management and the development of the American State: 1900–1912. J Policy Hist 17(4):375–403

Schurr SH, Netschert BC (1977) Energy in the American economy 1850–1975: an economic study of its history and prospects, 2nd edn. Greenwood Press, Westport

Shulman PA (2003) Science can never demobilize: the United States navy and the petroleum geology. Hist Technol 19(4):365–395

Smith GO (1920) Opportunity for the oil geologists. Oil Gas J 18:36, 74

Solow R, Wan F (1976) Extraction costs in the theory of exhaustible resources. Bell J Econ 7:359–370

Sorrell S et al (2010) Global oil depletion: a review of the evidence. Energy Policy 38(9):5290–5295

Steen HK (2001) The conservation diaries of Gifford Pinchot. Forest History Society, Durham

SPE/WPC (2001) Guidelines for the evaluation of petroleum reserves and resources. A supplement to the SPE/WPC petroleum reserves definitions and the SPE/WPC/AAPG petroleum resources definitions. Available via http://www.spe.org/web/store/index.shtml. Accessed 4 Feb 2013

Tsoskounoglou M et al (2008) The end of cheap oil: current status and prospects. Energy Policy 36(10):3797–3806

USGS (1934) Résumé of geology and occurrence of petroleum in the United States, In: USGS (ed.) Hearings before a subcommittee of the Committee on Interstate and Foreign Commerce. Government Printing Office, Washington, Part 1, p 884–907

Villarello JD (1908) Algunas regiones petroliferas de Mexico. Imprenta y Fototipia de la Secretaria de Fomento, Ciudad de Mexico

White D (1919) Mr White sees danger of exhaustion. Oil Gas J 17:8–9

Williams N (1928) Geophysics big factor on Gulf Coast. Oil Gas J 26:35, 88

Williamson HF et al (1968) The American petroleum industry. The age of energy 1899–1959. Northwestern University Press, Evanston

Chapter 7
Marginal Cost Pricing

Abstract This chapter describes the debate that evolved around the diffusion of electrical-rate systems in twentieth-century Europe and the clash that ensued between the competitive, promotional, and cost-based approaches. The major question addressed is the historical origin of marginal cost pricing. We show how the volatility of the costs characterizing hydroelectric production made this particular technology very sensitive to a cost-based approach toward pricing and to a seasonal and time-of-day perspective on rate systems. Subsequently, and after World War II, the concept of marginal cost came to be interpreted in the context of variations in demand by time of both day and year, the source of peak-load problems, and practically translated into a version of time-of-day pricing.

7.1 Rate Systems and Pricing

In late nineteenth century, several economists developed a theory in which the fundamental changes in the world did not occur through the displacement of averages but through adjustments at the margins. Society was conceptualized as an aggregation of independent, but interacting actions in which the value of things was determined by the last unit bought or the cost of producing an extra unit. The key concepts of marginal value and marginal cost embodied this vision.

The discovery of changes at the margins firstly proved useful as a principle for decision-making and optimization since they provided a yardstick for firms to set their level of production at the point where marginal cost equals the marginal revenue derived from sales. In ordinary terms, this meant that an enterprise should continue to manufacture through to the last unit that still yields a profit. Moreover, the theory also pinpointed the equilibrium wrought by the valuation agreed between the last buyer and the seller. Owing to the fact that in this particular situation, the producer may figure out exactly what the consumer is seeking and manufacture that which is needed—and no more or no less—the equilibrium is deemed to be an efficient allocation of resources.

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_7,

Overall, in decentralized market, economies where prices serve as communicational devices embedded in the flow of products and services, the rule of efficient pricing requires the marginal unit of the good be sold at marginal cost. However, what if transactions are not carried out according to decentralized transactions? What if there were no instantaneous exchanges between buyers and sellers? Is it even feasible to refer to marginal cost pricing under these circumstances? It is worth recalling that many transactions within the energy sector take place according to long-term rate systems rather than by any spot exchanges. In this new context, the bargaining closeness disappears; information on prices gets disconnected from the decentralized flow of transactions; the valuation of goods and services becomes conditional upon the categories established for the rates. In other words, this type of pricing scheme is characterized by an institutional background that shelters the transactions in long-term contracts so that information is removed from the edgy world of instant transactions. Moreover, there is discrimination according to various consumption categories and hence understanding the context of the value of goods and services proves sensitive, variable, and subject to learning-by-using.

Unlike the “flat”-price dimension of most consumer goods, the rate is bound to the complexity of a particular classificatory scheme: this may differentiate between social positions (e.g., the elderly, students, or the poor), activities (retail-shopping, large industry, small farming) or end usages (consumers in congested and non-congested periods). However, just what is the rationale behind such divisions? Why should someone pay according to predetermined categories of prices rather than simply pay for the quantity of goods and services demanded? After all, the consumer is buying the same product or service over and over again. Drawing upon this line of inquiry, historical research has found no single justification for the adoption of rate systems but instead actually three strands of arguments which have evolved over time: the competitive, the promotional, and the cost-based. This chapter describes the historical roots of marginal cost pricing in Europe taking as its guiding thread the whole-hearted debate around electricity rate systems.

7.2 Classification of Consumption or Classification of Consumers

7.2.1 The Competitive, Promotional, and Cost-Based Approach

7.2.1.1 Competitive Rate Systems

From the competitive perspective, a rate system is a tool to dissuade potential contenders. This means that the market environment still retains competitive traits even if there is an enterprise endorsed with some legal monopoly. The competition arises from new entrants that choose to invest in self-production or from enterprises that provide substitute goods. To keep the advantage over these contenders and to switch some of them back to clients, the central supplier must offer attractive and clear-cut prices. Neufeld (1987) noted the dissuasive role of rate systems in tracking the costs of operating isolated electrical power plants through the use of pricing schemes competing with the factors of self-generation—“standing” and “running” costs. On examining the electricity supply industry in the USA at the dawn of the twentieth century, Yakubovich et al. (2005) also advanced the explanation of pricing policies based upon the maximization of revenues in a period of fierce competition between the electrical utilities and competitors in gas lighting. This strategy resulted in the preemptive capture of more customers and more revenues.

7.2.1.2 Promotional Rate Systems

From the promotional view, rate systems are customized marketing devices to enhance individual consumption. Prices are generally set according to consumption levels with incremental discounts whenever the user reaches an upper plateau. In this respect, it is no accident that these pricing schemes often appear in conjunction with popular advertising campaigns. The overall purpose is to increase the supplier’s income through an increase in the volume of sales. Yet this goal is sometimes intermingled with others, namely the attempt to convince the users to shift their consumption from the system peak. In this more sophisticated version of promotional rates, lower prices are offered mainly in non-congested periods. Leslie Hannah (1979) discovered that most of the rate systems adopted by the English electrical utilities in the 1930s sought precisely to increase private consumption in the off-peak periods of very low system demand. Even though the practical achievements of these proposals brought unexpected consequences, the existence of variations by time of day and year in consumption combined with the non-storability of goods and services on the supply side provides a rational justification for using prices as harbingers for change.

7.2.1.3 Cost-Based Rate Systems

The cost-based view makes all clients pay for the real cost of their consumption without further attempting to sway their options: price is therefore used as an up-to-date signal rather than a hint for future adjustments. The existence of differential rates is justified by the asymmetric costs of serving customers which, in turn, results from the existence of variations by period in demand. This is a particular characteristic of many public utilities (electricity, transport services, telecommunications) greatly enhanced by the fact that the major operating costs are the costs of capital equipment. Since the goods and services must be produced at the same instant as they are consumed, the capacity of capital equipment comes to be determined by the size of demand at the peak load (or congested periods). The fundamental lesson drawn from the cost-based perspective pinpoints the differential between a client that uses the system during its peaks, thus influencing the quantity of equipment capital required, and another client whose consumption occurs in off-peak periods and therefore has little effect upon capital. Revealingly, the cost-based outlook proved to be historically identified with time-of-day rates and, subsequently, with marginal cost pricing. In a recent article, Martin Chick (2002) showed how the "Tarif Vert" (Green Rate) was introduced for the first time in 1956 by the nationalized enterprise "Électricité de France" (EDF), which involved the principle of price discrimination through time-of-day blocks (see also Hausman and Neufeld 1984a, b).

The three perspectives express different viewpoints of the time effects of managerial action: competitive rates involve the ex-ante recouping of costs (Hausman and Neufeld 1984b); promotional rates convey ex-post signals for consumption; and cost-based rates convey up-to-date signals for consumer choice. Concerns vary therefore between the medium-term consequences of pricing schemes, long-term consequences, and short-term consequences. What is more, each of these options entails distinct domains of information gathering: whereas in the first and third alternatives (competitive and cost based), the boundaries between price categories may be designed through an inward examination of accounting books; in the case of the second alternative (promotional rates), only an outward study that encompasses the consumers' needs might solve the dilemma of setting appropriate categories and appropriate prices. The shift toward rate systems that are grounded on prospective and behaviorial assumptions incurs the immediate effect of raising the costs of information gathering and data monitoring.

Finally, the measurement costs may also change whenever there is a need to replace the basic system of metering, closely associated with the competitive and the promotional rate systems (a meter based in electromechanical induction which counts the revolutions of an aluminum disk made to rotate at a speed proportional to the power), with more expensive multiple tariff meters so as to record the energy supplied at different times of the day (cost-based rate systems) (Lanphier 1925; Hausman and Neufeld 1984a; Gooday 2004: 219–262).

7.2.2 The Demand-Charge System

According to Hausman and Neufeld (1984b), the debate over pricing schemes gained momentum around 1910, a time when the American power industry was facing the prospect of monopoly regulation, "and was anxious to present a united front on rate structures so as to minimize possible conflict with regulators over that issue." The agreement that followed was endorsed by the special committee of the National Electric Light Association, a forum established by utility trade groups. With the purpose of avoiding the interference of regulatory commissions or appeals to courts, the committee moved speedily to uphold the demand-charge principle, thus providing a consistent position for the industry in subsequent years.

Formerly advocated by Arthur Wright, an engineer from Brighton, the demand-charge or Wright system—as it became known in the decades that followed—is a forerunner of two-part rates, in which the customer is charged not only for his maximum power consumption but also for the total quantity of energy supplied. The final bill therefore replicates the categories of costs relevant to the industry, setting a price structure that recoups both "standing" costs and "running" costs (fixed and variable costs). In a world where there was still little diversity in electrical applications, the entry of new customers may incrementally increase the demand in peak time and contribute to further expenses in equipment, machinery, and cables. Consequently, the demand-charge principle requires customers to pay both for the goods received (kWh) and also for the individual cost imposed over the system.

Since the consumer's bill is based on the size of individual maximum power consumption and not on its level of consumption during system peak, there are few incentives to shift private usages. The Wright system assumes that individual peaks have the tendency to occur simultaneously, so that each new client brings a linear increment to the overall fixed costs. Consumption is purely additive. Inasmuch as this assumption holds, the industry has few incentives to sharpen customer's classifications. Because the priority is to recoup costs under the prevalence of homogeneous demand, the relevant categories used in business accounting seem adequate to outline customers' profiles. Also, accounting categories provide a good dissuasive pricing toward self-generation, a feature of the above-mentioned competitive rate systems.

The key question reverts then to the diversity that a central power-station faces. What happens when the industry matures and the service is provided to industrial clients, railway lines, domestic customers, urban services, shopkeepers, and other activities? The answer is that in a more varied system (different types of usages and different timetables), the knot between individual peaks and system peaks is cut. Clients will continue to pay for fixed costs imposed upon the system, even when their maximum power is used outside the aggregate peak periods. This means that under a diversified load and non-additive consumption, the rationale for demand-charge collapses. Not by accident authors like Crew and Kleindorfer (1986) have described the adoption of the Wright rate system as a second best or imperfect form of peak-load pricing.

In any case, with the demand-charge principle, the consumer is still viewed as a black box, an entity whose individuality, behavior, and wishes are considered as irrelevant. The only active attribute that the rate system recognizes is the consumer's capacity to switch to alternatives source of supply whenever disadvantageous prices are offered. In practical terms, this results in a perception of the consumer through crude traits, basically given by the "exit" option and the "permanence" penchant, i.e., he can continue to be a customer, change to another energy carrier or change to self-production.

By the dawn of the twentieth century, the Wright system started spreading over wide areas. With this diffusion, a wave of social and political criticism ensued, as to whether the rate should be reformed and adapted, or upheld in its original form. Most complaints originated from pressure groups in towns that claimed an upper hand in prices and rejected the "tariff" (the English expression for "rates," henceforth used as a synonym). It was particularly in Europe that this tendency took shape, bringing the consumer's viewpoint to the forefront. The demand-charge rate reached the electrical utilities of England, the Netherlands, Sweden, and other countries, primarily as a surrogate for the use of undifferentiated flat-rate systems. Consumers voiced their opposition to what appeared as an arbitrary "fee" embodied in the double payment of the Wright system. In most places, it was the shopkeeper in the center of town who most disliked the rate, because he tended to use his lighting for only an hour or less per day in the winter evenings, but nevertheless was obliged to pay a sort of "tax" for the recovery of the standing costs of production. Even when he consumed only a very small quantity of electricity, the fixed component of two-part rates raised disproportionately the average price to be paid.

At this juncture, the institutional environment filtered down the public voices of discontent. The pressure for reviewing established pricing schemes appeared all the more urgent when public authorities endorsed the complaints, or when the business of electricity generation was directly managed and owned by the municipalities. In the first decades of the twentieth century, both conditions held, particularly in Great Britain, which had the largest public municipal sector in electricity generation (together with Norway) (Millward 2005), and also an influential body of electricity commissioners directly responsible for the overseeing of local developments and for promoting "progressive" tariffs (Hannah 1979; First Report 1950). This sensitivity of public powers toward public opinion, combined with the robust implantation of municipal enterprises, accelerate the attending of constituency claims, and the drift away from the demand-charge principle.

7.2.3 Downgrading of Peak-Load Concerns

The new rate systems that mushroomed thereafter used the Wright system's double structure (fixed plus variable payments), but transformed its content. Methods simpler than the maximum demand received clear preference. From among all the variations that surfaced, one important development occurred in Norwich, England, where the authorities took advantage of the existence of the universal tax levied on the rateable value of the houses, in order to forge a new standard. Publicans, drapers, and grocers formed the core of the retail sector of this industrial town and were the net beneficiaries of the change (Doyle 1995). Pioneered in 1908, the Norwich rate system popularized the idea of "two-part tariffs," in which the consumer pays the first x units of electricity at a high price, x being determined approximately by the rateable value of the house, and the rest at a low price. The rateable value of the house (or alternatively the floor area or the number of rooms) consequently started being used as a proxy indicator for the maximum demand. While this part of the rate served to recover capital and service costs, the remaining energy costs could be made sufficiently low to induce the expansion of demand. Sustained by a push in technological innovation and by the advance in economies of scale, two-part tariffs fed back the productivity gains to consumers. By the mid-1930s, about one-third of British users had opted for this pricing scheme, and at the very beginning of the sector's nationalization (1947), three-quarters of the electricity sold to domestic and commercial customers was provided through two-part tariffs (Hannah 1979: 199).

From the common people's perspective, the Norwich system established more tangible and comprehensible principles. From the utilities perspective, it brought a reduction in information and measurement costs: instead of periodic assessments of the maximum power demanded by each user, the fixed component of the rate could now be automatically estimated by looking at the rateable value of the house. The system took advantage of the previous existence of exhaustive and universalized (i.e., covering both residential and commercial houses) tax collections to reduce metering and administrative costs. Yet, where no such tax records were available, other methods had to be invented. Variations in this rate system soon emerged across Europe, leading to a degenerative recasting of the demand-charge principle. The utmost examples are the "Potsdam Rate," originally adopted in this German town, in which the fixed part is estimated according to the number of rooms of the apartment, and the "Dutch system," in which the fixed part is charged according to the floor area of the dwelling (Siegel and Nissel 1938: 160–161; 259–265).

Rateable value, number of rooms and floor area, all distorted the philosophy of the Wright system. Although the original purpose was to classify patterns of demand, the ultimate result was, in effect, a classification of customers. Under the new pricing schemes, access to property and the magnitude of rental dwellings came to constitute summary attributes upon which the bill was charged. In this manner, personal and social profiles, rather than consumption characteristics, made headway in classificatory schemes. The bill mirrored now the indoor habits of the families with a crude representation of consumer's differences.

Historically, it was assumed that "maximum demand" and "size of the house" were proxy estimates of the same thing. One indicator could be inferred from the other, and this allowed general equivalencies: larger houses require more lighting; more lighting, switched on at the same time, would produce an increase in the demand for maximum power. Although inaccurate, this forecasting method was relatively acceptable insofar as lighting was the dominant usage. But with the spurt of domestic appliances after the First World War (Bowden and Offer 1994; Frost 1993; Froelich and Morel 1949), the premises were irremediably modified. Particularly with the adoption of high-power appliances, like electric cooking and heating, there was less and less reason to suppose that the size of the house was still a good indicator of the demand charge. The idea of indivisible supply of a good (lighting power) was transformed into an array of divisible services (power for whatever activities the customer chose).

The drift from the classification of consumption toward the classification of consumers evolved out of the downgrading of the peak-load problem. Concerns with the recovery of standing charges, to pay for the high costs of machinery and equipment, dropped out as two-part rates penetrated further into the market. The new pricing schemes emphasized not so much the fixed part of the bill, but rather the extension of its variable component, which allowed sales at very promotional prices. A critical observer of this evolution described the goals of the electrical utilities in terms of "selling as much kWh as possible" (Houthakker 1951: 24).

Furthermore, all these developments appeared enfolded in social representations of the consumer as the main reason for entrepreneurial enhancement. Publicists and government officials transposed the image of technological progress into the core of households and announced a new era for mankind: the era of democratization of electrical progress, liberation, leisure time, and modernity. Politically, this ideal was inspired by median classes' aspirations, by compelling images of a "new domesticity" and by a wave of consumer activism (Frank 2001; Frost 1993; Hannah 1979: 201–208). From the perspective of the 1920s or the 1930s, the future looked bright, and the mission of public actors was to facilitate the access to new electrical improvements. Promotional rates therefore became a political bet for the democratisation of citizenship. Also, due to particular economic reasons, the central state progressively stepped up the supervision and regulation of local public utilities. Whether this regulation took the form of price indices—established to move electricity prices automatically up or down as input costs changed, special clauses in concessions (Besançon 2005; Varaschin 2002), or public investments in enterprises that secured price leadership as was the case of Finland, Sweden, and in some sense Great Britain (Myllyntaus 1991; Thue 1992; Kaijser 1995; Chick 1995), State interference allowed a national outlook of rate systems. The lessons drawn from such observations were that rate systems and prices varied enormously from one region to another. Consumers appeared like figures locked to local monopolies and obliged to pay whatever value the bill showed. Injustice was thus equated with "unfair pricing," and perceived through the heterogeneity of the rates. Along with regional asymmetries, price discrimination within the same locality reinforced the idea of arbitrary prices, business bias, and abuse of monopolistic power. In France, for instance, the issue of price solidarity between town and countryside played a key role in the course of collective action and in the institutionalization of consumer representation generally (Poupeau 2001, 2007). Through different streams, the question of pricing was increasingly seen as a political endeavor. Businessmen and managers could no longer disengage from what was going on in society and decide strictly on an accounting basis. Because classifications produced social effects, the building up of rate systems had somehow to classify consumers and determine the charges according to understandable social goals. Under the pressure of political, professional, and social forces, the awareness of levelling-up effects in electrical-rate systems placed the consumer at the center of classifications. While competitive pricing began to wane, promotional rates reached fruition.

7.3 Promotion Through Declining Price-Blocks

7.3.1 Network Expansion and Promotional Rates

We have seen how the principle of apportioning "standing" charges lagged behind the practices of the electrical sector in the twentieth century. Innovations in the domain of rate systems were quickly superseded by the evolution of markets and

Table 7.1 The roots of declining blocks rate systems

Rate system	Demand-charge rate system	Norwich rate system	Declining block rates Paris	Declining block rates UNIPEDE congress
	(1890–1910)	(1908–1911)	(1929–1930)	(1930–1936)
Description	Competitive	Promotional	Promotional	Promotional
Basic principle	Two-part rates	Two-part rates	One part rate with declining blocks	One part rate with declining blocks
Billing of customers according to:	maximum power consumption.	rateable value of the house	(household's size, serves to apportion the dimension of blocks of prices)	(blocks of prices fitted into patterns of consumption and personal utility)
	plus the total quantity of energy supplied	plus the total quantity of energy supplied	the total quantity of energy supplied in each block of prices	the total quantity of energy supplied in each block of prices

Table 7.2 Declining block rates: "Compagnie Parisienne de Distribution de l'Electricité" 1931 (Nissel 1938)

Number of rooms	Blocks of consumption			Price (francs by kWh)
	Block 1 (kWh/year)	Block 2 (kWh/year)	Block 3 (kWh/year)	
1	60	30		
2	90	45		
3	120	60	The surplus of consumption	Block 1: 1,551 fr
4	160	80		Block 2: 0,900 fr
5	200	100		Block 3: 0,257 fr
6	240	120		

technology: on the one hand, growth and diversification brought about non-additive individual peaks (quashing the aggregative assumptions of the demand-charge rate); on the other, the technical novelty of domestic appliances brought about divisible usages of electricity (invalidating the size of the house as a proxy indicator for two-part rates). The conclusion is that, although two-part rates could work out as historical legacies embodied in organisational routines, once the particular link with fixed and variable costs were broken, they became a hollow-shell structure.

The abandonment of this tradition was consequently a logical step under the assumption of promotional and democratising rates. Perhaps the most interesting development in all that followed was the attempt to make the consumer pay strictly for the units of energy consumed, while withholding the size of the house as a discriminatory device. Two-part tariffs were transformed into single-part rates with declining block rates, heightening the incentives for residential demand. The criteria of a household's size, widely accepted as a fair principle, were cast into a new mold, serving now to apportion the dimension of blocks of prices. A short summary of these historical shifts is presented in Table 7.1.

This major form of promotional pricing was initially introduced in Paris by the "Compagnie Parisienne de Distribution de l'Electricité" (Company of Distribution of Electricity), and its basic principle was that the greater the consumption the lower the average price: as more kWh are consumed, the household passes from a first block with standard lighting prices to a cheaper second block, and then on to an even more economic third block of prices. The borderline between one block and the next depends upon the number of rooms in the dwelling, so that larger houses will have proportionately greater volumes of high-priced electricity before reaching the thresholds of declining prices (Table 7.2).

Earlier, in the 1920s, the "Compagnie Parisienne de Distribution de l'Electricité" (CPDE) began experimenting with successive pricing schemes (1921, 1923, 1927, 1928) to increase consumption in segmented urban sectors. The "Compagnie" was one of three enterprises granted concession areas in Paris, but because it had no important industrial customers within its geographical boundaries, its main goal was to improve the discrimination between residential, professional, and commercial customers, and to create specific incentives to these sectors. It was basically a "lighting" company, interested in driving customers out of the afternoon peak period of 4–8 pm, thus stimulating non-peak-time usages (Anon 1935). The experimental and risk-taking attitude toward rates proved to be successful, and CPDE registered an annual 8.8 % consumption growth up until 1929, thus improving its problem of surplus capacity (Levy-Leboyer 1988: 254).

Then, under the shock of the 1929 crisis, the managers extended the policy of discrimination toward the sector of private households, establishing two rates tailored to the market of domestic appliances: one was a broad time-of-day rate in which price varied according to the time of consumption registered through a triple rate meter (able to differentiate between peak period, day hours, and night hours); the other was a declining block-rate based on the number of rooms of each household. Customers could opt for and choose the one that better suited them. Such measures came about precisely at the moment of signature of an agreement of interconnection and cooperation between the super-central stations of the Paris area (January 1, 1930). A local pool of thermal electricity production thereby became established, setting an example of voluntary cooperation between the three corporate giants with concessionary areas. In the following years, the Paris interconnection pool further benefited from the access to reserve hydropower brought in from the distant Alps and the Rhine, through lines of transmission at 220.000 V (Bernard 1986; Anon 1935; Levy-Leboyer and Morsel 1994: 584–587, 695–697, 703–707).

The rationale of rate innovation by the French "Compagnie Parisienne de Distribution de l'Électricité" must therefore be seen as part of an overall change in business environment and business strategy. With the access to thermal and hydro interconnection, the undertaking could improve the load factor through incentives toward "diversification" of usages and accommodate any possible aggravation of peak-load time. Periods of congestion could now be satisfied through the acquisition of reserve power at low marginal costs. Indeed, this is precisely a case where the interconnection acted as a long-term substitute for increasing capacity, since the CPDE diverted the investments planned for a new thermal central station into a program of installation of a superimposed alternate current distribution based in biphasic 5-wire network lines (Levy-Leboyer and Morsel 1994: 1197).

The step into declining block rates appears quite context sensitive: interconnection, access to reserve energy at low marginal costs, demand concentrated in residential, commercial and professional sectors, tendency to lower the ratio of running charges to energy costs are characteristics not easily replicated elsewhere. Some of these characteristics nevertheless represent on-going trends that are about to transform the European economic environment in the years to come. The 1920s and 1930s in effect constitute a period of regional network expansion fostered directly by the diffusion of alternating current. The investment in long-distance transmission lines allowed a flow of current both ways, connecting urban consumers to outlying centers of production. By increasing the cross section of conductor lines and the voltage of the current, distributors could improve the efficiency of transport and minimize the losses of energy. This was particularly important when transport lines extended over a great distance, interconnecting urban centers and distant hydropower stations.

With the building up of regional networks, the distance between a source of power and the final consumer becomes an additional element in the cost of supply.

Given that transport rates can vary with the length in kilometers of the trunk lines, but also according to the voltage of the transmission line, the subscribed power, and the volume of energy transmitted, further diversity is introduced in the techno-economic system of pricing (Berthonet 1988, 1999). Standard accounting practices thereafter start to specify "standing" and "running" costs of electricity production, together with the "standing" and "running" costs of transport (Bolton 1938). For the independent local distributors of low voltage, often under the ownership of municipal authorities, the reference price to delineate "progressive" rates to their customers becomes the bulk price of the electricity delivered by the regional network, sometimes increased by the costs of some self-generation potential. All this is accompanied by a shift in European entrepreneurial research and development (R&D) activities, with product and process innovations in the field of high-voltage technology coming to an end, while the areas of low-voltage technology receive a push of innovative ideas and investments (Giannetti and Lombardi 1995). The uncertain economic landscape of the 1920s brings into sight the relevance of the final nodes of electric systems—the distribution sector. While in the past, the power of influence had been located in the productive side of the industry, with electrical rates designed to closely track generation costs (demand charge), now prices had several margins of adjustment (generation, transport, distribution), and were driven by an intermingled web of interests.

Political and entrepreneurial visions of pricing policies could converge at this crossroad. Promotional and "democratising" rates, orientated toward segmented consumer groups, became all the more feasible as interconnection and access to alternative supplies of energy progressed. Particularly when interconnection could substitute investments in generation—the case of the French "Compagnie Parisienne de Distribution de l'Électricité," incremental sales could be made effective without boosting peak–load costs, at least in the short run.

Many things changed in less than a decade, and on the eve of the 1929 Great Depression, the industry faced a different political economy. Instead of independent local rules being based in ex-ante local regulation, now the Central State imposed ex-post supervision, enforcement, and even blueprints for integrated planning. Instead of direct market relations with customers, a growing number of transactions were carried on between enterprises. Instead of production prices as the main reference, wholesale prices of distribution became the yardstick for the industry. Finally, instead of pricing as an entrepreneurial issue, the relative (rate structures) and absolute (prices) characters of pricing were now subject to political and social pressures and placed under reserve.

7.3.2 The Invention of the Standard Consumer

The micro-Parisian experience flooded back into the macrostage of debate during the 1930s, a moment when the electrical establishments of the most-developed countries faced the interruption of a cycle of high consumption growth (1923–1929) and also bleak prospects for the future: under-utilization of the load capacity, postponement of investments, and accounting constraints. The idea of proactive policies to increase domestic demand could henceforth be justified as the right alternative to counterbalance the losses in the industrial sector. Within the industrial milieu, a focal point of convergence began to emerge around the principle of declining block rates based on the number of rooms of each dwelling. However, while the common detonator that prompted the urgency of decisions was the 1929 crisis and the plummeting of consumption, the propagation channel for the spread of the "Compagnie Parisienne de Distribution de l'Électricité" proposals was a recently established technical and business summit: the Congress of the International Union of Producers and Distributors of Electrical Energy, UNIPEDE ("Union Internationale des Producteurs et Distributeurs d'Énergie Électrique").

UNIPEDE was a private association, created in 1925 as a meeting point for the large undertakings of France, Belgium, and Italy. Its main goals consisted of the exchange of technical expertise, experiences, and points of view, which were debated systematically in a Congress that met every 2 years. As the decade unfolded, more and more countries joined the organization, so that it became largely representative of electrical engineering in Europe (23 countries) and received the additional participation of Turkey, Argentina and Chile (Persoz 1992). Along with this broadening of affiliates came also a renewed interest in new approaches to pricing. In fact, after the Brussels congress of 1930, the concrete model of declining block rates based on the number of rooms of each household became the benchmark for international comparisons. Instead of an assessment of benefits and drawbacks yield by different pricing schemes across Europe, the debate was tapered into a single-rate system. Some came down in favor of the idea, while others raised doubts about its efficiency. Whatever the case, the decision was framed (Lamoreaux 2001) by a single proposal and the agenda locked. In addition, some also remarked that this rate structure had been around for more than 40 years and was also well known in America (Verboud 1934). However, unlike past developments, its application to the technological innovations of domestic appliances, after World War I, created an entirely new framework. Declining block rates appeared as a reasonable balance between insufficient discrimination (setting equal prices for light and domestic appliances) and excessive discrimination (setting special prices for each appliance). In the catalog of historical pricing possibilities, this trade-off corresponded to a position between the demand-charge pricing and special rates (viz. for cooking and heating).

From 1930 to 1938, all routes seemed to lead to the study of consumer profiles. The Paris scheme was reinterpreted in terms of behavioral assumptions, and the rate structure gained a new meaning since the consumer's utility was shaped to fit into blocks of prices (see Table 7.1). A number of papers presented at UNIPEDE Congresses proposed that the decrease in prices must be accompanied by consumer satisfaction in the form of services delivered: in the traditional scheme of three declining blocks, it should start with sales of electricity for lighting (first block), continue with domestic appliances that require little electricity and low power (second block), and finish up with domestic appliances that require a high power demand, such as heating and cooking (third block). So, the fit between block rates and utility presupposes a standard consumer, a man who, after being satisfied with the quality of lighting, buys a radio and an electric iron and, thereafter, enjoys the benefits of a water heater, a refrigerator, or a space heater. The adjustment displays the association of the first block with the average consumption of domestic light, using as a proxy for estimation the number of rooms in a household. After recovering the income from the high-valued light-electricity, the rate can then stimulate additional uses of appliances through the proposal of good prices for the second and especially for the third blocks.

Rate systems and business innovation in the 1930s. (Verboud 1934; Staat 1924; Ott 1987).
Among the entrepreneurial profiles of the European electrical sector, Electricity of Strasbourg, a regional producer distributer formed with capital from A.E.G. and the municipality, corresponds closely to an "intensive searcher" of information, exploring new possibilities in the light of new evidence and investing in further research: in 1927, this enterprise launched a full branch of experiments and studies with declining tariffs, and in the early 1930s pioneered the launching of a monthly bulletin, with studies devoted to customers' behavior, household income distribution, propaganda, and statistics of impacts on load diagrams. Similar initiatives in Europe were at the time only carried out in Germany by Electricity of Berlin ("Berliner Elecktricitätswerke Aktien-Gesellschaft").

Declining block rates offered at this point a good solution, through the transformation of the puzzle of prices into a continuum of services. However, some critics like the engineer Verboud from Electricity of Strasbourg ("Elektrizitätswerk Strassburg A.G."), or the French engineer Jean Solomon, remarked that the borderline between successive regressive prices was randomly defined: only the consumption of light has been persistently studied for years, a fact that enabled managers to grasp sound estimates for the first block of prices. However, "the second and third blocks are mostly established through imitation and guess work, and based on prices paid for power force" (Solomon 1937; Verboud1934). In the view of some of the most reputable authorities in the field of intensive surveys of customers and groundbreaking commercial statistics in Europe the Parisian model of declining block rates fell short of rational justifications. After all, the study of consumer preferences and habits was an absolute novelty for most representatives of the European industry. The forthright issues that started being presented in the UNIPEDE's Congress included: the variation in demand according to the number of rooms in the house; the potential rate of diffusion of appliances; the price elasticity of consumer demand; the probable duration and timing of each electrical usage and its overall impact in the load charge; the demand profile of urban and rural customers.

Generally, basic uncertainties in the introduction of new rate systems required a breakthrough in consumer analysis, additional specialization within commercial departments, and the systematic gathering of evidence. And the most likely undertakings to risk this leap in the dark were those that had less to lose, i.e., which produced energy in thermal technological systems, with backward levels of energy consumption but located in the surroundings of an important urban core, often subjected to political strains to shift pricing systems. After the Paris experience of the "Compagnie Parisienne de Distribution de l'Electricité," declining block rates based on the number of rooms of each dwelling spread, on a voluntary adhesion basis, through Romania (Bucharest, Ploesti, Sibiu, Brasov, Arad), Czechoslovakia (Prague), Austria (Vienna), Poland (Warsaw), Latvia (Riga), and Portugal (Oporto, Lisbon and main cities).

The active promotion of appliances through rate restructuring evolved out of a context where low levels of consumption per capita and per household made the maximization of electricity revenues a condition for industrial development. The idea of recovering the standing costs of peak consumption was replaced by the more commercial-orientated principle of improving the load factor and charging the customer high prices whenever his consumption was strongly inelastic (light) and incentivise its demand whenever there were close substitutes, namely firewood and gas for space heating and cooking (for the price advantage of electricity in the 1920s—in terms of the content of energy delivered—see Barjot 1991; Myllyntaus 1999; Fouquet 2008). In this manner, the marginal utility of consumption can diverge widely from the marginal cost of production. Indeed, while electric lighting makes a small contribution to peak-load time and extends over the flat period of demand, but is burdened with high-value rates, space heating, which directly overlays the peaks (say, the time-frame from 4.30 to 6 pm on a cold winter's day), benefits from a low rate. The fundamental point to make clear is that when the starting point in per-capita consumption is very low, the distortions to efficient pricing are not immediately foreseeable, since most consumers are locked into a first block-rate of high lighting price. It is the achievement of a critical volume of demand and the diffusion of energy services that clarifies, from an engineering perspective, the consequences of the signals sent to customers.

7.3.3 Critical Stands

Alongside the overriding commitment to promotional-orientated rates, there grew a stream of skepticism that evolved into a critical stance. Since the IV UNIPED Congress, held in Paris in 1932, it was clear that the proposal for declining block rates based on the number of rooms of the house, did not gather much support from the undertakings that worked mainly with hydroelectrical sources. The arguments presented by the Italian representative of "Unione Fascista Industrie Elletriche" (Fascist Union of the Electrical Industry), Giacomo Fracazini, cast light on the flaws of the "excessive favoring of the domestic charge during the day-time period" and the risk of "superimposition of the industrial load charge with an increasing household consumption" in peak time (Fracazini 1932: 586). If something had to be done to stimulate the diffusion of domestic appliances, then the available alternatives of special rates or even two-part "tariffs" could give good results. The fundamental in this issue was the possibility of restricting the demand charge of consumers in congested periods. Inversely, "no reduction can be made with the commercial system that divides the overall consumption into several blocks, sold at declining prices" (Fracazini 1932: 586). By discussing electricity prices according to the load charge, the Italian engineer came to a proposal of discrimination based on three time-of-day periods: a maximum price for morning and evening peaks, medium prices for the other daylight hours, and low prices for night consumption, and dropping usages immediately after midnight.

An Italian in Paris, an Italian at the UNIPED Congress could not have felt entirely comfortable. Reliant upon hydroelectricity, Italian undertakings practised severe price discrimination in favor of industrial power and electrochemical industries. After the downturn of consumption in the 1930s, both State priorities and entrepreneurial strategies concentrated on extracting maximum prices from the relatively inelastic civilian consumption, to cross-subsidize the prices of supply to the transformer industry and, after 1936, the prices of electrochemical and electrometallurgical sectors (Giannetti 1993; Barluzzi 1962). To attend a Congress at which most speakers disputed the fast track to promote domestic consumption at low prices, it must have looked like an odd inversion of priorities, at least to Italian ears. In the same tone, Swiss representatives also raised doubts about declining block rates and argued about the necessity of billing the customers accordingly to the costs of production (Niesz 1934; Froelich and Morel 1949). Sales could not be made independent of time-of-day and seasonal factors, because even though marginal costs are equal to almost zero when watercourses overflow and headwater basins reach maximum storage capacity, the cost of production of an additional unit of electricity can be very high when the generating capacity of the dam is at its minimum (the dry season in southern climates or the winter-frozen season of northern mountains). In other words, the volatility of the costs that characterize hydroelectric production, made this particular technology very sensitive to a cost approach toward pricing and to a seasonal and time-of-day perspective on rate systems. On the other hand, the possibility of setting rates exclusively according to the criteria of the quality of the dwelling resembled a blind knot to the future. In particular, when the cycle of low waters coincided with peak private consumptions, as happened with space heating in Switzerland, a very cautious and reserved policy was envisaged (Devantery 1950; Ringwald 1940).

The technological differences of hydroelectric systems set them apart from the entrenched lines of debate. Over the familiar ring of what was rapidly becoming a conventional wisdom, some raised doubts and criticisms. The perception of the potential flaws from promotional rates, lead the engineers of Swiss and Italian enterprises to disclose the bedrock of the cost-based approach: prices should not be cut off from time-of-day costs.

7.3.4 Historical Overview

We have seen how the evolution of rate systems depicted the progressive downgrading of peak-load concerns. The idea of transposing cost accounting classifications into the bills of the customers, so as to make them pay for "standing" and "running" costs, faded away through a succession of small changes. Socio-political, technological, and market factors favored the gradual move away from competitive pricing. At the social and political level, social movements and political institutions endorsed proposals for discriminatory schemes based in consumer traits, rather than charges based in consumption characteristics. Likewise, they also supported promotional rates as an instrument for the democratization of modern life. The municipalization of electrical services and central state interference further eased the attending of constituency claims, and tipped the balance in favor of tangible principles of pricing, like the "size of the house." Technology was a second factor. Thanks to the diffusion of a full array of domestic appliances after World War I, the idea of indivisible supply of a good (lighting power) was transformed into a continuum of divisible services (power for cooking, boiling water, cleaning, heating, etc.). This diversification of demand reinforced the non-additive nature of individual peaks and rendered obsolete the previous rate systems that were tailored to lighting demand. Still on technological grounds, the growth of long-distance transmission and network interconnection eased the tendency to abandon peak-load principles, since some utilities could now satisfy the increase in peak-load demand at low marginal costs. Finally, a third factor was the overturn in the business cycle, caused by the 1929 crisis. Faced with a slump in demand, political distrust toward price discrimination and abuse of monopolistic power, the European power industry found a window of opportunity to restructure the rate systems. Major efforts were concentrated within the residential sector in an attempt to break even and to counteract the losses in the industrial sector. Basic uncertainties about the future, ignorance on fundamental issues of consumer behavior and inconsistencies across commercial departments impelled the major enterprises to instigate an urgent search for information and expertise. This motivation paved the way for a renewed interest in the existing forums of international debate, lifting its agenda, membership, and goals. Instances like the International Union of Producers and Distributors of Electrical Energy, UNIPEDE, were converted into genuine "foyers" of decision-making: not only did the organization frame the available options and taper the debate to a particular version of promotional declining block rates, it also contributed toward spreading its adoption. It followed that the process of decision-making became bound to a binary choice between those who were in favor of and those who were against the proposal of promotional declining block rates. At this point, a forceful reaction came mostly from undertakings that represented hydroelectric technological systems, which showed little enthusiasm for the idea that their task should be to sell as many kilowatt hour as possible. These undertakings feared what could happen with the promotion of low prices and incentives for private consumption independently of time-of-day and seasonal usages. Because hydroelectricity production displayed a high volatility of costs dependent upon climatic conditions, it was very sensitive to branching out prices from costs. By contrast, the feedback from thermal and mixed thermal electricity technological systems revealed a more commercial orientation, emphasizing the concern with the average load rather than with the peak load. In terms of the interpretation of rate systems, the 1930s, therefore, marks a moment of closure when the diversity of possibilities and commercial experiences gave way to a debate focused on a single pricing scheme. Due to the intervention of influential international groups, this “closure” or “framing” prompted the clarification of premises on both the promotional alternative and the cost-based alternative. Yet the most important point to note is how this turning point in the “social construction” of rate systems arose out of cumulative historical uncertainties which involved long-term changes in technology (new usages of electricity, shifts in inter-firm relations), medium-term changes in politics (progressivism and rationalization ideologies alongside enhanced central state interference) and eventful changes in the business cycle (plummeting of industrial demand after the 1929 crisis). Hence, the electricity supply industry experienced a major split. The volatility of costs that distinguished hydroelectric production made the users of this particular technology very sensitive to a cost-based approach on prices and to a seasonal and time-of-day approach on rates. In spelling out these principles at all the international meetings, some electrical engineers laid the foundation for a new understanding of rate systems. In fact, they unraveled sound rules for efficient pricing under the conditions of fluctuating demand: if a rate system must closely track both fixed and variable costs, then each additional unit of energy delivered in the market must reflect the time variation in costs.

7.4 Marginal Cost Pricing as Time-of-Day Pricing

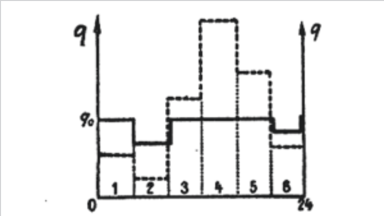
The formal demonstration and development of these principles would only be completed in the ensuing years. On the eve of the 1950s, the French economist Marcel Boiteux interpreted the concept of marginal cost in the context of variations by time of day and year in demand, the source of peak-load problems. In a sequence of seminal articles, Boiteux proved that marginal cost pricing could be practically translated in time-of-day pricing (Boiteux 1960 reprinted from an earlier French text of 1949; Chick 2002). From the 1950s onwards, promotional rates began definitively falling from favor.

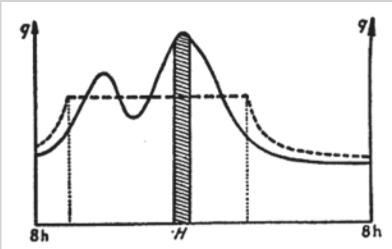
In practical terms, marginal cost was implemented by the nationalized “Électricité de France” (EDF) after negotiations with social interests and with government compliance. This new pricing scheme, based on the discrimination of hourly electricity usage, was made mandatory for industrial consumers as from January 1, 1962. To enhance public awareness, the rate was labeled the Green Tariff (“Tarif verte”) deploying a marketing strategy that set the tone for the straightforward identification of consumer profiles with colors (the Yellow Tariff was tried out initially in 1952, and there was a subsequent marketing campaign for the domestic household “Blue meter”). In this colorful framework, green became the proxy for time-of-day pricing in effect for industrial customers.

The assumption behind this was that the major public service obligation lay not in promoting strict price equality but rather in fully reflecting costs on prices so that any individual could choose the most economic form of consumption. As shown by the graphs in the box below, reproduced from the Boiteux article, marginal cost pricing envisaged the achievement of a minimum cost load charge, deploying optimal equipment usage. Clarity for the user, equality of choice for the citizen, and capital savings for the enterprise became the hallmarks of the public utility.

The load curve before and after peak-load pricing (Boiteux 1960)

The load curve before and after peak-load pricing (Boiteux 1960).





A set of factors came together to steer the course of innovation in pricing schemes, specifically: the greater relative weighting of industrial consumption over domestic consumption in France; the persistence of marginal cost concepts among the great French schools of engineers economists, especially the "École des Ponts et Chaussées," kept alive during the World War Two German occupation by a series of seminars directed by Maurice Allais for the staff at "Compagnie Parisienne de Distribution d'Électricité;" and the fact that the gross supply of industrial manufacturers was better able to absorb the higher costs of multiple tariff meters (Monnier 1983; Picard, Beltran, and Bungener 1985; Chick 2002;). In this respect, the major novelty in recent technological developments is the availability of smart digital meters susceptible to low-cost installation in every private household, expanding the range of choices both for consumers and for suppliers. At least from a practical viewpoint, marginal cost pricing is currently a feasible option.

References

Anon (1932) Rapport Génerale du Comité IV-A. In: AAVV (ed) IV Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique, vol 2. UNIPEDE, Paris, pp 239–241; 531–537
Anon (1935) Le développement de la production et de la vente de l' énergie électrique dans le département de la Seine et dans la ville de Paris. Revue Générale de l'Electricité (XXXVIII):670–675
Barjot D (1991) L 'énergie aux XIXe et XXe siècles. Presses de L' École Normale Supérieure, Paris
Barluzzi F (1962) La Tariffficazione dell' Energia Eléctrica. Comitato Nazionale Energia Nucleare, Rome
Bernard P (1986) Éléments pour une Rétrospective de la Tarification: Principles, Modalités et Procédures. In: Cardot F (ed) La France des Électriciens 1880–1980. PUF, Paris, pp 173–184
Berthonet A (1988) La tarification du transport de l' énergie électrique entre 1920 et 1945. Bulletin d'Histoire de l 'Electricité (32):53–84
Berthonet A (1999) La tarification du transport de l'énergie électrique en France avant la Nationalisation. Bulletin d'Histoire de l 'Electricité, special number 1999: 3–123
Besançon X, Caron F (2005) Bilan et perspectives de la concession dans les pays industrialisés. Entreprises et Histoire 38:120–132
Boiteux M (1960) Peak-load pricing. J Bus 33:157–179
Bolton DJ (1938) Costs and tariffs in electricity supply. Chapman & Hall, London
Bowden S, Offer A (1994) Household appliances and the use of time: the United States and Britain since the 1920s. Econ Hist Rev 47:725–748
Chick M (1995) The political economy of nationalization: the electric industry. In: Millward R, Singleton J (eds) The political economy of nationalization in Britain 1920–1950. Cambridge University Press, Cambridge, pp 257–274
Chick M (2002) Le Tarif Vert Retrouvé: The marginal costs and the pricing of electricity in Britain and France 1945–1970. Energy J 23:97–114
Crew M, Kleindorfer P (1986) The economics of public utility regulation. MIT Press, Cambridge
Devantery P (1950) Les Prix de l'Électricité. Coûts de Production—Tarifs de Vente. Imprimer Held, Lausanne
Doyle BM (1995) Urban liberalism and the 'lost generation': politics and middle class culture in Norwich. Hist J 38:617–634
First Report and Statement of Accounts for the period 13th August 1947 to 31st March 1949, 1950 (1950). His Majesty's Stationery Office, London
Fouquet R (2008) Heat, power and light. Revolutions in energy services. Edwards Elgar, London
Frank D (2001) Where are the workers in consumer-worker alliances? Class dynamics and the history of consumer-labor campaigns. Politics Soc 31:363–378
Fracazini G (1932) Rapport sur les Tarifications plus Appropriées au Développement des applications domestiques de l'électricité. In: AAVV, IV Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique, vol 2. UNIPEDE, Paris, pp 580–588
Froelich WL and Morel C (1949) Les Tarifs de Vente de l' Énergie Électrique en Suisse: évolutions et tendances actuelles. In: AAVV, VIII Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique, vol 2. UNIPEDE, Paris, pp 1–22
Frost RL (1993) Machine Liberation: inventing housewife's and home appliances in the interwar. Fr Hist Stud 18:109–130
Giannetti R, Lombardi M (1995) Behavioral rules in R&D strategies: German and Dutch electrical equipment between the wars. Bus Econ Hist 24(1):206–216
Giannetti R (1993) Dinamica della Domanda e delle Tariffe. In: Galasso G (ed) Storia dell' Industria Elettrica in Itália. Espansione e oligopólio 1926–1945, vol 3. Editori Laterza, Rome, pp 281–315
Gooday GJ (2004) The morals of measurement: accuracy, irony and trust in late victorian electrical practice. Cambridge University Press, Cambridge

Hannah L (1979) Electricity before nationalization. A study of the development of the electricity supply in Britain to 1948, 2nd edn. Johns Hopkins University Press, Baltimore
Hausman WJ, Neufeld JL (1984a) Time of day pricing in the U.S. electric power industry at the turn of the century. Rand J Econ 15(1):116–126
Hausman WJ, Neufeld JL (1984b) Engineers and economists: historical perspectives on the pricing of electricity. Technol Cult 30(1):83–104
Houthakker HS (1951) Electricity tariffs in theory and practice. Econ J 61:3–25
Kaijser A, Hedin M (1995) Nordic energy systems. Watson Publishing, Canton
Lanphier RC (1925) Electric meter history and progress. Sangamo Electric Co., Springfield
Lamoreaux N (2001) Reframing the past: thoughts about business leadership and decision making under uncertainty. Enterp Soc 2(4):632–659
Levy-Leboyer M (1988) The French electrical power system: an inter-country comparison. In: Mayntz R, Hughes P (eds) The development of large technical systems. Westview Press, Boulder, pp 245–262
Levy-Leboyer M and Morsel H (1994) Histoire de l'Électricité en France, vol 2. Fayard, Paris
Millward R (2005) Private and public enterprise in Europe. Cambridge University Press, Cambridge
Monnier L (1983) La tarification de l'électricité en France. Economica, Paris
Myllyntaus T (1991) Electrifying Finland. The transfer of a new technology into a late industrializing economy. Macmillan, London
Myllyntaus T (1999) Energian Hinta 1800–1898. Ministry of Trade and Industry, Helsinki
Neufeld JL (1987) Price discrimination and the adoption of the electricity demand charge. J Econ Hist 47(3):693–709
Niesz M (1934) Tarifications propres à favoriser la diffusion de la cuisine, du chauffe-eau et du chauffage électriques domestiques. In: AAVV, V Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique Zurich—Lausanne, vol 3. UNIPEDE, Paris, pp 120–138
Ott H (1987) History of electricity in Germany. In: Cardot F (ed) 1880–1980 Un Siècle d' Électricité dans le Monde. PUF, Paris, pp 135–150
Persoz H (1992) 40 Ans d'Interconnexion Internationale en Europe. Le Rôle de l'UNIPEDE. In: Tréde (ed) Électricité et Électrification dans le Monde. PUF, Paris, pp 293–303
Picard JF et al (1985) Histoire de l'EDF. Comment se sont prises les décisions de 1946 à nos jours. Dunod, Paris
Poupeau FM (2001) Libéralisation du service public et action publique localeLe département dans la recomposition du système de distribution électrique français. Sociologie du Travail 43(2):179–195
Poupeau FM (2007) La fabrique d'une solidaritè nationale. Etat et élus ruraux dans l'adoption d'une péréquation des tarifs de l'électricité en France. Revue Française de Science Politique 57(5):599–628
Ringwald MF (1940) De quelques Problèmes Intéressant l'Economie Électrique. Bulletin de l' Association Suisse des Électriciens 22:505–508
Siegel G, Nissel H (1938) La Tarification de l' Énergie Électrique. Électrodiffusion, Zurich
Solomon J (1937) Sur la politique de la Compagnie Parisienne de Distribution d'Électricité. In: AAVV, VI Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique, vol 1. UNIPEDE, Paris, pp 265–268
Staat J (1924) Une entreprise de économie mixte L' électricté de Strasbourg. Societe Generale, Mulhouse
Thue L (1992) The state and the dual structure of the power supply industry Norway 1890–1940. In: Trédé M (ed) Électric Électricité et Electrification dans le Monde. PUF, Paris, pp 227–234
Varaschin D (2002) De la concession en Espagne et Italie XIX–XX siècles. Entreprises et Histoire 31:55–70
Verboud M (1934) Expériences faites avec la Tarification Dégressive pour la Vente en Basse Tension. In: AAVV, V Congrés Union Internationale des Producteurs et Distributeurs d'Énergie Électrique Zurich—Lausanne, vol 3. UNIPEDE, Paris, pp 14–65
Yakubovich V, Gravonetter M, Mcguire P (2005) Electric charges: the social construction of rate systems. Theor Soc 34(5–6):579–612

Chapter 8
Energy Intensity

Abstract Energy intensity represents an indicator both of the sectorial economic composition and of the efficiency of national productive systems. This chapter discusses the argument that the historical pattern of development of energy intensity displays an inverted U-shaped curve in which the amount of energy per unit value of constant GDP rises but at some point in the economic development process begins to steadily decrease. The same observation is presented for another type of inverted U-shaped curve, the environmental Kuznets curve, which portrays the relationship between environmental quality and economic growth. To highlight the problems created by high energy intensity regimes, three historical cases of accelerated state-driven industrialization are presented and analyzed: the Soviet Union, Southern Europe, and Brazil.

8.1 The Decomposition of Energy Intensity

The amount of energy required to produce a unit of gross domestic product (GDP) is one of the most fruitful indicators of thermodynamic and economic efficiency. Called energy intensity, it describes the ratio between energy consumption (E) and a relevant measure of activity or output (Y). Owing to the ability to disclose the energy content of the wealth produced in but a single number, the E/Y ratio provides a first clue for the diagnosis of technological systems. Generally speaking, high energy intensities are likely to characterize backward productive systems, while low energy intensities are susceptible to making an appearance in advanced, modernized, and affluent technological systems. From this point of view, the difference between underdeveloped and rich nations lies in the amount of inputs that must be used to produce a given value of output and the environmental impacts associated with the E/Y fraction.

Aside from its usage as an indicator for national productive systems, energy intensity is also employed to measure sectorial and even equipment efficiency. In practical terms, this means that E/Y might express the ratio between tons of coal

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_8,

equivalent and 1$ GDP; the ratio of tons of oil equivalent to the value added per ton of steel; or the ratio between a kWh/year to the refrigeration produced by a system of a given size and features. Inasmuch as nations, industries and equipment require chemical, electromagnetic, or kinetic inputs to produce one unit of output, their relative productivity can be ascertained by the appropriate measurement scale. However, when aggregation takes place over a single technological system, encompassing for instance all the components of refrigeration or a heating system, energy intensity, and "specific energy consumption" becomes one and the same thing.

In the aftermath of Second World War, when national accounting systems improved their methods for estimating GDP by aggregating production through the value added by agriculture, industries, and services, the concept of energy intensity was widely acknowledged and even transformed into a benchmark indicator. The limitations to assessing outputs by resorting just to their physical production volumes were surpassed and alternative monetary assessments of Y became the rule. But what actually triggered the wave of interest around this concept was the discovery that developed nations displayed a cycle with two neatly distinguishable phases: an increase in energy intensity during the early stages of industrialization, followed by a smooth decline once maturity had been reached. The peak in the E/Y ratio embodied the pivotal historical moment in which a unit value of constant GDP demanded unmatched amounts of energy. As research proceeded, several authors confirmed the existence of differential time frames for the turnaround in intensity, reflecting the different paces of economic development: After a period of rising intensity, the decline settled in around 1880 in the UK, in the 1920s in the USA and Germany, just before 1930 in France and around 1970 in Japan (Schurr and Netschert 1960; Darmstadter and Dunkerley 1977; Percebois 1989). These conclusions were valid in the context of energy intensity appraised by primary commercial energy time series as the consideration of non-commercial sources such as firewood could significantly alter the basic framework.

Apparently, nations would tend to become more efficient once having past a certain point in income. As long as subsistence agriculture dominates economic activity, the bulk of commercial energy is consumed in the residential sector for basic needs such as cooking or heating. Then, when people and production activities move on to high energy intensive industries, the E/Y ratio naturally increases. For the same reason, when the less energy intensive service sectors widen their contribution to GDP, the E/Y ratio is reversed and likely enters into decline.

Intuitive as this basic explanation may sound, the sequential outlook is nonetheless insufficient to disentangle the causal factors behind intensity shifts and also to highlight the atypical path followed by less developed nations. From the outset, researchers sensed that changes in economic activity did not amount to a single explanatory factor. Instead, a factorial combination of causes appeared to be working independently to produce time-dependent outcomes. This awareness favored the methodological approach of breaking down energy intensity into components (Schurr and Netschert 1960). In this vein, the reduction in energy inputs for a given value of output could arise by means of shifts in the composition of economic activity, by shifts in consumer preferences, by shifts in the composition of the energy mix (structural compositional effects) and, furthermore, through the increase in thermodynamic efficiency accomplished with the adoption of economies of scale, improved equipment, technological innovation, reduction of losses in distribution, etcetera (energy efficiency effects). Put another way, when societies do apply their energy resources in a more sustained and efficient ways, that might be due to different combinations of effects: (1) Producers might shift production toward finished products that require less energy or abandon heavy industrial activities like blast furnaces and aluminum production, (2) producers might shift from lower-quality fuels to higher-quality fuels, for instance, replacing coal burning by natural gas or electricity, (3) consumers might shift to environment-friendly behaviors, for example, leaving the car behind for public transportation, (4) producers might adopt new machinery doing the same job but with less energy. Ultimately, the decomposition of energy intensity confirmed that the relative weight of compositional and efficiency factors have varied over time. Particularly for late comers that shortened the industrialization phase, the moment of entry into the process of closing the gap with more developed countries has conditioned the opportunities available while driving faster changes in E/Y ratios.

On the assumption that changes in energy intensity at the aggregate level result either from energy efficiency improvements in individual sectors or also from changes in the sector composition of the economy, the index number methodology for the decomposition of these factors may be summarized by the following formula:

$$\text{Aggregate Energy Intensity, I} = \sum_i \left(\frac{E_{it}}{Y_{it}}\right)\left(\frac{Y_{it}}{Y_t}\right) = \sum e_{it}s_{it} \tag{8.1}$$

where E_{it} is the energy consumption of sector i in year t, Y_{it} is a measure of economic activity in sector i in year t, and Y_t is the GDP in year t. Equation (8.1) simply states that aggregate energy intensity is a function of sector-specific energy efficiency e_{it} and of the share of this sector's activity within total GDP s_{it}. In this manner, the efficiency effect may be calculated as the weighted sum of energy intensity changes in individual sectors Δe_{it} while retaining the economy's sector composition constant. Equally, the structure effect can be calculated as the weighted sum of changes in the value-added share of individual sectors Δs_{it} while keeping the within-sector energy intensity constant (Ang et. al. 2003, 2004). Given the fact that the overall intensity in the developed nations, as much as in the world, has been dropping over the last fifty years, which effect proved dominant? For the majority of countries, the uppermost factor that contributed to the reduction in the ratio of energy use to GDP was the gains in efficiency. These were particularly striking in the case of Canada, the Netherlands, Germany, New Zealand, Sweden, and the USA. In contrast, structural changes in the composition of economic activity were by far the principal source of reductions in energy intensity in the UK, Norway, and Japan (Taylor et al. 2010; Mulder and De Groot 2012). It is moreover important to note that the above-mentioned factor of change in the composition of the energy mix and fuel quality is not explicitly dealt with by this standard decomposition model.

8.2 The Drive Toward Heavy Industrialization

8.2.1 Energy Intensity: An Understated Concept

In the first half of the twentieth century, capital, labor, and land were at the center of economic growth concerns. Compared with these primary factors of production, energy was an intermediate input of little interests. Capitalist and communist visions of development wagered all their betting chips on "land" and, especially, on "capital." In such a context, any constructive ideas about the role of energy in economic growth had to be only fairly incipient. One of the rare studies published in this period sought to highlight the role of energy productivity, defined as output per unit of energy input, i.e., $P = Y/E$. The agenda was set by a comparison with capital, labor, and land productivity with the crucial question surfacing around energy as an explanatory factor for economic growth. Strikingly, its author, Frederick Gale Tryon, an American authority on coal issues, provided the first account of energy intensity without ever noticing, let alone, mentioning it (Tryon 1927). The reason is that the essay departed from the theme energy productivity $P = Y/E$ but could not avoid reasoning in terms of its reciprocal $I = E/Y$, which, of course, is no less than energy intensity itself. Actually, the idea of energy productivity expressed as enhanced throughput with lesser or equal amounts of energy inputs matched the idea of lower energy inputs for a given output, or a decline in intensity—both being flip sides of the same coin. The obvious conclusion points to the veiled, deferred, and convoluted appearance of the concept: Energy intensity was a concept of bastard origins and borne without legitimacy and without a name.

The prevalence of productivity over intensity meant that the bright side of energy usage by far supplanted the potential dark side. Any concept that might represent a proxy of the environmental costs of capital formation was accordingly overlooked in economic thought so that the major problem in project management boiled down to sound control over energy costs. In view of that, the more energy intensity remained unnoticed, the larger the payback ascribed to large-scale industrialization. With everything but capital out of the equation, the benefits of shortening the gap toward mighty nations, through investments in heavy machinery and plant, appeared alluring. Heavy industry not only promised a high rate of return and, therefore, faster capital accumulation, but also the incorporation of import substitution technologies and the spread of efficiency gains throughout society. By the close of the First World War, metallurgy, machinery-building, electrical equipment, and chemicals constituted the four pillars of this fast "catch-up" model. In thermodynamic terms, the step ahead meant overcoming productive processes based on the low temperatures (up to 600°) needed for smelting copper, silver, or gold, making tiles or baking clay, through to the much higher temperatures (2000–2500 °C) and higher energy intensity industrial process needed to produce harder alloys like nickel and tungsten steels, or attain the high melting points required by aluminum oxide. Politically, heavy industrialization became the motto for rulers aiming to modernize their nations through great leap forward and in the meantime showing total impatience toward any slacking in tempo that might undermine their revolutionary and nationalist plans.

Under the banner of progress and modernization, it is hardly surprising that revolutionary industrializers looked on at the mass of peasants with the same regard that social Darwinists looked on at Africans and Negroes or how positivists perceived the superstitious beliefs of rural populations. They represented the leftovers of backwardness, a thorn in the side of progress. What was at stake was the massive dislocation of economic incentives so as to achieve a structural "compositional effect" in society and throughout the course inevitably sacrificing the most exposed sectors of the peasantry. Since this compositional shift was accomplished by design, all human and environmental consequences were brutal and long-lasting. While the modernizing shock envisaged by heavy industrialization turned out to be, by most historical accounts, generally accomplished, its deep consequences were only perceived after the fact: large-scale unemployment, poverty, mass migration, pollution, urban degradation, stop–go patterns of development, inflation, reinforcement of authoritarian policies, and, on occasion, income asymmetry.

8.2.2 Soviets, Electrification, and Industrialization

Rare sentences have been sharper in summarizing a global vision of the world than Vladimir Ilyich Lenin's renowned statement: "Communism is Soviet power plus the electrification of the whole country." Proclaimed in a speech delivered to the Moscow *Gubernia* conference, in November 1920, this declaration set the tone for the new challenges that ensued with the end of the civil war against the counterrevolutionary "white" Russians and asserted the transition in the economic reconstruction of Soviet Russia. In reality, Lenin already had in his hands parts of the report long since prepared by the State Commission for the Electrification of Russia (GOELRO), which provided the blueprint for the Bolshevik goals in the years to come. Placed in its historical context, "Soviet power plus electrification" was a reminder, particularly for communist activists, that the modernization of Russia was no less important than all the forthcoming political battles and the reconstruction of the Soviet State. No one could remain unmoved by this call which represented the driving force for the creation of the Russian proletarian state. On this point, the Bolshevik leader revealed a close resemblance to many of his capitalist counterparts who, in the West, also envisioned the benefits of switching to the high energy quality that electricity bestowed, transforming a technological possibility into an arousing program for nationalist modernization (see Sect. 8.2.3). Irrespective of the common approach, the urgency of the material foundation of socialism was much more pressing for the soviet regime, which struggled to survive amidst a fairly hostile international environment.

Strikingly, the first long-term public investment plan proved to be a balanced, unifying, and feasible enterprise. To start with, the most important regions earmarked for electrification, received their share of public investment, and were assigned separate programs, setting the rule of an inclusive policy toward the republics (Ukraine, Belorussia, Georgia, Azerbaidzhan, and Armenia) and within the Russian republic itself. In second place, power generation entailed the new goal of diversification of power sources, aiming to construct 10 hydroelectric stations and 20 coal-fuelled power stations. Through this approach, it became possible to consolidate the Russian school of dam engineering, neglect during the Imperial period, and accomplish the assimilation of cutting-edge foreign technology, most of which came from North America and Canada. In nearly a decade, construction technology evolved from the V. I. Lenin Volkhov hydroelectric station with a capacity of 58 MW (1926) to the achievement of the Dneprostroi Dam, with its overwhelming capacity of 558 MW (1932). Considered an international landmark in several aspects, Dneprostroi stood as the symbol of socialist endeavors owing to the challenge faced by engineers and workers in the erection of a reinforced-concrete dam in Ukrainian desert lands, and with extending across 760 and 60 m in height (1932). The remarkable scale of the Ukrainian Dam remained, in any case, the exception. Later, in the Khrushchev era, hydroelectric "gigantism" became subject to criticism owing to cost slippages and to the waste of thousands of hectares of arable farmland.

From the outset, the decentralized and regional approach of the first electrification plan pointed to separate networks based on large thermal power stations with an average capacity situated in the 50–60 MW range. However, even these thermal plants pushed the diversification of Russian power sources further ahead since high transportation costs and the Soviet policy of promoting recourse to cheap local sources stirred the extraction of lignite from the Moscow Basin, anthracites from the Donbass, peat from the Volga and North-West Districts, and even oil shale from the Leningrad and Volga fields. The core of Soviet energy reserves for thermal plants and industry supply continued nonetheless to rely on the hard coal of the Donets Basin. One of the first measures undertaken to revive the economy, in 1921, was precisely to give extra food to coal miners from this region so as to enable them to perform their work capably.

Finally, despite the clear priority given to the supply of manufactures, GOELRO also foresaw further goals, namely worker access to lighting services and the whole cultural brew conveyed by electricity, along with the reinforcement of political linkages between the working class and the peasantry, the town and the country (Lenin 1920). Bonding principles, diversified sources and multiple end-usages rooted the feeling of a balanced pattern of development after the chaotic war economy. For these reasons, the electrification plan filled a prominent role in the transition from the heavy-handed system of War Communism to the New Economic Policy (NEP) that ensued.

Following victory in the civil war and in the war against Poland, the Bolsheviks were confronted by the crude reality of a backward nation, beleaguered by destruction, beset by famine, hyperinflation, and epidemic outbreaks, ravaged by all sorts of bands, wary of authorities, ungovernable from city headquarters. Determined to hold onto power, Russian communists undertook a complete switchover of former policies and booklet principles in March 1921: Instead of the system for requisitioning peasant goods and the nationalization of all branches of industry, farmers were enticed to yield and sell their own products and private individuals were allowed to form small enterprises or to lease them from the state. More freedom was granted in internal affairs and the regime even attempted openness toward foreign investment and foreign relations, albeit unsuccessfully. Identified as Lenin's approach, the New Economic Policy set down a mixed economy system, in which large-scale enterprises, mining and banking firms, and foreign trade continued to be held by the state as public property, whereas small-scale business grew into highly successful cocoons of free enterprise. Through the control of key sectors and infrastructures, of which the utmost modernized was electrical generation and associated electrical equipment, the Bolsheviks deemed themselves to be preserving the "commanding heights" of the economy while driving the process of transition.

Strong at first, the recovery rapidly slowed to single-digit GDP growth in 1926/1927, precisely the moment when all physical energy indicators finally surpassed their pre-war levels (1913). By this stage, small-scale private enterprises dominated the economy, producing more than 50 % of the national income. A significant part of the Bolshevik party witnessed this revival of capitalism with increasing concern and alongside the ascension of the new landowner class, the urban intelligentsia and the in-group of traders nicknamed the NEPmen, who explored all the opportunities and market flaws to accomplish lucrative transactions. Beneath the surface of political power, socialism appeared to grow weaker with each passing day. Torn between the continuity of NEP and the need to resume the path undertaken by the Soviet revolution, the Bolshevik party experienced a major realignment of different groups and schools of thought just as soon as the first economic problems cropped up. By 1927, the economic slowdown combined with unyielding peasantry distrust toward the marketing of surpluses, triggered a procurements crisis which soon impaired the supply of grain to the cities. The reluctance to sell grain to the state, along with the tendency to concentrate on livestock and other crops, for which prices were more favorable, was particularly critical in West Siberia, the Urals, and the Volga, all regions where the harvest had been reasonably good. On top of that, Bolsheviks blamed better-off peasants, the so-called kulaks, of anti-Soviet practices, such as holding preferences for private traders rather than state procurement agencies, hoarding in expectation of higher prices, and wasteful self-consumption by feeding livestock with grains.

Bottlenecks in grain supply were the sign that Lenin's economic model for the transition toward communism was no longer able to reconcile agriculture and industrial development. Nevertheless, a much more radical interpretation also came into being according to which the supply breakdown disclosed the unfeasibility of any compromise between the Soviet State and free markets. The "emergency measures" to coerce the peasants into handing over their grain as adopted by the Politburo in the winter of 1927–1928 reflected this view and the perceived need for a major break with NEP. Senior party officials, backed by the army, henceforth embarked on procurement campaigns across vast agricultural regions, enforcing expropriations, the suppression of private merchants, and the indictment and arrest of any peasant resistance. Altogether, the criminalization of private ownership set a course of action from which there was no way back. Joseph Stalin, by then already a dominant figure in Soviet politics and General Secretary of the Communist Party, was riding the wave and securing his grip over the Politburo and the party as an unyielding leader. As matters turned out, the crushing of the peasantry and the suppression of private property were measures hard pressed for by Stalin, dealt with by his close political entourage, manipulated to isolate and discard his political adversaries within the Politburo and from which Stalin naturally reaped the benefits of undisputed leadership. Pushing forward at the first hint of crisis henceforth became the trademark of the Soviet Party. Retreat, a roundabout approach, and a new tougher push all complemented this behavioral pattern. More to the point, Stalin was as much the initiator of fundamental decisions as the supreme arbiter in the inter-institutional conflicts that ensued. Over and over again, he did play this double dictator role (Davies and Harris 2005). However, should a man's character be best revealed during a time of crisis, then the 1927 crisis revealed the traits of character most appreciated in Bolshevik leaders: resolve, firmness, and brutality (Gregory 2003: 17).

Communist thought was well aware that peasant support would be lost once the threshold of collectivization was crossed. In reality, the debate about agricultural policies loomed shortly after 1917 and bound up with the question of "super-industrialization." While the majority of the party agreed with the view of a balanced economy during the "transition phase," preserving private land ownership and slowly winning the peasants over to the idea of collectivities of production through the enticement of modern technology (tractors) and high productivity, the left wing (Trotsky, Preobrajenski) postulated the need for a fast pace of development based on accelerated industrialization, channeling capital to state-owned enterprises in heavy industry, creating mandatory collective farms so as to foster high rates of socialist "accumulation of capital." This was a theoretical debate, mostly about the pace of transition to the communist society, embroiled in the rhetoric of steps, goals, and phases. Formerly a defender of moderate policies toward the peasantry and an opponent of revolution from above, Joseph Stalin switched sides and appeared, in 1927, as the leader of bold and unhesitant responses. Upholding this position, the vast majority of cadres then also subscribed to the path of acceleration. For them, super-industrialization became a preventive policy aimed at strengthening Soviet muscle and exerting dissuasive leverage against any possible external aggression. This policy would enable the economy to jump the stage of gradually accumulating technology and knowhow and achieve in just 10 years what normally would take about 50 years to accomplish. Thus, the next step entailed the full nationalization of the economy, moving the peasants into collective units of production (kolkhozes), and the enactment of five-year investment plans. In a short while, the Soviet economy would be state-owned, nationally integrated, and centrally commanded. Spearhead by Stalin, the unfolding of events transformed the conjectural crisis and the need to cope with internal threats into a far-reaching change in the economic system so as to deal with forthcoming external threats. The linkage of microevents to macrosolutions was the bedrock for the defensive geostrategic concept of "socialism in a single country."

As the great breakthrough proceeded, economic decisions become more centralized at the top of the Soviet State. Ministerial and regional authorities replaced trusts, syndicates, and enterprises that lost their earlier microeconomic freedom. The first five-year plan (1928–1932) was to operate with effect from October 1928, albeit only in April 1929 was final approval granted at the sixteenth party conference. Along the process, the technical specialists in central planning (Gosplan) and the superior state institution for economic management (VSNKH) were under continuous pressure to adopt more ambitious growth targets. Culminating out of a long debate on whether socialist planning should base its forecasts on the extrapolation of empirical data according to objective economic laws or rather set targets for an optimal combination of productive forces (dubbed, respectively, the genetics and the teleological approaches), the first five-year plan tilted the balance clearly in favor of the latter methodology (King 1999). Socialism could set its own targets because it was free from the constraints of objective economic laws. Unlike capitalism, "subject in its development to the spontaneous laws of the market," and "determined by circumstances," the Soviet economy had, "to a large degree, freed herself from the complete domination of this blind spontaneity," so that economic growth could ultimately come to depend on the "collective will" (SG Strumilin quoted in Boobbyer 2000: 47). Targets, therefore, ought to supersede circumstances and reflect the greater rationality and more effective usage of resources bestowed by the planned socialist economy. Because the fully planned economy was superior to spontaneous markets, every physical or technical constraint could ultimately be overcome. With the fundamentals of economics taken over by political will-power, higher growth targets could not but reflect the commitment to socialism and the belief that larger amounts of physical production could be achieved in five years. In contrast, less ambitious "optimal targets" revealed a "wreckers" mindset.

The outcome of this teleological disposition was the upward revision of targets soon after their adoption so that during 1929 and 1930 the five-year plan was altered to the most eccentric of goals, most of which were only actually attained after Second World War. In reality, not only were the physical amounts of production consecutively raised but also political resolutions additionally called for the fulfillment of production targets in just four or even three years. Another consequence of the overoptimistic ideology that pervaded the political leadership came with the arresting of high-level specialists from central agencies charged with running the economy. Renowned scientists such as Kondratiev, Makarov, Chaianov, Bazarov, Groman, and others, who called for technical caution and who drew attention to difficulties and bottlenecks, were accused of being counterrevolutionary wreckers (Khlevniuk 2008).

The disregard for objective constraints to economic growth moreover translated the view that natural resources could be obtained at low costs and that nature required taming and reshaping according to a socialist mold. Massive capital investments and the takeoff of metallurgy, machinery, and electrical equipment industries required huge amounts of fuels. It was thanks to the success obtained in the exploration of Russian, Ukrainian, and Azerbaijan riches that Soviet industrialization fared reasonably well, even if falling far short of the overstated targets. In economic terms, a large part of the achievements were explained by the fact that, since the First World War, the exploration of Russian natural resources hinged on the availability of spare capacity leaving a greater margin for progression through investment in new discoveries and new capacity. In oil production, for instance, the largest increases in the Stalinist period resulted from the exploration of relatively new fields (the oilfields of *Surakhany* and *Bibi Eibat* Bay in Azerbaijan; New *Aldy* in North Caucasus; *Temir, Novo-Bogatinsk,* and *Karaton* in Kazakhstan) augmenting the number of wells per field, to counterbalance the declining production of aging platforms. Likewise, new discoveries hit during the execution of the plan were scheduled for immediate production, the case of *Nafta Chala* in the south of Baku, and the new pool in *Maikop*, nearer the Black Sea. Despite the revitalization of the industry, all exploration remained concentrated in the regions where proven reserves had long been identified. Indeed, few test-drillings were carried outside Azerbaijan and the North Caucasus, keeping the geographical profile of the Soviet industry unchanged until the 1950s. Under these conditions, oil experienced stabilization in output after an initial boom, in sharp contrast with coal's pattern of steady growth throughout the 1930s. So important was this fuel that one may even say that coal was the real cradle of heavy Soviet industrialization.

Strikingly, coal enterprises avoided the orthodox gigantism of Soviet economics to embrace the scale of medium- and small-sized units, more adapted to the diversity in geological conditions. It was precisely within this milieu that work efforts could be individualized and rewarded as happened with Alexei Stakhanov, a coalminer who cut 102 tons of coal in a single session instead of the prescribed 7 tons and for this reason becoming an example of the "new Soviet man." Expansion in the foremost productive area of the Donets Basin, in Eastern Ukraine was achieved through rapid mechanization and particularly following the import of coal-cutting machinery. In other mining centers, coal progression followed in the footsteps of investments in new metallurgical and machinery equipment. In reality, the areas of development were located nearby pools of ore and coal, including the Urals-Kuznets (the Kuzbass Basin in south-western Siberia) with its excellent coal for coke, and the Karaganda and Moscow basins. In these pivotal areas, the installation of giant enterprises exemplified by the Magnitogorsk iron and steel combine, the tractor factories concentrated in the Stalingrad, Kharkov, and Chelyabinsk industrial belts alongside steel mills, machinery factories, and power stations brought about soaring demand for hard coal and brown coal. All the more so, when most of these iron and steel production facilities were established from scratch (Russia's Five Year Program 1930; Elliot 1974; Krylov 1998).

Lenin's and Stalin's policies held little in common from the point view of the relationship between energy and economic growth: The former thought in terms of guaranteeing access to high-quality energy by turning electricity into the key driver for change and modernization, while the latter conceived of increasing

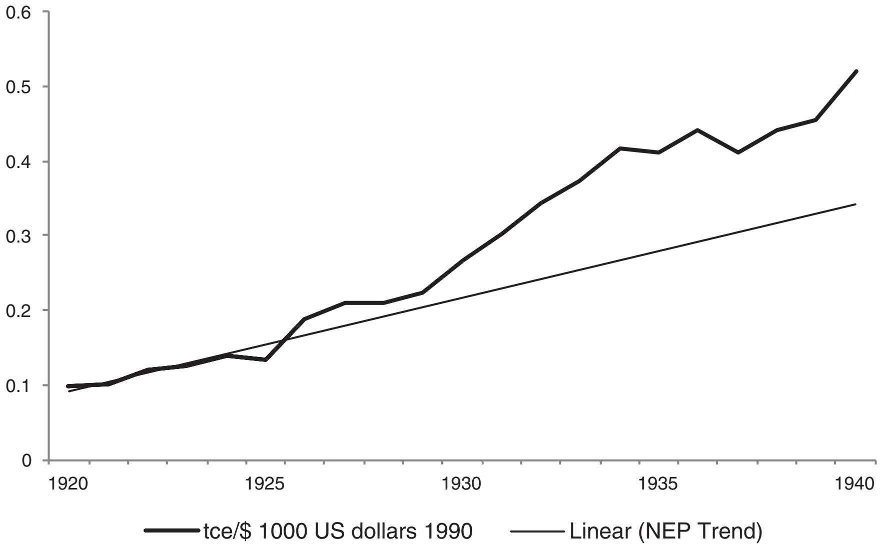


Fig. 8.1 Energy intensity in Soviet Russia (1920–1940). Tons of coal equivalent per USD 1,000, estimated by 1990 Geary–Khamis dollars. (Energy data: Etemad and Luciani 1991; Mitchell 1981; population and GDP data: Davies et al. 1994; Markevich and Harrison 2010; Bolt and van Zanden 2013). *Note* Energy intensity (E/Y). Energy consumption = Soviet production (hard coal, lignite, peat, oil, natural gas, hydroelectricity) + imports (coal, oil) − exports (coal, oil). Conversion coefficients to tons of coal equivalent based on Etemad and Luciani (1991)

the amounts of primary energy in order to support accelerated industrialization. One strove to infuse the grand axioms of improved useful work, electrification and the power of the Soviets; the other, raw inputs, coal–iron development, and advancement through a major break with the existing policies. Should this contrasting interpretation holds, there should be a sharp difference in the intensity of Soviet energy consumption reflecting the divergent stands toward the relationship between energy and economic growth.

Figure 8.1 depicts the energy intensity of Soviet Russia. The graph shows that the country moved on from 0.1 tons of coal equivalent (tce) for each USD 1,000 produced, to 0.52 tce, or half ton, for each USD 1,000 in 1940, a fivefold intensity increase in just twenty years. Drawing on the initial NEP phase, the straight line in Fig. 8.1 represents the linear trend extrapolated from the partial time series of 1920–1926. In a simplified form, it shows how energy intensity would have evolved had there been no large changes introduced into the Leninist policies. The upshot of the soft NEP's path should have led to about 0.33 tce for each USD 1,000 in 1940 as against the 0.52 really observed. In the starkest terms, this means that the leap forward in heavy industrialization added 0.2 tons of coal equivalent/USD 1,000 to the Soviet energy intensity. However, the most remarkable lesson to be drawn from the graph is the surge that took place between 1929 and 1934. This skyrocketing in energy intensity during the First Soviet Plan and its immediate aftermath plainly pinpoints the exceptionality of the period. As several authors have stressed, the early phases of super-industrialization are characterized not only by shifts in the composition of economic activities but furthermore by the occurrence of unexpected wastes, generalized contradictions, and the mismanagement of resources. In the Soviet Union, a large part of the inefficiency stemmed from the flood of unqualified peasants into industrial manufacturing alongside the drawbacks and knowledge shortages caused by the mass repressions of specialists, engineers, and technically skilled workers. Thus, despite the phenomenal rate of growth in nominal investment, labor productivity fell at an annual rate of close to 10 % by year. To this, one must also add the constant bottlenecks stemming out of the central allocation of resources, the generalized lack of construction materials, industrial and agricultural goods, excessive outlays invested in unfinished constructions, the importance of non-economic factors in apportioning supplies, delays to timetables, and the breakdown of industrial discipline. Altogether, this lack of preparation, inefficient management, and project gigantism combined to aggravate the wastage of labor and physical resources. Energy began being consumed without the proper return in throughput, prompting an overarching rise in intensity. Under these conditions, Stalin and the Politburo proposed the first cuts to investment in 1932, accepting with resignation that the economy lacked the physical ability to continue with investments at the pace set by the plan. Reflecting this political learning, the second five-year plan (1933–1937) become much more moderate and realistic. Its original targets, which were introduced at the 17th Party Congress, were in fact later subject to further reduction (Gregory 2003: 94–111; Davies 2003; Ellman 2008).

While the human cost of Soviet industrialization may have been reasonably quantified, much less is known about its environmental costs. One recent historical account showed that problems with biodiversity, the collapse of ecosystems, and water and air quality issues spanned the heavily industrializing regions of Ukraine, the Urals, Volga, and Southern Siberia to the difficult settlement areas, located in northern Siberia and the Arctic and where the hardships of climate, ice-covered land, and cold tundra did not dampen the exploration of nonferrous metals and timber. These were the more energy intensive economic activities because supporting life in such adverse conditions required huge amounts of energy (typically ten heating plants per 10,000 inhabitants), even taking into account how the lives at stake were chiefly those of political prisoners compelled to forced labor (Josephson 2013: 103–106). The low priority attributed to factors other than the plan's fulfillment opened the way to the introduction of contaminants into the environment, the destruction of natural habitats, soil erosion, and the alteration of the makeup of soils and flora in areas crossed by rivers and very long-lasting problems.

To sum up, the increase in energy intensity stood as an unforeseen consequence of heavy Soviet industrialization. However, judging by the way the Bolshevik leadership envisioned the geostrategic transformation of backward Russia, the outcome of the 1930s turned out to be an indisputable success as demonstrated by GDP growth at 5.4 % year. Furthermore, according to Stalin, there was no true alternative to the path taken. In a notorious speech delivered at a conference held in February 1931, Stalin asked: “Do you want our socialist fatherland to be beaten and to lose its independence? If you do not want this, you must put an end to its backwardness in the shortest possible time and develop genuine Bolshevik tempo in building up its socialist system of economy. There is no other way” (Stalin 1931).

8.2.3 Southern European White Coal

National survival and collective beliefs provide powerful arguments for speeding up the pace of change in society. On a far softer basis, the conviction that a fantastic technological revolution might be dawning might cause the same effect. The will to abruptly change the composition of the economy is particularly strong in the initial phase of development cycles, whereupon many aspects still remain unknown and the feasible is muddled in with the likely. Extreme cases of technological utopias include the promise of cost-free nuclear energy, production automation redeeming mankind from the servitude of work and intelligent robots replicating human action. Apart from these visionary letdowns, more common, down-to-earth misjudgments result simply from the wishful assessment of costs, of applications and market conditions. By the dawn of the twentieth century, for instance, the initial expansion of hydroelectricity raised the hope of rapid industrialization. However, this hope was not always grounded on equal comparative advantages regarding the endowment of natural resources (water flowing in high altitude rivers; with high precipitation and an abundance of water courses). Whereas in northern Europe, the open trajectories for industrial innovation became a function of economies of scale based upon high hydrological potential, low population density, and reasonably advanced levels of development, expressed in the fact that industrialization was already in progress when electricity first appeared, Mediterranean Europe faced much more adverse conditions. In these circumstances was it beneficial to engage in high intensity industrialization? Could backward countries take advantage of the new dam technology to catch up with industrializing nations?

Investment in super-central hydropower stations would soon lead to an excess in supply that might prove valuable in terms of the low-priced energy available to industrial manufacturers, following the US pattern at the Niagara Falls corporate complex. Quite often, it was the manufacturers themselves who built their own power stations and power lines, fully integrating the self-generation and distribution of energy into their main industrial business. Such microembodiment of electricity in the capital goods of firms was particularly intense in Scandinavia and opening up the way for a split between the market for commercial and domestic clients on the one hand and the motive power market on the other hand. The comparative advantage of cheap electricity prompted the substitution of productive factors such as capital and work by energy, thus stimulating a wide range of intensive industrial applications for hydroelectricity. First, there emerged the new electrochemical and electromechanical manufacturing sector, which took advantage of competitive energy inputs to produce nitrates, carbide, carborundum, corundum, and cyanamid. At the beginning of the twentieth century, electrochemical production consumed 50 % of the hydropower produced in Norway, 34 % in Switzerland, and 30 % in Sweden. The advantages gained from the exploration of vast amounts of hydropotential also stimulated the emergence of companies specializing in the production of power and transmission equipment and of electrical companies that became outstanding competitors in international markets. The Swiss Compagnie de l'Industrie Électrique/Sécheron and Brown Boveri, the Swedish A.S.E.A., Stal-Laval and Electrolux, the Finnish Tampella, Srömberg and Kone, and the Norwegian Tamdberg are the best-known examples of the maturity achieved by this electrical machinery and equipment sector. Finally, there was the electrometallurgical industry, strongly based on the manufacture of aluminum, where Norway and Sweden soon ranked among the largest exporting countries (Glete 1986; Thue 1992; Myllyntaus 1992, 1995).

In Southern Europe, the diffusion of electricity represented an opportunity and a turning point both to counterbalance the dependence on imported coal and the high distribution costs to inland regions and to reduce the scope of action of foreign enterprises. At a more ambitious level, electricity was also perceived as potentially instrumental in changing industrial output and its mix through the development of new clusters of electromechanical equipment, metallurgy, and electrochemical production. Opportunities thus arose both in the replacement of general-use technologies and in the specific applications to industry. In this context, investment in energy production became the touchstone for proposals of rapid industrial development, establishing a link between modernization and national resources a fact that explains the technological pitch of twentieth century economic nationalism. It was after the shortages and disturbing times of the First World War that the governments, public opinion, engineers, and entrepreneurs of Southern Europe turned their attention to the potential of hydroturbines to generate power, exploring the abundant, renewable, and still relatively unexplored resources provided by water flows and rivers. The developmental tendency was further favored by the mainstream perspective of autarkic growth that characterized the regimes of Salazar in Portugal, and Primo de Rivera and Franco, in Spain.

Starting with this latter nation, the 1920s were, in effect, the golden age for the expansion of hydroelectric capacity, pushed onward by relative price advantages over the alternative use of coal (both imported and national), by an industrial spurt that boosted the demand for energy products and services, and by the escalation in foreign investment. The doubling of capacity and the fourfold increase in total electricity production within ten years (1,027 kW of capacity and 2,433 GWh produced in 1929) awakened the idea of a demand-side push to foster economies of scale. The role of energy in industrialization was nonetheless approached from different angles:

The forethoughtful view held that the real hydroelectrical potential of Spain was ascertained not by its natural riches but by the competing price of coal-fuelled engines. Contrary to the then currently aired assessments, Spanish hydrological resources might not be so overwhelming and perhaps attaining the productive potential of 2 million kW. The industry should correspondingly avoid risks and invest only when sufficient demand was in sight, exploiting the most productive sites for hydroelectricity in each region, and deploying every possible engineering skill to cut back on capital costs and build cheap stations (for instance, falls without a dam or integrated run-of-river systems and storage dams capable of maximizing water flows and heights).

The import substitution perspective held that the hydroelectrical potential of Spain was worth expanding as a means of cutting imports of foreign coal. To this end, every measure susceptible to increasing internal production would be welcome: extending the railway network, for instance, seemed indispensable to fostering the commercialization of Spanish coal and overcoming geographical bottlenecks; promoting new electric blast furnace technologies and nurturing new markets for hydroelectricity were other paths open to take full advantage of interfuel substitution. Naturally, the import substitution point of departure was consistently more optimistic toward resource assessment, foreseeing a total hydrological capacity of 7.5 million kW, even if only in some distant future. A final concern of this stance, largely imbued by economic nationalism, was the distrust toward regional policies then ongoing in Catalonia and the objective favoring of French, Belgian, and North American interests in electrical enterprises.

Finally, the super-industrialization view deeply committed to a bold strategy of cross-fertilization between hydroelectricity and industrial capital. The key sectors singled out for this industrial spurt included the electrification of railways and the advancement of the incipient electrochemical sector. Such an ambitious strategy understandably required the enlargement of generation capacity through central state involvement in setting up a high-voltage distribution network and fostering a leap forward in production. The expansion of coal extraction in the Asturias region and hydroelectric expansion on the Ebro and Duero rivers in this regard seemed the most promising options. Interconnection plus increased capacity thus came to represent a joint issue justifying centralized schemes of nationalist development and pull-over planning. Significantly, these ambitious policies were put into practice irrespective of the appraisal of the actual hydrological potential.

Juan Urrutia, the head of the Hidroeléctrica Ibérica business group, the geographer Huguet del Villar, and the physicist and academic Perez del Pulgar were leading representatives of these streams of thought (Antolín 2003; Rodriguez 2007; Villar 2010). While the 1920s witnessed the expansion and concentration of major hydroelectrical business groups and the postponement of heavy industrialization strategies, what loomed afterward was a crude and rudimentary version of the import substitution thesis.

General Franco's victory in the civil war and the implementation of a rough autarkic style of traditional interventionism (1939–1951) sidetracked the hydroelectric sector from having any possible influence over the capital goods industries. In an atmosphere of unrelenting political repression, the regime embraced the solution of a controlled economy characterized by isolation from foreign markets, limitations on foreign investments, interference in the allocation of resources, price ceilings, and rationing of goods. The twisted economy of coupons and black markets that surfaced encouraged the diversion of scarce energy resources to state defined priorities, favoring administrative services and import substitution industries such as cement, iron, steel, and metallurgy. Overall strict controls over production, prices, and the consumption of coal, combined with the freezing of electricity tariffs, resulted in excess demand, the blockage of investments and overutilization of industrial power in certain regions, principally Catalonia (Sudría 1994; Delgado 1994). On the other hand, the overt hostility toward foreign enterprises, embodied in the linkage between the Franco government and the forerunners of the corporative–nationalist leaders of the National Review of Economics (*Revista Nacional de Economia*) reflected many of the feelings of those political economists subscribing to the above-mentioned import substitution and super-industrialization theses. The outcome of the autarkic inward-looking state was a moderate rise in energy intensity due to the reinforcement of heavy industrial sectors (Bertoni et al. 2009).

As matters turned out, it was in neighboring Portugal that a more radical program of accelerated industrialization came to the fore. By all accounts, its contemporary supporters praised the program as combative, groundbreaking, and thriving on adversity while overcoming all the traditionalist rural-based opposition within the "Estado Novo" regime, headed by the dictator Oliveira Salazar. At the forefront of the battle were a group of young technologically minded engineers who foresaw the advocating of hydroelectricity as part of an integrated blueprint for industrial modernization based on the launching of new manufacturing activities and on the restructuring of traditional production through concentration, rationalization, and competitiveness. Electricity consequently became an argument in favor of fostering economies of scale through a global strategy to clean up the myriad of small-sized firms, while low-priced energy prices became the necessary incentive for the introduction of new industries such as the electrochemical and electrometallurgical sectors. Instead of the gains achieved through import substitution, the engineers appealed to the gains achieved through a closer embodiment of energy in capital and intermediary goods. This change of vision also brought about a redistribution of the institutional roles of market actors: While previously the focus on replacing coal imports had led to the favoring of municipalities and private enterprises as appropriate actors in development, these bodies now seemed too weak to drive the strong surge actually necessary. Only the central state contained the conditions and the capacity to attain these achievements and embark on a path different to the developing municipal and private undertakings (Dias 1945; Campos 1998).

After a belated path of ups and downs, the modernizing wing finally succeeded in getting its plans within the authoritarian regime. Thanks to the enactment of two new electrification laws in 1944, clear synchronization was established between gross public investment in increasing the generation capacity, through the construction of a far-reaching system of integrated hydrographical basin power stations, and complementary measures undertaken on the distribution side, for example building a high-voltage transmission network connecting the hydroelectric stations to the main centers of consumption.

The surge in supply through large hydropower investments drew the country closer to European levels of commercial primary energy consumption; from lower than 200 kg of coal equivalent (kgce) per capita before the Second World War, there was a rise to over 500 kgce in the 1950s and over 1,000 kgce after 1969. This period has been seen as the precise moment of acceleration in economic growth, resulting in real convergence with the more developed nations, signaling Portugal's entry into a phase of "modern economic growth," according to Kuznets's classical definition (Batista 1997). But if the multiplier effects of the diffusion of modern energy seem undisputable, the surge of hydroelectricity was nevertheless short-lived, particularly when compared with Nordic countries. Thus, the goal of feeding the productive sector with cheap electricity soon reached its limit, and only through direct state cross-subsidization of industrial prices could the competitiveness of intensive sectors be maintained. By the 1960s, the effective abandonment of hydroelectricity-intensive industrial clusters already lay on the desks of ministers and entrepreneurs. As the price of cross-subsidization went up and the periods (days) of supply grew shorter, the government decided to authorize the production of nitrates, sulfates, ammoniac, and other goods using chemical/oil-based processes and dumping the hitherto existing priority of the electrolytic technologies responsible for the high consumption of electricity. After the shift from the electrochemical industries to chemical/oil-based processes, the major railway line also became "fuelized," and investment was concentrated thereafter in thermal power stations that burned liquid fuels, most notably the *Carregado* (1964) and *Setúbal* (1973) power stations, with installed capacities of 500 and 1,000 MW, respectively. In aggregate terms, the demand for hydroelectricity reached its peak in 1967 with a 45 % share of total primary energy consumption with its importance thereafter only ever declining: in 1973, hydroelectricity accounted for 30 % of consumption while only 15 % in 1982. Emulating the most-developed nations, the emphasis on a high intensity regime and the target of swift industrialization led a southern country, with hydroresources very different to those of Nordic countries, to follow the path of stimulating the production of cheap electricity (Madureira and Teives 2005; Madureira 2008). The paradox is that while the "nationalization of inputs" caused the economy to drift toward more intensive industrial sectors, what really sustained these intensive sectors shortly thereafter was the import of oil. Scarcities in hydropotential set a ceiling on electricity-adding strategies. Industrial diversification could thereby only proceed as long as the international price of oil remained low and accessible. The bill to be paid for this path to development suddenly reached a dramatic point in 1974 when the Organization of Petroleum Exporting Countries (OPEC) hiked the price of crude oil.

8.2.4 Energy Intensity in Brazil

Brazilian singularity in the domain of culture and way of life has fostered the allure of a creative and hybridized mix of modern and traditional society. No less original has been the country's trajectory in the domain of energy development.

Although Brazil pursued a course of heavy industrialization, for over thirty years, it managed to do so without raising the overall energy intensity. Even more remarkable, the industrialization process actually strengthened the country's position as the leader in advanced non-fossil fuel technologies like charcoal production, sugarcane processing, and giant hydroelectric power stations. Indeed, it was only after the consolidation of this inward-looking green economy that a new turn in fossil fuel exploration took shape and first triggered by the discovery of deep, pre-salt oil in the Santos Basin, the largest find worldwide in the twenty-first century (*Tupi* field 2007). The singular historical path resulted in an exceptionally diversified energy portfolio that widened the range of options available to decision makers. In sharp contrast with the standard agenda of developed nations, Brazil set the ground for the exploration of its wood, sugarcane, and national water resources when everyone else was living on cheap oil and nuclear expectations. Later, and just when the world turned to renewable energies and started talking about oil exhaustion and the spiral increases in fuel prices, Brazil hit one of the largest offshore reserves ever.

At the innermost heart of this "tropical" course of development lay the plentiful forestry riches and the historical importance of wood consumption in the overall energy balance. In reality, Brazil displayed one of the lowest levels of per capita consumption of commercial fossil fuels in the early twentieth century and barely surpassing countries such as El Salvador, Ecuador, Guatemala, and Nicaragua (Rubio 2010). Entrenched in agro-livestock production and in the exports of tropical goods, the whole economy relied on wood to serve the basic needs of its 40 million inhabitants, supplemented by minor imports of kerosene, fuel-oil, and hard coal in the coastal regions. During the early decades of the twentieth century, the political expression of the agricultural export model was summed up by the ironic expression "*política do café com leite*" (a milky coffee policy) in reference to the two major states that took turns holding office, *São Paulo* with its coffee plantation-based representatives and *Minas Gerais* with its cattle barons.

The first signs of political agitation and clamor against the status quo surfaced in 1922, with a military movement known as lieutenantism. Led by young army officers, lieutenantism sought to rouse the country around demands for political democracy and economic nationalism. Fair elections and a new electoral system, protection for domestic products and Brazilian industry, restrictions on foreign economic activities, and the safeguarding of natural resources became the banners for large sectors of public discontent. Defeated, albeit not curtailed, the movement experienced a revival when Getúlio Vargas took power after a coup that set aside the liberal–constitutional government. The event marked a complete reorientation of the Brazilian economy away from export agriculture and toward industrial expansion before later evolving toward aggressive import substitution industrialization (ISI). As of 1930, coffee and livestock interests were already on the wane and it was recognized that the Brazilian agricultural export economy had floundered. As in much of Latin America, the 1929 world depression abruptly accelerated the drop in the international prices of tropical exports, opening a cycle in which the terms of trade become clearly favorable to industrial goods. The supply of coffee was now running a surplus with income losses for plantation landowners and pressures on balance of payments to substitute high-priced imports and overcome impeding deficits. Despite government attempts to rescue the São Paulo plantations by resorting to quite unorthodox measures (buying and burning excess production), more than half of the land cultivating coffee was soon to fall out of production. Along with this collapse in trade, the downfall of the international monetary system, based on the gold standard, moreover forced Brazil and other large Latin American countries to opt for dual-exchange rates. In the short term, the depreciation of the official Brazilian currency made imports more costly, thereby providing an incentive to focus production on the domestic market; on the other hand, the non-official rate allowed free fluctuations for a variety of transactions, including capital exports and profit remittances (Bulmer-Thomas 1994: 198). In all these domains, the previous pattern of agricultural export growth held little potential: the shrinking international trade, the decreasing value of agricultural goods, and disconnected financial systems shifted the focus toward inward-looking policies. The accumulation of these economic factors, in turn, built up the momentum for a structural breakthrough as subsequently capitalized on by Getúlio Vargas.

While the old Republican regime perpetuated the oligarchic–liberal society, the new administration pursued a developmentalist policy, with a nationalist bent, authoritarian and with a state-centered drive. With the demise of the liberal state, the main instruments for economic policy had to be built from scratch and deliberately substituting spontaneous adjustments. Hence, an active monetary policy driven by budget deficits and the central bank rediscount policy substituted the flow of financing of external origins, while fiscal reforms replaced part of the state revenues arriving from tariff rates with internal taxes levied on income, property, and consumption. Public utilities, prices, and interest rates were controlled as of 1933. Once the state's hold on the economy was well underway, the Vargas government started to draft bills for industrialization projects, the majority of which targeted the building transportation infrastructures, the construction of new electric power plants, oil prospecting, and the founding of basic industries. Indeed, what distinguished the first phase in the Brazilian import substitution process is the fact that the state envisaged the promotion of light industrial goods, basically consumer non-durables in conjunction with the first definition of energy and energy infrastructure policies (Leite 2009; Mattei and Júnior 2009).

The "heavy" phase only arose out of the difficulties experienced during Second World War. Supply shortcomings thereafter pushed import substitution to new heights leading to the establishment of national steel, mining, caustic soda, and automotive corporations. Progress toward the first integrated metallurgical plant further stirred the creation of an iron–coal combine with the backing of coal mining industrialization in Santa Catarina linked to the National Metallurgical Company, in the Volta Redonda municipality. Despite the low quality of this fuel (maximum energy value of 4,500 kcal/kg presenting moreover a high content of ashes and sulfur) and despite the troublesome transport route from the coal basin in the South, the government made its use mandatory. Full exploration of national resources ensued, grounded on the extraction and transformation of mineral-coal for coke production, the exploration

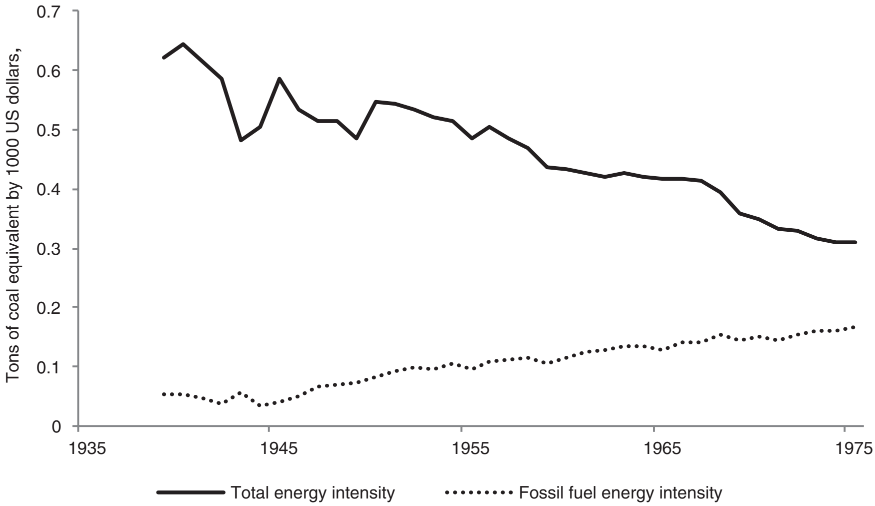


Fig. 8.2 Energy intensity in Brazil (1939–1970). Tons of coal equivalent by USD 1,000, estimated at 1990 Geary–Khamis dollars (Loeb 1953; Haddad 1975; IBGE. 1990; EPE 2007; Bolt and van Zanden 2013). *Note* Time series for firewood and charcoal used in industry estimated from linear interpolation from 5-year data intervals provided by EPE (2007) and according to vegetal extraction production indices provided by Loeb (1953)

of the large reserves of iron ore and the planning for giant hydroelectric power stations, set in motion by the São Francisco Hydroelectric Company (1945) and continued shortly thereafter with the launch of the first public utilities licensed to Brazilian states (*centrais estaduais*). The dramatic demise of Vargas in 1954 (suicide) did not trigger any change in either government policy or in the import substitution program, which was even taken a step further under the new presidency of *Juscelino Kubitschek* (1956–1961), who systematized public investment targets in a medium-term program known as the Goal's Plan (*Plano de Metas*).

In order to gauge the evolution of energy intensity over this period, Fig. 8.2 displays total energy intensity in a solid line aggregating time series detailing firewood and charcoal, ethanol, hydroelectric energy, hard coal, and brown coal, crude oil, natural gas, plus fuel-oil, gasoline, and kerosene imports. On the other hand, the dashed line represents only the fossil fuels components of Brazilian consumption adding oil by-products and crude, natural gas and coal. Starting with this latter indicator, it seems quite significant that the amount of non-renewable energy per unit of GDP remained flat during the Second World War and only initiating its ascendant trajectory once the surge in heavy industrialization had begun in the second phase of the Vargas economic policy. For the rest of the period, fossil fuel demand was always greater than the value added by Brazilian output, a fact that supports the idea of modernization as a switch to fossil fuel energy. Conversely, should all energy forms be taken into account, the overall outlook takes a notable turn. What the Fig. 8.2 solid line illustrates is how, in effect, total energy decreases in intensity during the light, heavy, and heavy-planned phases of the Brazilian import substitution industrialization. Wood consumption and hydroelectric energy, key drivers for this trend, continued to grow in absolute terms during the overall time frame (wood alone, however, reached its plateau in 1965). The crux of the matter is that, unlike fossil fuels, the demand for wood grew at a slower pace than that of GDP. Logically, given the success of Brazilian industrialization and modernization, publicly acknowledged as the “Brazilian miracle,” the net outcome could only be the global decline in energy intensity. Another consequence of the slower pace in wood consumption is that oil outpaced this traditional energy form in the 1970s. To a large extent, the “tropical” historical path was driven by the leveraging of opposite forces: the slow rise in wood consumption cancelled out the initial effects of mounting fossil fuel consumption.

Rather than being just a sign of backwardness or underdevelopment, the persistence of wood in the global energy balance was, in itself, as much a sign of tradition as a hallmark of modernity. Tradition because it is well known that firewood and charcoal have subsisted particularly in the poor and destitute segments of society who, for decades, resorted to inefficient stoves adapted to consuming wood waste, and very often because no other cheap power source was available as was the case in the states of North-Eastern Brazil. Tradition also in the cultural sense because well-off social milieus maintained a reasonable per capita demand for wood in the form of charcoal consumption stemming from the popularity of private gatherings for barbecues (*churrascos*). Every grocery shop or supermarket in Brazil invariably has a corner of their retailing space stacked with bags of barbecue charcoal.

On the opposite side, wood subsistence is moreover a hallmark of the modernization prompted by the Vargas-Kubitschek conviction that the metallurgical industry was vital to the success of industrialization and correspondingly had to be fuelled by national energies. Defense of the vegetal route (charcoal) for pig iron production was likewise endorsed by technicians and engineers, who foresaw in the abundant national forests a natural advantage worth harnessing. The promotion of eucalyptus plantations, the technological improvement of the old charcoal stoves so as to enhance the wood/charcoal conversion ratio (with the replacement of *medas* by *fornos de rabo quente*), the introduction of scientific reforestation practices and the attribution of fiscal incentives to reforestation were among the most visible signs of the public sponsoring of this wood-fuelled metallurgical program. In the last thirty years, the overall tendency has been for a slow decline in wood demand in the domestic milieu counterbalanced by the recovery in metallurgy-related industrial usages. As of 2005, core wood demand arose from the charcoal-fuelled industries (43 %), followed by residential consumption (29 %), and then other industrial uses (20 %) Perhaps still more surprising, the globally declining trend in the years between 1965 and 1998 has recently been reversed and wood demand for energy production is again on the rise (Brito 2007; Uhlig 2008; Morello 2009).

Finally, it is worth remembering that systematic recourse to sugarcane alcohol as an automotive fuel in Brazil dates back to the 1930s and that the first attempts at energy exploitation via vegetable oils and fats (babassu, coconut, castor seed, and particularly cotton seed) range back as far as the 1940s. Although these endeavors at first remained relatively marginal, for instance, just about attaining 5.5 % of petrol demand, production boomed after the first and second oil shocks (1974 and 1979), and sheltered under the umbrella of state financed programs like "*Pro-álcool,*" "*Pró-óleo,*" and "*Produção e uso de Bio-Diesel.*" The environmental costs of the massive reconversion of forests and agricultural lands to "green" energy production have nonetheless subsequently been criticized within the scientific community and by environmental groups. Regarded through the lens of the long term effects the "green" was darker than expected (Pousa et al. 2007; Leite 2009).

Ultimately, most of the events herein presented occurred before the concept of energy intensity had itself become soundly established in political, business, and scientific circles. Through rather different processes and circumstances, the story of Soviet Russia, Spain, Portugal, and Brazil reveal a shared intent on embarking on a process of accelerated industrialization so as to swiftly catch up with developed nations. Perhaps it is also not just a coincidence that in all these historical cases, the targeted shift in the composition of economic activity was paralleled by a reinforcement of the authoritarian role of the central state. The causal linkage derives from the fact that bold policies require strong state apparatuses and strong state action. Likewise, the sequence of events demonstrated how heavy industrialization projects were the chief arguments deployed by leaders such as Stalin and Vargas to strengthen their personal grip on power.

Another common trait was the reinforcement of a nationalist vision and the consequent emphasis upon the maximization of natural resources. This proved crucial in the unfolding of events, because while the Soviet path relied on coal–steel combines, Portugal and Spain dreamed about the full exploitation of their hydroelectricity, while Brazil was betting on harnessing all the immense potential of its agricultural lands and forests. The consequences in terms of pollution and environmental damages were accordingly quite different. Stirred by the nationalist discourse a new group of skilled technicians, engineers, agronomists, and geologists came to the fore extolling the capacities of modern technology and quite often in rather delusionary ways. With the exception of Soviet Union, ravaged by the 1929–1931 and 1937–1939 purges, this group formed a core of social support for the continuity of import substitution policies, ingraining networks of technical skills in the public administration and in state-owned enterprises. However, by the 1960s, in all of the above-mentioned nations, the background to support for visionary industrial plans grew less favorable as cost competitiveness considerations, the tightness of natural endowments, balance of payments constraints, and a more liberal or "open to the west" approach precluded the scope for changes in mega projects.

8.3 The Environmental Kuznets Curve

The study undertaken by one of the first post-war environmental economists, Harold Barnett (1950), on the estimation aggregate energy/GNP ratios marked a turning point in the recognition of the energy intensity concept. From the USA, the indicator spread to Canada, Europe, and onto developing countries, filling a noticeable role in the groundbreaking field of energy economics. Coincidentally, energy intensity began asserting the importance of efficient industrial practices just as the Middle East oil boom loosened most concerns about conservation-oriented practices and resource savings. Unprecedented low prices coupled with technological utopias about nuclear power conveyed an erroneous impression about endless energy and boundless resources. Such ideas continued running against the grain until the first oil shock brought clear evidence on the need to think in terms of barrels of oil per wealth produced. The disquieting message was that energy ought to be productively put to good usage and all the more so when the finiteness of uranium and oil dawned on public opinion and political circles.

Alongside the changes caused by the brutality of economic fact, people were also more attuned to ecological issues in the 1970s. And, as mentioned before, the intensity indicator grasped better the energy burden of income levels while quite badly expressing the environmental costs of development. High intensities, for instance, though generally associated with anti-ecological development patterns could nonetheless evolve to a relatively environment-friendly economy. It suffices to think that should a country increase its share of "clean energy" in the overall productive capacity, the impact of energy consumption upon pollution emissions would be lessened. This was the case with Canada that managed to achieve high levels of throughput by resorting to hydroelectricity, nuclear power, and renewables. Thus, the point remains that energy intensity represents only a flawed indicator of pollution levels.

Within this line of reasoning, a more direct means of disclosing the environmental costs of energy usage involves the substitution of "Energy" in the basic formula E/Y by "Pollution," where P would stand for a relevant measure of pollution and Y, for a relevant measure of activity or output. Although this change makes the effects of economic growth upon environmental quality clear, a further problem stems from just what indicator to choose out of the different aspects characterizing ecosystems. Clean energies such as hydropower or nuclear power, albeit almost free from pollutant emissions, incur other types of negative impacts upon the environment. Currently, the issue tends to somehow get simplified owing to the overbearing influence of climate change, which pushes all other indicators into the background to attribute greatest prominence to the direct cause of global warming: greenhouse gas emissions. Therefore, where the purpose is to measure the pollution caused by the full range of human activities, and not strictly locally determined, relevant to greenhouse gas effects, the best indicator for P/Y would become the emission of metric tons of carbon dioxide (CO_2) per capita/GDP per capita. Furthermore, it hardly needs saying that this specific indicator has displaced others in applied research. Yet, the issue remains that only one aspect of pollution effects is captured by this statistical analysis. A completely different breadth of pollution problems would be grasped by applying the ratio of Kg of sulfur dioxide (SO_2) per capita/GDP per capita. In this case, the P/Y formula underpins the impact of sulfur dioxide, which does not contribute to greenhouse gas formation but rather to acid rain precipitation. Moreover, since coal and petroleum contain sulfur compounds, the sulfur dioxide/income ratio is a good gauge to measure the concentration of fossil fuel industrial pollution at a regional level, as opposed to industrial and residential pollution over wider geographical areas (through emissions of carbon dioxide).

Overall, it has proven difficult to find a statistical synthesis able to encompass the several dimensions of environmental quality. Whatever the dominant choice, the matter of fact is that energy intensity (E/Y) has yielded to its gloomy kin: pollution intensity (P/Y). And this institutionalization of yet another energy indicator brought in its wake a wholly new argument: the environmental Kuznets curve (EKC).

The EKC is a theoretical hypothesis for explaining the empirical discovery that environmental quality falls as economic development proceeds, but, at a certain point of income, environmental quality begins improving again and in a sustainable way. Over time, pollution indicators consequently reveal an inverted U-shaped curve that is a function of the evolution of income per capita. The main justification for dubbing the phenomenon the "Kuznets curve" lies precisely in the design of the inverted U-shaped form, which was first proposed in 1955 by the leading economist Simon Kuznets to explain the historical unfolding of income inequality. Briefly during the first phase in "modern economic growth," income inequality increases up to the moment when a trend reversal opens up a phase of continuous income leveling out that persists into more mature stages of economic development. Forty years after the Kuznets stylized discovery, two other economists, Grossman and Krueger (1991) found the same inverted U layout in environmental time series: a curve depicting environmental degradation and pollution increasing in the early stages of economic growth and the reversal of this trend past a certain level of per capita income. Because this pathbreaking discovery was made in the context of the debate over the NAFTA trade agreements before then being further popularized by the 1992 World Bank World Development Report, the concept immediately acquired a political reach subsumed in the message that economic growth and trade liberalization ultimately prove beneficial to the environment. In this vein, an opinion column in an Australian newspaper filtered down the argument to its shortest form: "growth is the key to protecting the environment, not its enemy" (quoted in Carson 2010: 5).

The standard theoretical argument for why pollution levels might fall as GDP per capita rises somehow repeats the aforementioned causes of structural change in the economy (composition effects) and technological improvements (efficiency effects). However, a new argument also enters the scene in the form of changing social preferences: With mounting incomes, the demand for environmental quality increases at a more-than-proportional rate. In other words, affluent societies display an income elasticity of "willingness to pay" for environmental quality greater than one (Jacobsen and Hanley 2009). This translates into political pressure for tougher environmental policy driving down the level of pollution per unit of GDP.

Major critics of the EKC theory suspect that the model misses the countries of greatest interest and pays little attention to the actual statistical properties of the data used. On theoretical grounds, two types of objections have been made to the idea that environmental quality tends to be positively, not negatively, correlated with income in wealthier countries: First, EKC neglects the fact that economic activity is inevitably environmentally disruptive in some way. "Satisfying the material needs of people requires the use and disturbance of energy flows and materials stocks. Therefore, an effort to reduce some environmental impacts may just aggravate other problems" (Stern 2004: 1426). Second, developed countries might simply further encourage the outsourcing of dirty production to developing countries. In this "race to the bottom" scenario, developing countries will experience greater difficulties in efforts to reduce emissions (Jaffe 1995).

8.4 Toward a Heavy Post-Industrial Society?

Overhanging much of the considerations above is the conception of a post-industrial society. Research is still attempting to figure out the consequences of the shift from manufacturing-based economies to service-based economies, especially the extent to which continuous innovation fed by theoretical knowledge and information technologies can counterbalance the dynamism of job creation in agriculture and industry. One of the core ideas in this approach is that the change toward service industries brings about the decoupling of economic growth and energy usage. By most accounts, the post-industrial has ample room to reduce energy intensity and, throughout, prop up environmental friendly productive activities. Even though this expectation is not false, Stern and Cleveland (2004; Stern 2011) call into question the sustainability of energy decoupling from economic growth. According to them, the service and household sectors might, after all, prove as highly energy intensive as the other sectors of the economy. The gadgetry, digital, intensive facet to services, and residential activities help to explain why late comers to modernization experience the shift from agriculture and industry into services at the critical juncture in which precisely the service sector is increasing its own energy intensity. By this means, there is a real chance that the compositional saving effects are somehow cancelled out in the transition. The country moves toward a post-industrial society but sees little improvements in its energy intensity. Were this pattern identified in some economies (Ramos-Martin 2001) to be replicated elsewhere, we might be heading toward a heavy post-industrial society.

References

Ang BW, Liu FL, Chew EP (2003) Perfect decomposition techniques in energy and environmental analysis. Energy Policy 31(14):1561–1566

Ang BW, Liu FL, Chung H (2004) A generalized fisher index approach to energy decomposition analysis. Energy Econ 26(5):757–763

Antolín F (2003) Global strategies and national performance: explaining the singularities of the Spanish electricity supply industry, business and economic history on-line, Vol 1. Available at https://www.langinnovate.msu.edu/~business/bhcweb/publications/BEHonline/2003/antolin.pdf. Accessed 14 June 2004

Barnett HJ (1950) Energy use and supplies, 1939, 1947, 1965. Bureau of Mines information circular 7582. US Department of the Interior, Washington
Batista D et al (1997) New estimates for Portugal's GDP 1910–1958. Banco de Portugal, Lisbon
Bertoni R, Román C, Rubio M (2009) El desarrollo energético de España y Uruguay en perspectiva comparada, 1860–2000. Revista de Historia Industrial 41(3):161–196
Bolt J, van Zanden JL (2013) The first update of the Maddison project; re-estimating growth before 1820. Maddison Project Working Paper 4. Available at http://www.ggdc.net/maddison/maddison-project/data.htm. Accessed 12 June 2013
Boobbyer P (2000) The Stalin era. Routledge Press, London
Brito JO (2007) O uso energético da madeira. Estudos avançados 21(59):185–193
Bulmer-Thomas V (1994) The economic history of Latin America since independence, 2nd edn. Cambridge University Press, Cambridge
Campos E (1998) Textos de Economia e Política Agrária e Industrial 1918–1944. Banco de Portugal, Lisbon
Carson R (2010) The environmental Kuznets curve: seeking empirical regularity and theoretical structure. Rev Environ Econ Policy 4(1):3–23
Darmstadter J, Dunkerley AJ (1977) How industrial societies use energy: a comparative analysis. Johns Hopkins Press, Baltimore
Davies RW (2003) The Stalin–Kaganovich correspondence: 1931–1936. Yale University Press, New Haven
Davies S, Harris J (2005) Stalin: a new history. Cambridge University Press, New York
Davies RW, Harrison M, Wheatcroft SG (1994) The economic transformation of the Soviet Union, 1913—1945. Cambridge University Press, Cambridge
Delgado JG (1994) La industrializacion y el desarollo económico de Espana durante el Franquismo. In: Nadal J, Carreras A, Sudria C (Eds) La Economia Española en el siglo XX: Una Perspectiva Historica, Ariel, Barcelona, pp 164–189
Dias JF (1945) Linha de Rumo. Clássica Editora, Lisbon
Elliot IF (1974) The Soviet energy balance. Natural gas, other fossil fuels and alternative power sources. Praeger Publishers, New York
Ellman M (2008) The political economy of Stalinism in the light of the archival revolution. J Inst Econ 4(1):99–125
EPE—Empresa de Pesquisa Energética (2007) Balanço energético nacional 2007: ano base 2006. Ministério de Minas e Energia, Rio de Janeiro
Etemad BJ, Luciani P (1991) World energy production 1800–1985. Librarie Droz, Geneva
Glete J (1986) Demand pull or technology push: preconditions for the development of the Swedish heavy electrical industry. In: Cardot F (ed) Un Siècle d' Électricité dans le Monde. PUF, Paris, pp 243–251
Gregory PR (2003) The political economy of Stalinism: evidence from the Soviet secret archives. Cambridge University Press, Cambridge
Grossman GM, Krueger AB (1991) Environmental impacts of a North American free trade agreement. National Bureau of Economic Research Working Paper 3914. NBER, Cambridge
Haddad C (1975) Crescimento do produto real brasileiro—1900/1947. Revista Brasileira de Economia 29(1):3–26
IBGE—Instituto Brasileiro de Geografia e Estatística (1990) Estatísticas históricas do Brasil: séries econômicas, demográficas e sociais. IBGE, Rio de Janeiro
Jacobsen JB, Hanley N (2009) Are there income effects on global willingness to pay for biodiversity conservation? Environ Resour Econ 43(2):137–160
Jaffe AB et al (1995) Environmental regulation and the competitiveness of US manufacturing: what does the evidence tell us? J Econ Lit 33(1):132–163
Josephson P et al (2013) An environmental history of Russia. Cambridge University Press, Cambridge
Khlevniuk O (2008) Master of the house Stalin and his inner circle. Yale University Press, New Haven
King F (1999) The Russian revolution and the idea of a single economic plan 1917–1928. Revolutionary Russia 12(1):69–83

Krylov NA et al (1998) The oil reserves and resource base of Russia. In: Krylov NA, Bokserman AA, Stavrovsky ED (eds) The oil industry of the former Soviet Union. Gordon and Breach science Publishers, Amsterdam, pp 1–68
Leite AD (2009) Energy in Brazil: towards a renewable energy dominated system. Sterling, VA
Lenin VI (1920) Our foreign and domestic position and party tasks. Speech delivered to the Moscow Gubernia conference of the R.C.P.(B.). Available at http://www.marxists.org/archive/lenin/works/1920/nov/21.htm. Accessed 28 April 2013
Loeb GF (1953) Números Índices do Desenvolvimento Físico da Produção Industrial no Brasil—1939—1949. Revista Brasileira de Economia 7(1):31–36
Madureira NL (2008) When the South Emulates the North: energy and nationalism in the twentieth century. Contemp Eur Hist 17(1):1–21
Madureira NL, Teives S (2005) Os ciclos de desenvolvimento. In: Madureira NL (ed) A história da energia. Portugal, 1890–1982. Livros Horizonte, Lisbon, pp 15–45
Markevich A, Harrison M (2010) Great war, civil war, and recovery: Russia's national income, 1913 to 1928. Working papers working papers from center for economic and financial research, w0146. Available at http://www.cefir.ru/papers/WP146.pdf. Accessed 12 June 2013
Mattei L, Júnior JAS (2009) Industrialização e Substituição de Importações no Brasil e na Argentina: Uma Análise Histórica Comparada. Revista de Economia 35(1):93–115
Mitchell BR (1981) International historical statistics: Europe, 1750–1975, 4th edn. Facts on File, New York
Morello TF (2009) Carvão vegetal e siderurgia: de elo perdido a solução para um mundo pós-Kyoto. Dissertation, University of São Paulo
Mulder P, De Groot HLF (2012) Structural change and convergence of energy intensity across OECD countries, 1970–2005. Energy Economics 34(6):1910–1921
Myllyntaus T (1992) The role of industry in the electrification of Finland. In: Tréde M (ed) Électricité et Electrification dans le Monde. PUF, Paris, pp 235–249
Myllyntaus T (1995) Kilowatts at work. Electricity and industrial transformation in the Nordic Countries. In: Kaijser A, Hedin M (eds) Nordic energy system. Watson Publishing, Canton, pp 114–116
Percebois J (1989) Economie de l'energie. Editions Economica, Paris
Pousa PAG, Santos ALF, Suarez PAZ (2007) History and policy of biodiesel in Brazil. Energy Policy 35(11):5393–5398
Ramos-Martın J (2001) Historical analysis of energy intensity of Spain: from a "conventional view" to an "integrated assessment". Popul Environ 22(3):281–313
Rodriguez IB (2007) La industria eléctrica en España (1890–1936). Banco de España, Madrid
Rubio M et al (2010) Energy as an indicator of modernization in Latin America, 1890–1925. Econ Hist Rev 63(3):769–804
Russia's Five Year Program (1930). The Oil and Gas Journal, December 11, (1930): 128, 151
Schurr SH, Netschert B (1960) Energy in the American economy, 1850–1975—Resources for the future. Johns Hopkins Press, Baltimore
Stalin J (1931) The tasks of business executives. In: Stalin J (ed) Works, vol 13. Foreign Languages Publishing House, Moscow, pp 31–44
Stern DI (2004) The rise and fall of the environmental Kuznets curve. World Dev 32(8):1419–1439
Stern DI (2011) The role of energy in economic growth. Ann N Y Acad Sci 1219:26–51
Stern DI, Cleveland CJ (2004) Energy and economic growth. Rensselaer Polytechnic Institute, Working papers in economics number 0410. Available at http://www.economics.rpi.edu/workingpapers/rpi0410.pdf. Accessed 12 April 2012
Sudría C (1994) Un factor determinante: la energía. In: Nadal J, Carreras A, Sudria C (Eds), La Economia Española en el siglo XX: Una Perspectiva Historica, Ariel, Barcelona, pp 313–363
Taylor PG, d'Ortigue OL, Francoeur M, Trudeau N (2010) Final energy use in IEA countries: the role of energy efficiency. Energy Policy 38(11):6463–6474
Thue L (1992) The state and the dual structure of the power supply industry in Norway, 1890–1940. In: Tréde M (ed) Électricité et Electrification dans le Monde. PUF, Paris, pp 228–234

Tryon FG (1927) An index of consumption of fuels and water power. J Am Stat Assoc 22:271–282
Uhlig A (2008) Lenha e carvão vegetal no Brasil: balanço oferta-demanda e métodos para a estimação do consumo. Dissertation, Universidade de São Paulo
Villar EH (2010) [1921] El valor geográfico de España. Universidad de Barcelona, Barcelona

Chapter 9
Levelized Electricity Costs

Abstract The concept of levelized energy costs responds to the necessity of disclosing the future consequences of present decisions. This incorporates what price should be charged per unit of energy sold in order to recover the total life cycle cost of energy production. This chapter charts the effectiveness of levelized cost formulas over time depicting how the perspectives for electricity generation in the UK and in France have evolved. While cost comparisons in the 1950s were almost exclusively focused on the eventuality of nuclear power catching up with the competitiveness of coal power, in the 1960s, things became thornier as nuclear hopes fell short of expectations and also because a new competitor—oil-fired power stations—made significant strides in the marketplace. The final conclusion points out that whereas in periods of relative price stability, all concerns center on reductions in the part of the formula accounting for capital costs, in periods of higher volatility, the focus switches to trends in the forecasting of fuel prices. Depending on the historical circumstances, the levelized energy costs formula was viewed either through the lens of "yesterday's costs"—the most up-to-date estimates of existing equipment and installations—or through the lens of "tomorrow's costs"—the likely evolution of fuel prices.

9.1 Engineering the Economy

The levelized electricity cost concept responds to the necessity of disclosing the future consequences of present investment decisions and thus reflecting what price should be charged per energy unit sold in order to recover the total costs of the energy production life cycle. In practical terms, the levelized costs indicator has proven an important tool in the domain of investment planning as it enables cost-benefit comparisons between different generations of technologies, between competing fuels, and between alternative expenditure schedules. Owing to the mix of

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_9,

engineering knowledge and economic skills required to assess likely future energy costs, this methodology became a standard in engineering economics.

The core idea is to discount the total lifetime costs of delivering energy by converting the total investment outlay into annualized costs. Equation 9.1 displays the relevant parameters with the inputs of capital costs, operation and maintenance costs (O&M costs), alongside fuel costs in the numerator and with the output of energy production in the denominator. To convert total capital costs into a constant stream of annualized payments over the lifetime of the fixed investment, a static annuity method called the capital recovery factor (CRF) is applied. The CRF accounts for the specific investment time-cycle and the interest rate (t and i in Eq. 9.2, respectively).

$$\text{Levelized Energy Costs} = \frac{\text{Costs}_{\text{Capital (CRF)}} + \text{Costs}_{\text{O\& M}} + \text{Costs}_{\text{Fuel}}}{\text{Energy Production (by year)}} \quad (9.1)$$

$$\text{Capital Recovery Factor (CRF)} = \frac{i(1+i)^t}{(1+i)^t - 1} \quad (9.2)$$

The word “levelized” therefore expresses the principle that all quantities of inputs and all quantities of outputs should be divided into equal proportions over the plant’s operational lifetime. The sum of the annualized capital, operation, maintenance, and fuel costs divided by the expected amount of energy to be supplied yearly enabled prospective investors or state agencies to forecast; for instance, the costs per KWh of each power station under commission as well as which of the technological options available (gas turbines, coal plants, oil plants, or nuclear stations) presented the most competitive prices.

Competitiveness, however, only set the background to such decisions. Factors other than economic engineering costs also came into play and all the more so when decision-makers proved sensitive to the balance between corporate interests, regional interests, political networks, and environmental concerns. Furthermore, because some of the basic premises of the cost formula could be called into question, the debate sometimes drew technological-minded individuals into the discussion of democratic options. In effect, the levelized energy costs equation relied not only on data from construction and equipment costs, labor, and raw materials but also on educated hypotheses about the rate of interest (i), the plant lifetime (t), and the capacity factor, which determined the amount of energy due to be produced (*Energy production*). As with parametric assumptions in which the more the assumptions prove correct, the more accurate and precise the estimate becomes, the “breakeven” value that a power provider promises to charge is also largely dependent upon how sharp these points of departure are. Owing to this, the rate of interest, the plant amortization lifetime, and the capacity factor are called “ground rules”. They have varied according to techno-political conditions, the maturity of the technology and the different tax and amortization guidelines in effect among countries.

Historical ground rules for power plant amortization (1968)		
	Plant lifetime (years)	Plant lifetime (years)
	Nuclear power stations	Coal power stations
UK	20	25
France	20	30
USA	30	35
Japan	20	16–17
West Germany	16–17	16–17
Canada	30	30
Italy	20	20
Spain	20	25

9.2 Nuclear Power Costs: Struggling with a Bad Reputation

9.2.1 The Policy of Secrecy

Energy forecasting received a tremendous boost in the wake of World War II. Strong growth, trade liberalization, enhanced economies of scale, industrial innovation, and the unprecedented diversification in energy sources made forecasts not only more necessary but also more feasible. This was the period in which experts maintained that electricity consumption would double each and every 10 years, a rule deduced from the observation of a steady growth rate of 7 % per year. Moreover, a new era was about to dawn in engineering planning with the competing technology of nuclear power arriving to dispute the electricity generation markets.

Hitherto, the necessity to compare alternative investment decisions was much less explicit since the different technologies were in effect complimentary. For instance, dam produced hydroelectricity tended to meet the base load and required additional investments in coal power stations to meet peak capacity demand. In contrast, the first civil nuclear power program aimed directly at seizing the most remunerative electric system payoffs—generation for base load—at the expense of other power sources. In countries otherwise not well endowed with exploitable rivers, such as Germany, Japan, the UK, or the USA, the nuclear option posed a threat to established interests in coal mining and/or coal stations; while in nations that had fully developed dam construction programs, such as France, Italy, Spain, and Sweden, nuclear power programs threatened the position of hydroelectricity as the mainstay for base load and even the scope for future investments in coal stations. No wonder that the refinement of the methodology of levelized energy costs was particularly endorsed by nuclear engineers eager to prove that, sooner or later, nuclear would overcome the competing technologies in the supply of cheap energy.

This chapter traces the evolution of cost analysis among the forerunners of nuclear power in Europe: the UK and France. In both countries, the end of the war set in motion a race to master atomic power, whose ultimate goal was the production of atomic bombs as a means to ensure long-term security, while also fostering cutting-edge technology with an eye to exports and national grandeur. These national goals were endorsed by a wide majority of political forces, first set out in the UK by the war cabinet of Winston Churchill and pursued enthusiastically under Clement Atlee Labour's government (1945–1951) and advanced a step further when the Tories returned to power with Churchill again serving as prime minister (1951–1955). Likewise, French atomic plans were set in motion by the broad reaching coalition led by General de Gaulle (1944–1946) and pressed ahead with by the Socialist President Vicent Auriol (1947–1954) with the backup of prime ministers as diverse as the Socialist Paul Ramadier, the leader of the center-left *Mouvement Républicain Populaire* Robert Schuman, or the head of the Conservative Party Antoine Pinay.

In spite of the far-reaching political convergence around the issue, the true military goals of these civilian atomic programs were for a long time concealed from public opinion. Even more strikingly, when detailed investment plans were subsequently submitted to parliaments, the basic principle that "fissile material produced in a nuclear power station might be used for military purposes" was swept under the catchall general declarations about the "peaceful applications of nuclear energy" and the inescapability of developing "the energy of the future" (NA. EG-1-60 1955; NA. Cabinet Papers 1955; Gowing 1978; Hecht 1998). At this stage, competitive nuclear power represented a still distant pledge and the development of civilian applications was, above all, a cover for the real goal of advancing nuclear weapons technology. This subterfuge resulted in an odd situation in which parliaments voted for investment plans in atomic generation while ignoring that the costs were ultimately based on experimenting with dual-purpose reactors optimized for the non-commercial military production of plutonium. From a governmental standpoint, however, the secrecy seemed justified by the escalation of bipolar rivalries and the willingness to conceal key strategic information from the Soviet Union. Curiously, all precautions, behavioral confidentiality, and compulsory rules for information exchange did not prevent the mysterious disappearance of secret files from government offices. Worse still, British nuclear facilities had been infiltrated at the highest levels by scientists that passed on top secret information to the Soviet intelligence services (NA. POWE 14-740 1954–1955; Cooke 2009: 70–84).

9.2.2 Atomic Piles and Dual-Purpose Reactors

The evolution from experimental atomic piles to dual-purpose reactors and to single-purpose electricity power stations took place in a singular atmosphere featuring a characteristic blend of developmental optimism and anxiety. Technological utopia and history-making awareness were very much in the air, at least as far as the politics of international competition and military rivalry. The upshot was swift technological development in which each technological stage began before the former had been fully consolidated. Such a pattern of rushing for results was counterbalanced by a tactic agreement to leave the fundamental technological paths open and exploring a wide range of options through the extension of basic experimental research.

As early as 1946–1947, the British started running experimental atomic piles at the Royal Air Force facilities in Harwell, Oxfordshire (named GEEP and BEPO) to further test irradiation and materials, check operating procedures, and produce radio-isotopes. Almost simultaneously, the authorities moved ahead to establish a large plutonium-producing pile and associated chemical plants for the separation and purification of irradiated uranium in a factory located on the west coast of Cumberland, known as the Windscale Works. This military agenda pressure on the scientific field yielded only a first round of ill-founded outcomes: The chemical engineers swiftly found that the irradiated uranium slugs obtained from Windscale piles contained less plutonium than theoretically expected. Worst of all, the plutonium recovered was heavily contaminated with dangerous radioactive fission products and far from attaining the purification levels required for the stringent specifications in effect for weapons-grade plutonium (Roberts 1999: 40). What ensued was a cluster of learning by doing adjustments aiming mostly at improving the enclosing of the fuel elements, preventing the deformation of uranium under irradiation and the better tuning of irradation times (Hinton 1957).

When weapons-grade plutonium production was back on track and with the deadline set for the explosion of the first British nuclear device fully achieved (October 1952), the military and political circles became convinced that the time was ripe to push ahead with a new goal for the nuclear program. This incorporated the fulfillment of the promise of large reactors and their commercial takeoff. By the end of 1952, a perfect constellation of factors prompted the decisive go-ahead for the building of sizeable nuclear reactors.

On the demand side, the military and the government developed new approaches to world security aired in a report written by the Chiefs of Staff of the three branches of the British armed forces and whose main conclusion thoroughly supported the deployment of atomic weapons and air power to act as a deterrent against the country's massive numerical inferiority in conventional forces. More plutonium and the expansion of the atomic bomb making program was the logical corollary of this strategy. Meanwhile, on the "supply side", a small research design team working at Harwell, had just finished a technical blueprint for a 35 MW nuclear power station, resolving troublesome issues such as the choice of the coolant, the can's materials, and the device for loading fuel rods. On top of that, the industrial group director, Christopher Hinton, responsible for Windscale facilities, bid to take on the job, lending his managerial weight to the civilian atomic power project. Finally, in the political realm, the nuclear cause secured a powerful ally within the highest spheres of government thanks to the return of Churchill, along with his inseparable scientific adviser, Lord Cherwell, to the halls of power. In fact, Lord Cherwell proved to be one key decision-maker in the overall institutional arrangement that finally settled consent for the new nuclear station to go ahead (NA. POWE-14-741 1953).

At the crossroads of all these developments, the decision to build the first dual-purpose reactor at Calder Hall proceeded swiftly along the principle that the reactor's design should be optimized for extracting the largest possible output of weapons-grade plutonium, with electricity generation a by-product of this goal. The installed capacity was boosted to a total of 92 MWe and distributed across twin reactors identified as Calder Hall "A", of which, 65 MWe were destined to supplying the commercial electrical grid. In September 1956, the final power station components, the turbines, were coupled up to the atomic unit, and the final tests completed. The nuclear facility was prepared for the delivering of energy to the electric grid and the Queen herself was on hand to pull the switch in a public celebration proclaiming "the threshold of a new era."

The French evolution confronted similar technical challenges even while leapfrogging to the completion of experimental atomic piles for dual-purpose reactors, without passing through the middle stage of military plutonium production prototypes.

After a learning experimental phase, the appointment of a young state secretary for atomic energy, Felix Gaillard (later to become Prime Minister), coupled with the guidance of a group of engineers from the institution of the Corps de Mines, spurred the engineering–nationalist vision of atomic power and succeed in getting parliamentary approval for a 38 million francs budget to be invested over a five-year plan (1952–1957). In detail, this envisioned the construction of two large atomic piles, G1 and G2, yielding the heat equivalent to 40 and 80 MW and contemplating the introduction of an air/gas circulation refrigeration system around the uranium bars (Dürr 1996).

Just as the installation of the G1 facilities at the Marcoule site in Côtes-du-Rhône region were almost finished, the overall strategy took an unexpected turn. Pierre Ailleret, the head of *Electricité de France* (EDF) Research Division, stepped in and suggested the addition of a classical power plant to the military-scientific plans in order to recover some of the heat that would otherwise be uneconomically dissipated. This goal was achievable without interfering with the military target of obtaining 15 kg of plutonium by the end of the five-year plan. Furthermore, as long as the add-on further justified the interests of atomic funding, it came to be accepted without opposition by the nuclear authorities. Hence, the French nuclear industry experienced a leap forward, scoring a major success when the first supply of "peaceful" energy fed into the EDF's distribution grid and almost simultaneously to the Calder Hall reactor on the other side of the channel coming onstream. However, the dual purpose of G1 remained quite uneven. The nuclear station had been designed for plutonium production and little other than the detached generation facilities were reshuffled anew. As a result, the global energy balance turned out negative because G1 consumed more energy than it actually produced: The auxiliary blowers used to refrigerate the uranium rods required 8 MWe of power while the generator feeding the national grid held a capacity of just 5 MWe.

Regardless of the last-minute adaptations of the French G1, its conception resembled the British in terms of optimizing weapons-grade plutonium, meaning that "dual-purpose" reactors could but recover only a small percentage of the potential energy contained in the natural uranium. In other words, the technology of dual reactors assumed the interruption of fission in the fuel rods when there was still a large amount of energy capable of release and a large amount of enduring chain reactions. To understand this, one must bear in mind that French and British reactors made use of the available fissile natural uranium material, which contains 0.7 % of the isotope U_{235}, the element bearing the property of releasing energy through the division of the nucleus. It is this substance that is the object of "fission" whenever a neutron hits an U_{235} atom causing the split of the nucleus and releasing more neutrons. Some of these additional neutrons are, in turn, absorbed by other U_{235} atoms producing more fission. The prospect for sustaining a chain reaction of this kind is greatly enhanced when the neutrons are slowed down by means of contact with a material with a light nucleus known as moderator (graphite in France and Britain). Once the velocity is slowed down to thermal energies, neutrons are more readily absorbed by the fissile U_{235} atoms. However, even resorting to a moderator, some neutrons come to be absorbed by a different isotope, U_{238}, which does not produce fission but rather undergoes a transformation to become U_{239} before eventually changing again to fissile P_{239} or plutonium. Overall, the number of P_{239} fissile nuclei created is less than the number of U_{235} destroyed and thus resulting in a conversion factor of less than unity.

As the logic above suggests, a "dual-purpose" technology aims at sustaining the U_{235} chain reaction as long as it feeds the U_{238}–U_{239}–P_{239} conversion. However, while that formed the bottom line of initial endeavors, scientists soon realized that it was difficult to maintain P_{239} stability during irradiation, because, over time, P_{239} absorbed ever more neutrons and changed into "poison" isotopes P_{240} and P_{241}. This issue became the major drawback since too many of these "poison" isotopes might render the bomb itself unpredictable (Hecht 1998; Roberts 1999). Taking furthermore into account that there were pressing time schedules and that no straightforward solution through chemical separation was in sight, what ensued was a patchy quick fix based on the removal of the fuel rods before too much P_{240} and P_{241} accrued in the irradiation slugs. Naturally, the flip side of a short irradiation period was wastage in the final amount of energy with the rods withdrawn before the natural uranium U_{235} isotope had been completely depleted. In practical terms, this design implied the permanence of fuel rods at the reactor core over a period of 250 days in G2 (France) and 350 days at Calder Hall (Britain) (Hecht 1998; Cockcroft 1957). These figures compare with the unloading and reloading of nuclear fuel in the forthcoming electricity generation nuclear reactors for periods only after about 1,100–1,500 days. The chemical purity of weapons-grade plutonium seems to justify the dissipation at least two-thirds of the potential heat released by fission.

How then could the cost of electricity be accurately forecast? How could the value of energy be ascertained when the costs were indivisible and energy merely a misused secondary by-product? Even assuming the complete redesign of energy optimization, difficult issues persisted since many of the equipment costs involved prototype costs tailored for the production of weapons-grade plutonium (for instance, equipment like the fuel loading system or the desuperheaters placed in front of the steam generator to absorb excess heat). One thing was the addition of a cooling circuit to remove leftover heat and transfer it to steam turbines and thereby generate electricity; another completely different the conception of an integrated, thermo-dynamically efficient nuclear power station capable of extracting the largest possible amount of energy from each U_{235} atom. Throughout this entire process, cost accounting was but a dupe of the primacy ascribed to military nuclear programs, and wavered between concealment and legitimization.

9.2.3 The Indivisible Costs of Dual-Purpose Reactors

Given the infeasibility of a strict cost accounting outlook for the price of nuclear energy, the issue ended up being willingly ironed out as a price arrangement between public institutions. On the one side stood the respective atomic energy agencies (*Commissariat à l'Énergie Atomique* in France and the United Kingdom Atomic Energy Authority), responsible for carrying forward the installation of atomic piles and the early reactors; on the other stood the late comers of the nationalized utility companies [Électrictité de France (EDF), the British Electricity Authority (BEA), and the Scottish Electricity Boards (SEB)] who claimed the right to cope with the market for nuclear power generation. Somewhere in between these competing claims (and clearly not from above), governments attempted to find some form of middle ground.

The French hammered out a guest-cost approach in which the leading role of the Commissariat for Atomic Energy (CEA) prevailed. Equipment costs were envisioned as if divisible, splitting the costs of the nuclear reactor (ascribed to the CEA) from the cost of the electricity production equipment (ascribed to EDF with the technological borderline to split the charges established by the heat exchangers). Then, the next step in the arrangement entailed the reimbursement of the CEA for EDF's share of the operational and maintenance costs plus interest, by means of successive payments bestowed by electricity supplied at the price of "the usage value for the distribution grid" (Lamiral 1988: 11). Under this contractual scheme, the EDF stake became a type of a guest at the major facility owned and run by the CEA and the electricity by-product somehow considered equivalent to a monthly rental payment.

The British, in turn, preferred a more sophisticated bilateral trading arrangement in which indivisible equipment costs were held and managed by the utility—the British Electrical Authority (BEA)—but the cost of inputs and outputs for the nuclear reactor were allocated separately to the BEA and to the United Kingdom Atomic Energy Authority (UKAEA). In practice, the first entity ran the power station and purchased the fuel rods prepared by the second in their chemical factories. Then, the client BEA turned into the position of supplier and, in this role, offered up its irradiated fuels for sale, on a fixed-price contract. Upon receiving this irradiated fuel "output," the UKAEA undertook the chemical separation of the plutonium contained in the uranium slugs. As for the price per unit of electricity generated by the first generation of nuclear stations, the authorities announced that this was "expected to be only slightly higher than the cost of electricity generated in the most advanced conventional plant… assuming present coal and oil prices." Yet, the costs of the later generations "should soon be fully competitive with conventional power plants" (UKAEA 1957).

Taken as a whole, resource allocation in dual-purpose reactors attempted to minimize production and transaction costs. This was achieved by assigning management and ownership rights to the organization that accounted for the most significant nuclear output: When weapons-grade plutonium dominated and only low levels of energy were fed into the grid, as in the case of France, the "market" for transactions was centered on the Atomic Energy organization; however, when electricity generation proved as important as plutonium production, as in the case of Britain, the core market for transactions was run by the utility. Not by accident, once the first single-purpose nuclear reactor to generate electricity was commissioned in France, the EDF company took over the role of the CEA in terms of the management and ownership of nuclear facilities. Despite these patterns of institutional agreements, tensions and differences in interests between electrical utilities and atomic organizations cropped up in both countries.

9.2.4 Prototypes and Series

While experimental atomic piles were unique prototypes designed for learning purposes, dual-purpose reactors already represented potentially replicable prototypes. Once the major technical difficulties were superceded, the particular design under scrutiny might become the foundation stone for a new technological path, narrowing the range of future options. Unable to enrich uranium overnight, pressed by the unfolding nuclear race and unwilling to depend on the United States or Canada for future supplies of enriched uranium or heavy water, the British and French converged around a distinct type of dual-purpose reactor based on the immediate availability of solutions: They chose to deploy rods of metallic natural uranium clad in a light alloy, in an array within a graphite pile and cooling it by pressured gas (first air, then CO_2). One key technical characteristic involved the selection of the moderator material for slowing down the neutrons (graphite) that provided the key identifier for this whole branch of graphite gas-cooled nuclear reactors.

The decision to move ahead with a "series" of nuclear stations of the same type did not foreclose the development of alternative designs, which remained under study in experimental facilities. In particular, French engineers felt the option of building reactors incorporating heavy water as the moderator still remained a feasible option. In any case, even with acquired experience and blueprint prototypes, devising the costs of a civilian nuclear program based on series of similar reactors was no easy task. Specific items in the levelized energy cost formula fell entirely beyond the scope of the existing calculations with engineers, thereby feeling compelled to fill in the blanks with some fairly wild shortcuts. For instance, to calculate the "operational and maintenance costs" of the future series of nuclear stations, they assumed these expenses would fall approximately within the range of those in effect for coal power stations, thereby extrapolating a percentage of the current conventional technology into the forthcoming technology. Overall, this accounting formula was applied somewhat forcefully in Britain but quite thoughtfully in France.

Different authors arrived at the figure of at 0.60–0.66d nuclear KWh for the British Isles, a price which pushed the impending reactors to the edge of competitiveness with coal power stations. Almost as if a target price, the 0.60d/nuclear KWh lingered in the minds of engineers ever since the debates held in the Harwell research branch in the 1950s. The same figure resurfaced again on the occasion of the fundamental Treasury Working Group document, authored by Burke Trend (1954) and which set the stage for the government's decision to replicate the successful Calder Hall reactor in a new eight power station investment program with each running two reactors and deploying an installed capacity of 1,500–2,000 MW to be built over a 10-year timeframe (1955–1965). Shielded by the authority imparted by the British Treasury, the Burke report foresaw a good investment opportunity because nuclear energy could, at the very least, certainly produce cost-effective electricity in the near future. Hence, the transition from dual-purpose prototypes toward power stations of the same design type contained all the advantages necessary for operational roll out. Whatever doubts there might be, the prevailing attitude was that the time was right for resolution (NA. POWE 14-740 1954–1955; Jukes 1955; Williams 1980).

Throughout all this, the cost accounting methodology resorted to a tricky procedure that somehow reflected the technological mindset of the 1950s: The figure of 0.60d was reckoned as a practicable energy price but only after compensating for the plutonium leftover in the reactor, technically referred to as the "plutonium credit." Without such credit, the real KWh price would approach 1.0d. The justifying reasoning was that in a nuclear power station optimized for electricity generation, in which U_{235} undergoes fission through to the release of all energy, the fissile plutonium (P_{239}) by-product still remains in the spent fuel. Albeit not of the same chemical quality as weapons-grade plutonium, this substance was still recoverable through chemical extraction. Among other aspects, the recovered plutonium might undergo fission more readily than uranium and, in the process, produce more excess neutrons for continuing the chain reaction. It was precisely this property that engineers tried to put to best effect in a second generation of nuclear reactors in which irradiation was based on fast neutrons and soundly termed "fast breeder" reactors. In the 1950s, experimental fast breeder prototypes entered into development not only in the US but furthermore in the then Soviet Union and the UK. Thus, the foundation for ascribing a market-like value to the plutonium credit was none other than the promise of the likely achievements of these second-generation technologies. Even though plutonium held a conditional, hypothetical value, experts were ready to discount this future value according to present circumstances at 0.35d per KWh. Unfortunately, mastering breeder reactors proved more difficult than foreseen and, right through to the current day, no commercial applications of this technology have achieved technical–economic viability (Cochran et al. 2010). For that reason, the price set for the plutonium credit plummeted in the 1960s and vanished in the 1970s. Rather than an asset, plutonium instead became a liability with the correspondent problem of siting a geological repository for the storage and disposal of spent fuel.

In sharp contrast to the detailed accounting found in Britain, French engineers branched the civilian nuclear power program without a clear figure for the future electricity price. In the same year that the British White Paper announced the commissioning of nuclear power stations (1955), the recently established Consultative Commission for the Production of Nuclear Electricity (the Commission PÉON in French) suggested the installation of EDF1 with 50 MW and an ensuing twofold increase in nuclear capacity every 3 years. Although there was no official commitment toward a target price, forecasts made by an atomic engineer suggested that the nuclear KWh might cost around 7.30–6.20 Fr in France, which was a clearly uncompetitive benchmark for the cost of coal station produced electricity (4.20 Fr/KWh) (AEDF. 800629 File 27327 1956).

Besides many unknowns in the data, the cost figures brought to light how nuclear was far from justifying comprehensive investment plans at least for the time being. This proved all the more so as these systematic programs began even before their dual-purpose reactor predecessors had been rated a success. In effect, what this means is that prior to a single KWh having been sent to the grid (Calder Hall and G1 only commenced operations in the fall of 1956), the next stage, the reactor-type replication stage, was already in motion. Such a risky strategy of forging ahead with technological innovation endowed cost accounting with a bad reputation because engineers were compelled to set values for equipment costs when the basic technological solutions were still under examination. In this respect, we should note that the critical step in developing nuclear power stations implied a whole new approach centered not just on the improvement of materials but also on maximizing thermo efficiency. The reactors had to perform flawlessly as well as economically and this drove a chain of learning by error adjustments. The sticking point was to retain sufficient reactivity to enable higher fuel burn-up: Table 9.1 shows that whereas the first atomic piles permitted the production of 0.1–0.15 MW from 1 tonne of natural uranium, nuclear power stations pushed this ceiling to 2.1–2.4 MW per tonne (Table 9.1, average rating MW/T). However, these aims could not be attained without first solving the problems caused by the increase in the can's temperature and the more likely deformation of the materials undergoing irradiation (Table 9.1, maximum can temperature). Devising new metal alloys, new arrangements for vertically stacking the uranium and new designs for the canned fuel elements were necessary steps to withstanding the temperature increase created by higher fuel burn-up rates. After that, recovering the heat produced at the core and transferring it with minimal losses to the turbines was furthermore vital. Once more, this step entailed increasing the plate thickness of the reactor vessel and improvements in welded seams so as to augment

Table 9.1 Technical characteristics of early nuclear reactors in Great Britain and France (Yvon 1955; Hinton 1957; ANF Box 19910737 File 3 1959)

Reactor type	Experimental atomic piles	Plutonium-producing piles	Dual-purpose reactors	Nuclear power stations
Reactor	(Bepo)	(Windscale)	(Calder Hall)	(Berkeley)
Operational launch	1948	1950	1956	1961
Average rating MW/T	0.15	0.75	1.4	2.4
Maximum can temperature (°C)	200	300	408	450
Maximum gas outlet temperature (°C)	90	211	336	345

Reactor type	Experimental atomic piles	Dual-purpose reactors	Dual-purpose reactors	Nuclear power stations
Reactor	(EL-2[a])	(G1)	(G2)	(EDF1)
Operational launch	1949	1956	1959	1962
Average rating MW/T	0.1	0.37	1.8	2.1
Maximum cladding/can temperature (°C)	126	–	525	550
Maximum gas outlet temperature (°C)	70	104	340	349

[a] Zoé (1948) was the first French experimental atomic pile. Data from EL-2 based on the pile operating at 800 kw thermal power

gas pressure, gas temperature as well as advancing with the uniformity of channel gas heat outputs (Table 9.1 maximum gas outlet temperature). As may be deduced from Table 9.1, the major leap forward on all these thermo-dynamic fronts came with the changeover from plutonium-producing piles to dual-purpose reactors and power stations. Inasmuch as this leap occurred quite suddenly and with the superimposition of technological stages, it heightened the hazards already inherent in forecasts and provoked delays in the overall plans (Ailleret 1954; Hinton et al. 1960). Moreover, the track record of British gas-cooled power stations in the 1960s disclosed several weak points in their technological operational performance.

9.3 The Dawn of Super-Technologies

9.3.1 The Upscaling Cycle

The post-Second World War conjuncture created a sense of stability, prosperity, and progress. The combination of low interest rates, controlled inflation, moderate wage growth, cheap raw materials, and strong investment prompted the idea that most variations in levelized energy costs were caused by changes in technology. In fact, experience showed that while operational and maintenance costs, as well as fuel costs, did not vary greatly from year to year and were straightforwardly predictable, capital costs proved much more fickle and uncertain. Thus, from the 1950s and up to the first oil shock, capital equipment captured the spotlight of cost forecasts.

Judging from the views of decision-makers, the significance of capital costs stemmed from the possibility of pushing the technological frontier forwards, increasing the capacity of power stations and achieving additional cost reductions. At the crossroads of the 1950s, every condition seemed to favor the motto: Bigger is cheaper; bigger is more efficient. Technically, the increase in size or in the performance capacity of a given technology and the expected reduction in average cost is summarized by the concept of "scaling-up" or "upscaling". Among the developed nations, France and Britain benefited from prospects quite advantageous to pursuing the systematic upscaling of electricity generation as both countries had experienced dramatic bottlenecks in energy supply up to 1955 and were furthermore eager to prove the scope of opportunity for nuclear power competitiveness. Furthermore, the nationalization of the electrical, coal, and gas sectors following World War II prompted a technocratic-political vision of state owned enterprises in which the public sector acted on behalf of the long-term interests of consumers, defending the general interest at large (a view personified by the political stand of ministers like Herbert Morrison in Britain and Paul Ramadier in France). According to this view, the major energy supply obligation was to guarantee universality of access, equality of prices, and the reallocation of efficiency gains. All technical "conquests" of public ownership such as the integration of the national grid, the coordination of public investment, R&D and productive activities, long-term planning, and the availability of intensive capital emboldened the attempts to capture economies of scale units. Techno-political elites translated the idea that the technological moment was ripe for nationalization, meaning that public property could step up the obvious trend toward increasing sizes and levels of performance so as to redistribute its benefits throughout society. This common ground of technological legitimacy largely explains why both the left and the right supported the nationalization acts in France and in Britain alike.

Historically, scaling-up has been identified with a phase in the lifecycle of technologies in which a "radical innovation becomes embedded as the dominant design" with core concepts settled for components such as boilers, alternators, or turbines and with their resultant integration in an assembled product. According to Frenken and Leydesdorff (2000), once a technology reaches the "dominant design" stage, upscaling is likely to occur. In the same vein, Wilson (2012) confirmed an "upscaling" cycle in both coal and nuclear power stations spanning the period after World War II and characterized by the most rapid increase in size/capacity ever seen in terms of maximum unit capacity. What appears to be at stake is a critical juncture somewhere in the course of development, whereupon it becomes possible to set out a research agenda driven by the ambition to produce "super-technologies" and fully capture the respective economies of scale. Once in motion, this agenda follows a trajectory of further increases in capacity until a unit scale frontier is reached beyond which diseconomies of scale impose additional costs. At this moment, the technology encounters its upscaling limit, and a new cycle, with a new agenda, begins. More to the point, the research agenda involved in the upscaling of electricity generation drew upon the emergence of a renewed "Materials Science", since the crux of the matter in the 1950s and 1960s turned out to revolve around the development of new alloys and new welding techniques to cope with higher pressures and higher temperatures. In this respect, the metallurgy state of the art proved an impediment to the achievement of quick progress. Until maximum temperatures of 565º–600 ºC had been attained in the steam circuit of coal power stations, engineering efforts were concentrated on testing and experimenting ferrite steels with chrome, molybdenum, and vanadium. Above that ceiling, matters became far thornier due to the need to test and apply a different class of high tensile strength steels, with their crystal structure altered from ferrite to austenite, containing chrome and nickel. Austenite steels proved much more difficult to manufacture, to weld, and to pull together with other steels. From the point of view of historical actors, the upper limit of 565 ºC, roughly corresponding to the 600 MW power capacities, was attained by coal power stations at the dawn of the 1970s and thus constituting the first serious obstacle to unremitting upscaling.

Moved by the quick adoption of central investment plans that channeled money into the nationalized industry, France was one of the first nations to press ahead with deliberate upscaling. The breakthrough was accomplished when EDF's equipment management team set forth the idea of vertical standardization, devising the installation of integrated equipment sets with a single boiler for a

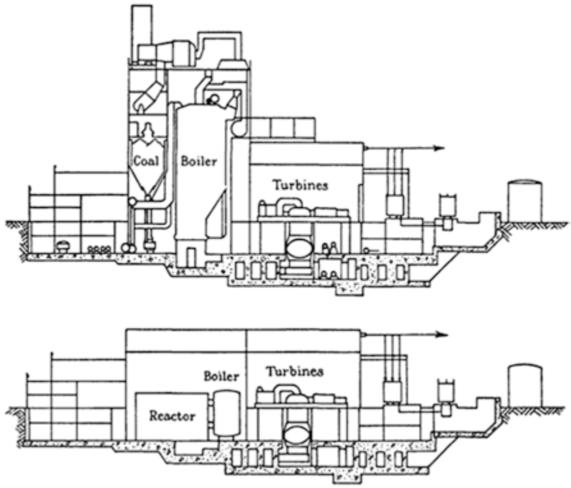


Fig. 9.1 Relative sizes of nuclear and coal-fired plants (reproduced from Cockcroft 1953)

single turbo-alternator, and transformer. This cluster was coined a "tranche" and designed for the *Nantes-Chevriré* coal power station due for completion in 1953. Then, the next step followed suit, adding the concept of upscaling to the principle of replication and standardization. Correspondingly, a whole new blueprint of expansion through "*palier techniques*" or technical plateaus, materialized for the first time in the *Creil* and *Porcheville* stations, both located in northern France (1956). Instead of a single tranche, each power station was henceforth equipped with one to four identical groups, each with a turbo-alternator capacity of 115/125 MW, and so each station machinery room contained only equipment from the same manufacturer. To enhance unit cost reductions resulting from the cumulative experience of production, the 115/125 MW pattern was then replicated, paving the way for the installation of 36 similar tranches in other districts that added a total of 4,500 MW of cutting-edge electricity generation. Most significantly, the *Porcheville* power station proved a real milestone in standardization as it set forth an array of innovations reflecting this new "ergonomic" concept of power stations. To detail only the most important advances: adoption of team circuit reheating to improve the thermo-dynamic cycle; autonomy of each tranche ensured by separate control instruments; normalization of the electric voltage and of the auxiliary devices; horizontal architectural design of the buildings based on semi-outdoor construction aimed at introducing cheaper and more functional designs; and mechanization of coal unloading and loading operations. Notwithstanding the clear benefits soon reflected in costs, the idea of producing more of the same raised a few eyebrows within the state-owned enterprise. Some sectors resisted the idea of building a new power plant without at least somewhat improving on prior versions. After all, perfection was a quality highly prized by engineers. Indeed, only after time did the tranche concept become fully accepted as a standard worthy of replication (AEDF. 943408 File 39726 1958; Picard et al. 1985) (Fig. 9.1).

Once in motion, the upscaling trend soon became overwhelming: the 115/125 MW coal power plant blueprint gave way to a new plateau of 250 MW in the 1960s, before the advent of 500–600 MW capacities in the 1970s–1980s with the corresponding maximum steam temperatures of 565º. Meanwhile, nuclear stations progressed still further moving on from 200 MW in the 1950s to 480 MW in the 1960s, 900 MW in the early 1970s before attaining 1,300 MW by the end of that decade.

By a range of different means, the upscaling trend spanned much of the world and, along with abundant fossil fuel supplies, contributed toward decades of unprecedented low energy prices. Among other aspects, the era of super-technologies ensured the production of more energy from less fuel, merging gains from economies of scale with gains from standardization and economies of learning. Yet, these drivers of productivity were not always interlocked. British industry, for example, secured fewer benefits from the upscaling trend owing to a more troublesome process of learning and standardization. Two points stand out in contrast with the French development process: First, upscaling unfolded according to a policy of “small steps,” meaning that the leap forward from 200 to 500 MW capacity plants was bridged by the installation of a small number of transitional coal power units of 275 and 350 MW, thereby multiplying the quantity of tranches to be standardized and thereby mislaying the economies of learning gained from replication. Secondly, the incentives to competition in the manufacture of turbo-alternators and boilers fostered diversity of coal and nuclear power station designs that correspondingly complicated system integration, system maintenance and repair, as much in terms of technological plateaus as in terms of tranches. The worrying fact was that while at the outset, there were decreasing costs with the replication of generation sets, this trend soon lost ground and when the time did come to upscale toward 500 MW tranches, replication revealed increasing prices (NA. POWE-14-2744 1972). The economies of scale squandered the economies of learning. For these reasons, the British electricity industry fared worse at the close of the 1960s in experiencing the first signs of diseconomies sooner than other nations (see below Sect. 9.3.3).

9.3.2 The 0.3 Power Law

Out of the upscaling experience, a key question crossed the minds of those involved in planning: How much could be spared by skipping straight to the next plateau? At least up to the 1970s, it seemed as if the way ahead could only bring very worthwhile increases in both thermo efficiency and in cost. Hence, moving on fast to larger and larger sizes held a self-justifying appeal and ministers were often requested to examine reports in which technical progress fed a spiral of decreasing marginal costs. With a considerable amount of data available, engineers in the meantime devised an equation to gauge the cost–capacity gains. Its advantage lay in the ability to extrapolate future capital costs from those presently in

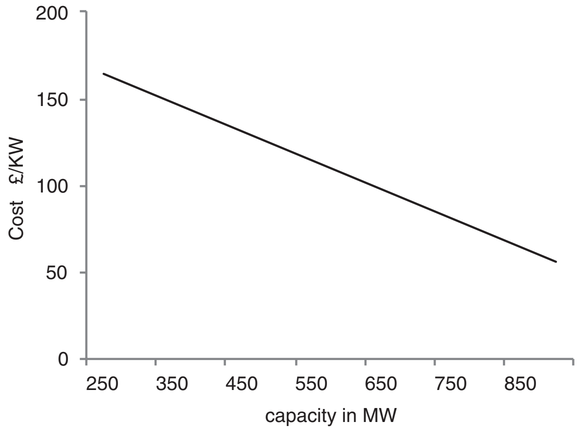


Fig. 9.2 The cost–capacity relationship estimated according to the 0.3 power law. British nuclear power plants 1963 (NA POWE-14-1407 1963–1965)

effect. Known as the "power law," this formula relates the effect of increasing the size of a power unit or a tranche (Q_2/Q_1) with the previously known cost (Cost_1) by means of the exponent (x).

$$\text{Cost}_2 = \text{Cost}_1 \left(Q_2 / Q_1\right)^x \tag{9.3}$$

Provided that everything else remains constant, the power law is a powerful tool as it enables practical research into the correlation between the cost at capacity Q_1 and the cost at capacity Q_2. Once a determined exponent is accurately set for various plants, for technical systems or even for single components, it becomes a yardstick to gauge future costs and often included in engineering handbook reference tables. Moreover, the formula allows for the singling out of clear concepts in upscaling forecasts distinguishing between cases where $x < 1$, indicative that economies of scale are present, and smaller values of x, demonstrating greater cost advantages in moving to an upper technological capacity; the case of $x = 1$, indicative that a linear relationship is present; and cases of where $x > 1$, which signals that diseconomies of scale have begun to build up and, therefore, the larger the size increase, the more costly the upscaling becomes.

One of the most startling and rousing discoveries of the late 1960s was that the power law projected quite sharply the costs of power plant growth. Particularly in the domain of nuclear power, engineers found that $x = 0.3$ provided a close fit to the capital cost data. Somehow the figure reflected something that was already well known in the professional milieu: The existence of very strong electricity generation unit scale economies. Figure 9.2 represents the equation $\text{Cost}_2 = \text{Cost}_1(Q_2/Q_1)^x$ in a linear adjustment based on revised estimated costs, that is, cost revised by the need to take account of design changes as of 1963. As expected, the 0.3 power law appropriately fits the empirical accounting data and discloses the potential for new investments in plants with enhanced capacity. More to the point, the graph illustrates a moment in technological history and a particular time frame when progress was not only a prized target but also apparently very much in sight. Departing from the 275 MW plateau, under the aforementioned rule, one might expect a 38 % reduction in costs when the reactor capacity is doubled. However, from the viewpoint of studies undertaken later in the US, average cost reduction values of between 15 and 30 % following the doubling of capacity appeared to be closer to the mark (Komanoff 1981; McCabe 1996).

Needless to say, much of the accuracy resulting boiled down to that clause: “everything else remains equal.” In a perfect and stable world, we might perhaps posit that the costs vary according to linear increases in capacity. However, what if the ladder-like dynamic of doubling capacity does raise unforeseen technological issues? Furthermore, what if society changes and requires another plateau of safety or environmental norms be added to accompany the oncoming burst of super-technologies? What if the economy or the energy policy enters a new cycle? In all of these situations, the pattern of singular correlation between capacity and cost withers and the picture becomes much more complex. Such shifts, actually experienced in the 1970s, overturned many of the forecasting techniques in effect. Among other consequences, the changes paved the way for a multifactor approach to cost accounting (Komanoff 1981).

9.3.3 The 1970s Alterations

Before the traumatic Organization of the Petroleum Exporting Countries (OPEC) October 1973 declaration triggered an increase in oil prices large enough to cause a significant reduction in real global gross domestic product, cost pressures had already been gathering pace within electricity systems. Thanks to enhanced capital costs, producing more energy was not becoming any cheaper. Step by step, the trend of upscaling, standardization, economies of learning, and economies of networking was gradually crawling to a halt. The different streams contributing to the rise in capital costs unfolded quite independently and asynchronously but for whatever reason did seem to collectively snowball in the 1970s, first in Great Britain and then, subsequently, in France. Cumulatively, they offset the progress toward decreasing costs brought about by the age of super-technologies. Within this framework, the first oil shock took place in a critical juncture between the deceleration of gains in fixed capital costs per KW and the momentous rise in fuel (variable) costs.

Five major factors may be pinpointed as wielding an influence over the capital cost trend. In chronological order, they are as follows:

9.3.3.1 Increase in Interest Rates

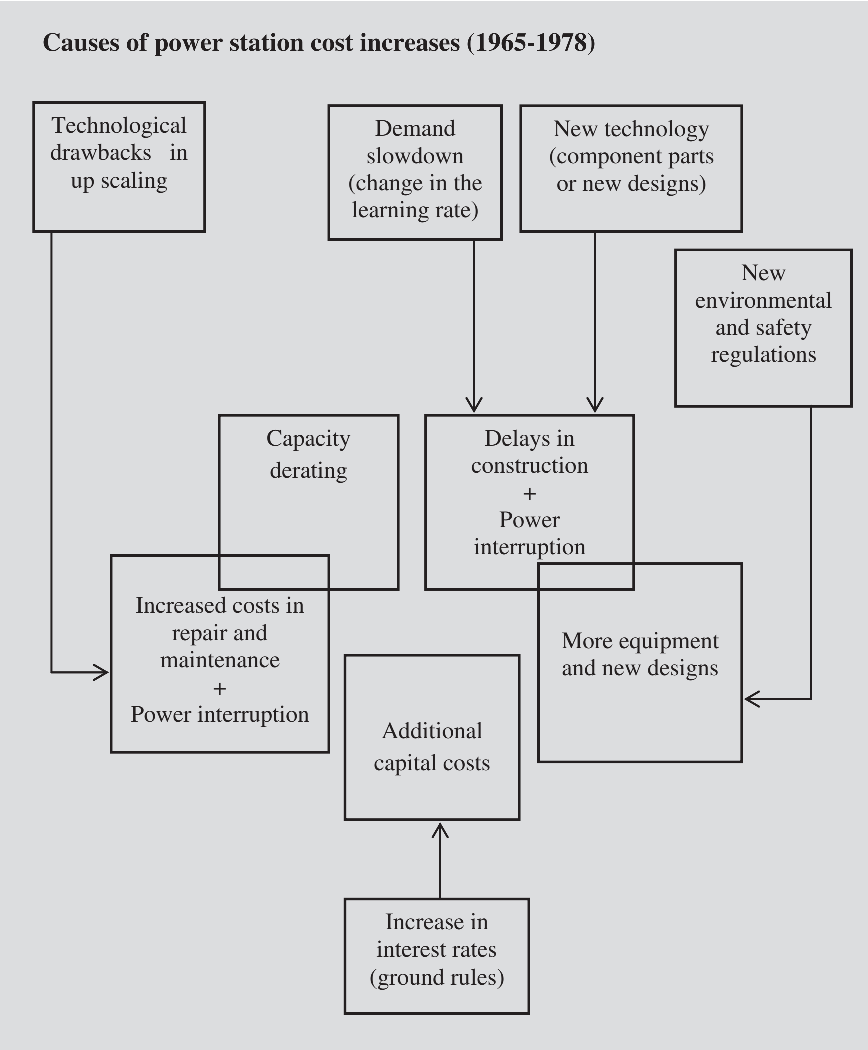
The fact that nationalized industries had been obtaining a low return on their capital in relation to that gained by commercial or industrial capital had generally been widely debated since the 1960s, whether by the British Treasury and Commissions in charge of issuing “White Papers,” or by the Plan’s Commission and the Government in France. Even though there were a number of good arguments for maintaining low interest rates, there was also the concern that such a policy might jeopardize the competitiveness of public enterprises and set a low benchmark for entire sectors of the economy. In order to overcome such apprehension, it was decided that electricity generation and distribution should improve its net return from a rate of 4.5 % per annum to 5.5 %, then to 7–7.5 % and finally up to 8 %. In 1974, the ground rule in Britain was 8 %, but engineers stuck to the custom of always incorporating a double scenario into accounting forecasts, not only with the 8 % rate but also with a 10 % rate. Equally, in France, the trend over the same period after World War II evolved from a 7 % to a 10 % net return. Overall, the burden of higher capital costs weighed mostly against the competitiveness of nuclear power as this technology required a greater percentage of capital costs within the overall total cost.

9.3.3.2 Technological Drawbacks in Up-scaling

Technological problems that only reveal themselves after weeks or months of operation prove a nightmare to any engineering team. Ex-post faults not only cast serious doubts about those in charge but furthermore result in time-consuming endeavors in tardy diagnostic efforts. From a cost accounting perspective, unforeseen technical problems exert a twofold pressure over costs: First, they entail the interruption of work and repairs to the dysfunctional electricity generation equipment. Secondly, whenever the flaws cannot be completely circumvented, the only feasible solution is to operate the machine at less than its rated maximum power, thereby avoiding both critical temperatures and critical atmospheric pressures. Technically, this option of last resort is called "derating." Both developments, interruption and derating, occurred after the jump forward to 500 MW generating sets in British power stations. In the testing phase, a problem with the erosion of low-pressure rotor blades was detected, with droplets forming on them at high speed, a problem highly difficult to simulate in laboratorial settings. Likewise, regular inspections of nuclear power stations based on the Calder Hall prototype (with its graphite gas-cooled reactors) showed an unexpected degree of corrosion in some steel components in the carbon dioxide coolant section and in all power stations in operation. The dangers posed by this situation raised fears that a temporary or even a permanent shutdown of some or all of the affected reactors might become necessary. After a period of investigation, it became clear that corrosion problems could be circumvented by firstly replacing mild steel by stainless steel in certain core fittings and then secondly by reducing operating temperatures. This occasioned the permanent derating throughout nuclear reactor life of 10 % in some stations and by 20 % in others.

Comparatively, the French industrial sector encountered fewer shortcomings. The closest to the British experience was the protracted interruption to the Chinon EDF III nuclear reactor (also graphite gas-cooled) in 1966, followed by its temporary derating by 50 % in 1968 due to faults in fuel rods control devices. Prolonged stoppages at the Chooz Nuclear Power Station became a second trouble spot (AEDF. 800545 File 27243 1970).

Clearly, these interruptions and deratings raised total energy costs by diminishing the amount of energy produced—the denominator in the levelized costs equation (see Eq. 9.1) while increasing the costs of repair, maintenance, and research. As always, the longer the time frame with lower levels of output and the greater the inputs, the more expensive the average life cycle cost becomes.



9.3.3.3 Demand Slowdown

Most of the expectations of high growth in electricity demand were not borne out. By the second half of the 1960s, decision-makers were already aware that demand growth had been clearly overestimated in Britain and was unfolding behind expectations in France. In the 1970s, demand plummeted to unthinkable levels. With excess capacity in sight, it became correspondingly necessary to review medium term plans and slow down the pace of investment by widening the gap between the commissioning of new plants. The effects of this downturn were felt by private consortia, electrical equipment producers, and the civil engineering contractors that worked for the public electricity utilities. Economies of learning based on the know-how picked up through experience regressed in tandem with the pace of demand. In a meeting held in June 1972 between the French and British authorities to examine this issue, the latter asserted that “the low rate of ordering for new plant had in itself increased costs” (NA. POWE-14-2744 1972). Aside from the likely multiplication of small flaws, fewer opportunities for learning held major implications for extended construction times with the subsequent drag on capital costs and interests. To counteract the wastage of know-how, the governments embarked on a thorough policy shift, favoring merging and the concentration into one or two supplier enterprises instead of competitive tendering among numerous market rivals. After some steps toward industrial integration, complete supply chain concentration was achieved between 1973 (Britain) and 1975 (France).

9.3.3.4 New Technology: Introduction of Component Parts or New Designs

The introduction of innovation into either systems or their component parts generates contradictory capital cost effects: while in the long run, they contribute to enhanced security and efficiency, lowering average costs, in the short term, they often drive the return to a phase of trial and error more likely to raise costs than otherwise. This is particularly evident whenever these innovations are embroiled in longer construction times and/or with power interruptions. French economists singled out this reality under the concept of “*frais de prototype*”—prototype charges—to illustrate the gap between the first tranche of a série or component and the whole tranche or component with identical characteristics (AEDF. 800624 File 27322 1971). As a result, upscaling becomes not just the continuos transformation of costs through economies of scale but rather a process punctuated by abnormal leaps and starts brought about by innovations. The concept of “prototype charges” therefore proves relevant to grasping the capital costs of in effect for some particular power stations. To mention just the most prominent cases of drifting investments and construction delays: the Wylfa British nuclear power station, commissioned in 1971, with a new integral concrete vessel concept that enclosed an once-through boiler; Dungeness B, the first advanced gas-cooled reactor system (AGR) nuclear power station, ordered in 1965 but which only entered operational service in 1983; *Choz*, the first commercial French Pressurized Water Reactor, operational in 1967 and managed by a joint French–Belgian enterprise.

9.3.3.5 New Environmental and Safety Regulations

The upscaling trend came hand in hand with the slowdown in demand that, in turn, drove profound alterations in the industrial landscape: while the number of power stations decreased, their size and their impact upon their respective local environments became, in contrast, overwhelming. Part of the more strigent environmental, conservation, safety, and security regulations derived from the novel scale of the problems the authorities began having to face up to; while another part, and no less important, stemmed from middle class concerns about nature, landscape, the amenities, and resource preservation, as well as community health and the broader wellbeing. As the 1960s ended, all such apprehensions coalesced around the label "environment," which moreover acquired political significance. The changeover was particularly acute in the collective feelings toward atomic energy, no longer regarded as the hallmark of science, technology, and progress.

Almost every aspect of electricity generation was subject to narrow regulations, reflecting the awareness that energy usage was not without harmful impacts: the emission of grit, dust, CO_2, and sulfur dioxide from the use of fuel; the emission of smoke from the incomplete combustion of bituminous coal; the discharge of hot effluents from direct cooling power stations; the control over solid radioactive disposal and the amount of radioactivity released into the atmosphere; double certification and quality control in the manufacture of boilers, turbines and tubing; and visual amenity and noise limits.

The most obvious effects of tighter regulations and technical prescriptions were felt in the increase in equipment costs. In the 1960s and 1970s, these embraced the rescaling of chimneys to taller heights; the installation of adequate grit and dust electrostatic collectors; improvements in scrubbers; conversion from river cooling to tower cooling; the extension of major plant structures in nuclear stations with the installations of barriers and increased separation distances; the expansion of instrumentation, control, and monitoring devices; and the increase in redundant safety-related equipment.

Besides these direct effects, environmental standards also had less evident consequences upon variable costs, in particular, the default pricing on coal and oil with high sulfur contents and the preference for premium priced sulfur-free fuels. Until the theme of acid rain took over the European agenda, the usage of low sulfur fuels or the removal of sulfur before combustion represented the cheapest solutions for combating air pollution. More sophisticated responses based on regenerative systems only came to the fore in the 1980s. Finally, there were also overhead time costs with the increase in bureaucratic controls at every single stage of production, from the certification and control of electrical equipment through to the preliminary studies for the siting and licensing of power stations. Enhanced security drove up costs through extended construction times, more personnel in the workforce, and a broader range of tasks requiring completion (Sheail 1991; Clapp 1994; Foasso 2003; AEDF. 800621 File 27319 1971).

Taken as a whole, this was a period of goal-oriented endeavors, technocratic in outlook, thrilling in discoveries, and rewarding in results. Major scientific breakthroughs in electricity generation and nuclear power had already been accomplished. Drawing on the pool of accumulated scientific knowledge, engineers, politicians, and latterly economists felt that progress was at hand. This awareness framed and sharpened the ambition to produce "super-technologies," to fully capture scale economies and to redistribute the benefits of cheap energy. By the 1970s, however, the technological and thermodynamic evolution was approaching a troublesome threshold. Further gains from heading down the same path henceforth became hard to achieve. Moreover, the era of technical sovereignty began to fade. Technology was becoming far too important to be left to engineers. Public airing and activist protests not only impinged additional costs on decision-making and on preliminary siting phases but also disclosed how engineering and economic criteria might be completely at odds with community interests. Two major events, which gained exceptional and hitherto unprecedented press coverage, came to represent a milestone in the democratization of energy decisions: The first was the appointment of a totally independent commission constituted by academics and members of private associations to issue a report on the security of the Fessenheim nuclear reactor in Alsace-France during its operational shutdown in 1989. The final outlook of the "counter-experts" disavowed the restarting of operations, opening up the battleground between the long-standing Alsatian activist associations and the electrical utility EDF. The second milestone was the approval of a public hearing for the installation of a pressurized nuclear water reactor at Sizewell B in Suffolk, England in 1982. By most accounts, the hearings costs amounted to about £10 million, a figure that portrays the sheer scope of the enterprise. Structured as an open-ended, legally based process for technological decision-making, from the outset the inquiry had very broad terms of reference, opening up the field to the participation of a large number of objector groups with their respective different insights. All this added up to a non-stop venture, breaking new ground for airing alternative views and profile-raising public dissent. Despite wide repercussions in the academy and in the press, the debate nevertheless got little attention from the public at large (Davies 1984; Foasso 2003).

What had begun as an exciting transposition from laboratory practices onto manufacturing innovations ended up being debated in the press and, while not by commoners at large, at least by experts and by political activists. Super-technologies would henceforth be imbued by social, political, and environmental assessments.

9.4 The Costs of the First Oil Shock

9.4.1 Oil Power

For a long time, the choice among alternative technologies was effectively limited to coal-nuclear options. Taking into consideration, the course of events during the heroic years of European reconstruction, marked by chronic energy shortages, one might expect the blossoming of power stations equipped with fuel–oil boilers. However, in effect, the swift adoption of oil as a source of electric power proved to be a false start. In Great Britain and especially in France, fuel–oil receded into the background after the nationalization of the Suez canal by Egyptian president Nasser on July 26, 1956, and the ensuing invasion of the canal zone by British, French, and Israeli troops. During the conflict and its aftermath, the canal was closed to commercial shipping and the Levant pipelines, carrying oil from Saudi Arabia to the Mediterranean, were also shutdown, causing wide-ranging disruptions to three quarters of European supplies. The last quarter of 1956 and the first months of 1957 witnessed a governmental nightmare, with rationing of motor fuels and rationing electricity supply to both domestic households and to companies and manufacturers. The return to normality left a sense of trauma as regards the way the security of energy supplies had hitherto been planned and designed, with "caution" as the watchword to be stressed into the future. Dependence on Middle East oil required solutions and European governments steadfastly decided to resume the installation of oil-burning equipment and step up investment in nuclear power stations.

Revealingly, the orientation of governments did not change greatly when oil prices dropped to all-time lows in the 1960s. The Suez memory still remained fresh and ministers resisted pressures from the electrical utilities to generate power resorting to highly competitive fuel–oil imports. Moreover, the defense of a moderate and controlled rundown of the industrial workforce at coal mines was also at stake. In Great Britain, two oil-fired stations were authorized (Fawley and Pembroke) but the first Labour government of Harold Wilson refused consent for the installation of a third oil-fired station at Littlebrook D and issued a virtual ban on conversions from coal to oil in the Fuel Policy White Paper released in 1967 (Cmnd 3438). The Central Electricity Generation Board—CEGB, however, continued its unyielding pressure and did not allow the issue to lapse as oil-firing power stations displayed the lowest levelized costs of all classical thermal stations, presented the highest thermal efficiency and provided twice the average calorific value of coal per unit of weight. Ultimately, labor had to give up some conversions (CEGB 1973–1974). In France, some oil-firing stations (mostly dual firing, which meant that they were switchable between oil and coal over a 15-day schedule) were authorized, albeit with only a limited weighting in the overall energy system.

After a decade of caution and compromise, the switch to fuel–oil quickened significantly in 1970. In just four months, about 3,250 MW of coal-firing stations, involving a total of 30 boilers, were converted in Great Britain to either oil or natural gas-firing while France experienced the same process one year later with the adaptation of 1,750 MW, plus another 1,000 MW, to fuel–oil burning. With this alteration, the technological plateau of 250 MW power stations was switched from coal to oil in France. The trend proceeded apace in the ensuing years, backed up by long-term contracts signed with the oil business giants: BP, Shell-Mex, Esso and Texaco (then called the "Regent"), on the one hand, and Elf-Aquitaine (*Union Générale des Pétroles*) on the other. On the eve of 1974, 39 % of the electricity produced by *Electricité de France–EDF* and 22 % of the electricity produced by the Central Electricity Generating Board–CEGB was generated by oil-firing stations.

Inasmuch as levelized costs demonstrated that oil-firing was cheaper outside of base load periods, it became increasingly difficult to justify political bans couched in arguments of coal protection, strategic security of supply, or balance of payments stability. In this respect, the 1970 turning point indicates that the equilibrium of powers had tilted definitely in favor of the electricity utilities. Many reasons seemed to have accrued to this turnaround: Growing concerns over air pollution and the fact that fuel–oil power stations released less grit, dust, and CO_2 than coal stations; the expansion of national refinery capacities and the ensuing local availability of larger quantities of fuel–oil, along with the boost to demand for gasoline; full-on optimism linked to the hot technological revolution thinking surrounding British North Sea gas and oil; the bridging role played in France by Pierre Guillaumat, simultaneously the head of EDF and of the major French oil company. On the other hand, the conjuncture proved not just favorable to oil-firing but also adverse to its competing technologies: There was widespread uncertainty over the technological paths for the future development of nuclear power with a deadlock between graphite gas and pressurized water reactors; the expectations of forthcoming coal shortages prevailing over the course of the winters of 1970 and 1971; a decrease in productivity of coal mining; mounting doubts around the rationality of energy investments voiced by the treasuries of both countries. Nevertheless, within this mix of structure and action, the pivotal driver of change toward electricity fuelization seems to have derived more from the side of action. However, important the aforementioned factors might have been, what actually triggered the swift fuelization was the restatement that state owned enterprises should be judged by their commercial success. In practice, and not just in theory, governments accepted the independence of public boards. This policy of public sector competitiveness involved an overture to foreign competition, technological modernization and support for a policy of "national champions" was put into practice in the final phase of the British Labour government of Harold Wilson (1964–1970) and pressed ahead with, but not without problems, by the Conservative administration of Edward Heath (1970–1974). In France, such a strategy was pursued during George Pompidou's presidency in a retreat from the previous Gaullist orientation (1969–1974) (Robens 1972; Hall 1986; Brussière 2003). As a consequence of the liberal-commercial approach, the electricity utilities expanded their freedom of choice between technologies, between fuels, and between investments. When free to choose oil-firing, they grabbed at the opportunity.

9.4.2 The Aftershock

Most of the concerns over the security of oil supplies seemed to vanish at the beginning of the 1970s. As always, individuals foresaw the future as a variation of known patterns—in an echo that reverberated with ever less amplitude. These mindsets led Western decision-makers to spotlight core events in oil-producing countries. To European eyes, the crux of matter lay in the changeover in Libya where the newly arrived leader Muammar al Qaddafi compelled the western concessionaire companies to concede higher taxes and higher prices to the government, under the mantle of anti-Zionism and anti-imperialism. After that, the common agreement established at the Teheran and Tripoli conferences by the OPEC member states spread the benefits of the Libyan example far further, forcing the oil majors to concede much higher host-government "takes." These developments drove oil price increases: The posted price in the Middle East rose from $1.80 to $3.00 per barrel between 1970 and 1973.

Looking at the reports and the scenarios predicted at the time by French and British analysts, one nevertheless gets the sensation that they tended to understate the risks associated with both the political as well as with the market conjuncture. Within the respective ministries of industry and power as well as the electrical utilities, the experts consistently agreed that there were a "number of factors likely to restrain OPEC members from pushing claims for increased prices to extremes," or that the "tax yield of oil-producing nations will tend to diminish (at least marginally for the extraction of crude exceeding the normal extraction plans)…." This very same view was aired by energy expert "gurus" like Maurice Adelman or Paul Frankel (NA. EG-20-3 1974; ANF. Box 19910737 Dossier 1 1970). On the geostrategic level, one finds the very same underestimation, if not outright disregard, about what was going on in the Middle East in the high-ranking decision-making spheres of the US administration where President Richard Nixon apparently seemed more alarmed over the emerging Watergate case scandal (Howland and Daigle 2011).

Within this context, all that might reasonably be expected were minor and incremental increases in the price of Middle Eastern crude, since the market and bargaining position of Arab countries had been exhausted. It was then believed that the upper hand now belonged to the West and the petroleum-based electricity generators would bask in a glow of growth and development. More troublesome were perhaps the long-term prospects. Should consumption continue to mount and the level of reserves not be supplemented by new discoveries, the threat of impending depletion had to be taken seriously. Although this sort of expert wariness was not imbued with the critique of technological optimism that pervaded some ecological thinking (for instance the "Limits to growth" report released in 1973), it nonetheless anticipated the enhanced scarcity would reflect in higher prices: In the approximate forthcoming 15 years, a shortage premium would thus push prices upward. The levelized cost forecasts of oil-firing power stations therefore envisaged stability in the cost of crude, perhaps with certain marginal increases over the period 1973–1985, followed by a rising trend there afterward. Such a future, however, represented a variation of the already identified patterns.

The oil embargo thus came as a complete surprise. On October 17, 1973, 11 days after the outbreak of the Egyptian–Israeli war, Arab oil ministers meeting in Kuwait agreed to institute a total oil embargo against the United States and other countries friendly to Israel. Additionally, they decided to cut back production by 5 % from the September level and to continue cutting by 5 % in each succeeding month until their objectives were met. The ensuing escalation drove oil prices to the unprecedented level of $11.65 per barrel.

The "oil shock" concept encapsulates the psychological sense of perplexity alongside the rippling effects felt throughout the entire extent of the economy. Furthermore, the disturbances in the oil market were no short-lived phenomenon but were bound up with the turnover generated throughout the business cycles and impacting on inflation, wage-price spirals, rising interest rates, falling investment, increasing unemployment, decreasing productivity, and long-term decreases in economic growth. The "first oil shock" therefore refers to the sequence of events that spans the Kuwait embargo declaration in October 1973 right through to October 1978, when a second shock rippled through the system (Iranian oil workers walking out on strike at the outbreak of the Iran–Iraq war).

In response to higher energy prices, firms and households tend to consume less energy while enterprises making recourse to energy-intensive capital stocks may experience a faster depreciation of their stocks. In addition, the cost structure changed in explosive and erratic ways, owing to the drag effects of oil upon coal prices combined with an upswing in the price trend for natural uranium, which first began rising in 1972. Equally, the consequences of inflation upon equipment, facilities, and workforce remuneration were hard to grasp, let alone to forecast.

For over 25 years, experts had mostly focused on capital cost-related issues. More to the point, they tried to ascertain the sensitivity of costs to modifications in upscaling. Tactically, it was assumed that other factors in the levelized cost formula (see Eq. 9.1) remained to a greater or lesser extent stable. Now, in the midst of the 1973 upheaval everything in the formula seemed to change overnight: capital costs and fuel costs as well as operational and maintenance costs. Without a past anchor to form the foundation, it proved really hard to convey an accurate picture of the best future choices. Yet, at no other stage was such information more urgently needed. Public utility policymakers and managers turned their attention to altering the energy mix and shifting the balance away from petroleum. This was furthermore crucial because the most recent equipment commissioned by France and Great Britain was for oil-firing power stations and flexible dual firing to allow for a quick switchover between oil and coal. However, it remained to be ascertained not only just which technology was more competitive but also the most appropriate energy mix.

Figure 9.3 represents the levelized costs of electricity generation for coal, oil, and nuclear power stations in Great Britain and France. The gross domestic product deflator was applied to account for nominal costs and express all values in constant 1978 prices (Economic Statistics Database 2010). The data were compiled from historical archives and printed sources, in particular, the reports and accounts issued by the electricity utilities. Each graph discloses the prospective levelized costs and thus the cost of the last power station either commissioned or constructed for each respective technology in use. Insofar as any decision is undertaken at a specific point in time, this is the most relevant data that historical actors seek out, simply because it represents the most updated information available

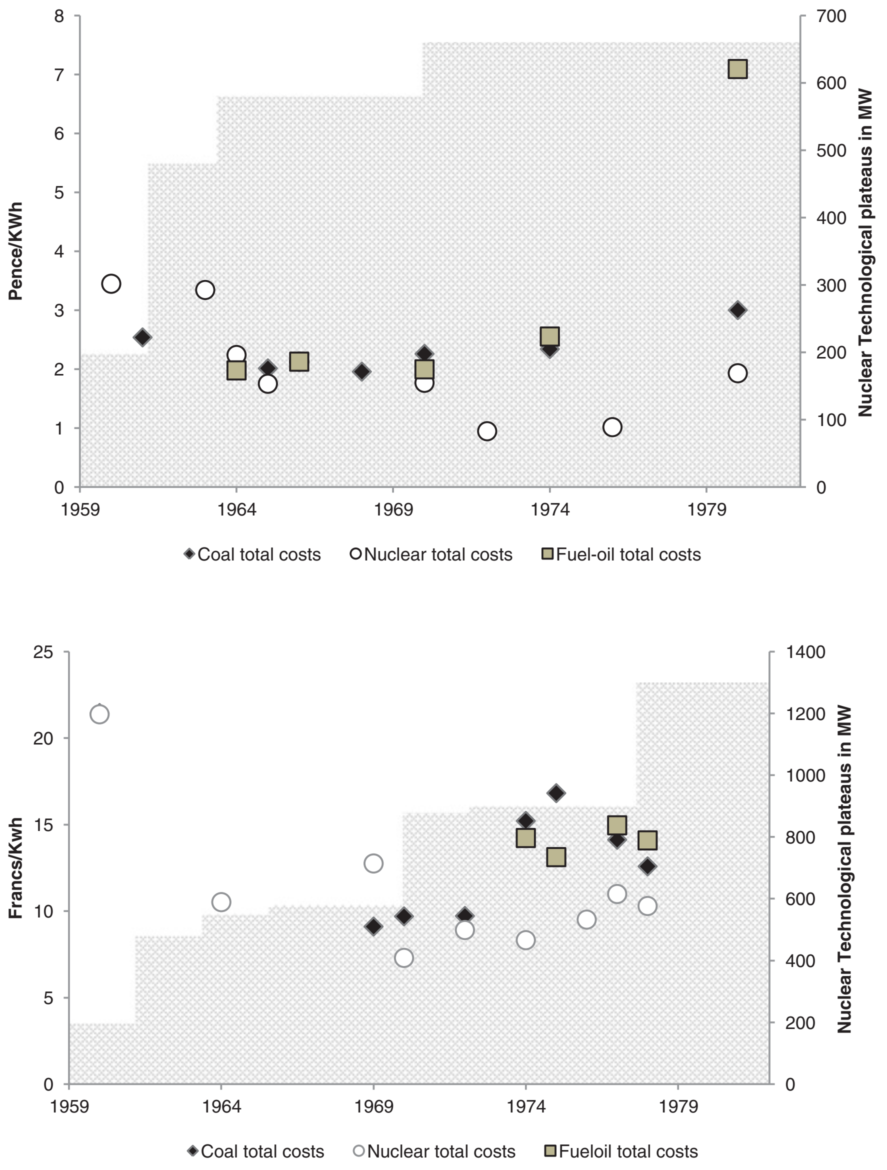


Fig. 9.3 The prospective levelized costs of electricity generation in Great Britain and France (1959–1980). 1978 pence per KWh and 1978 francs per KWh. (see the archival sources in the final bibliography)

at that time. However, if prospective costs are pertinent clues to understanding actions and decisions, they do not cover the total real cost. More often than not, several contingencies occurring during the siting, construction, and/or test phases add further costs to the overall project. For this reason, the results depicted in Fig. 9.3 do not capture the final accounts signed off for the project but rather the conditional evidence upon which the decisions were actually grounded. Finally, the upscaling trend in the evolution of nuclear power's technological plateau is shown in the gray background area of Fig. 9.3 and represented along the right vertical axis.

As all values are expressed in constant 1978 costs, the slope of the downward cost trend since the 1950s reveals the pace of technological progress, economies of scale, and standardization. As previously stated, by the close of the 1960s, this descendent trajectory flattened out. Herewith nuclear costs appear positioned below oil and below coal, indicating the expectations of a breakthrough in atomic energy power costs per KWh. According to the perspective of the historical actors depicted by the graph, two events took place in the aftermath of the 1973–1974 crisis: First, there is the inversion in the overall cycle, ushering in an age of high energy prices; second, the investment prospects for nuclear power become increasingly bright while, further distant in the appraisal, oil and coal exchange relative positions, with petroleum relegated to an uncompetitive ranking.

Although both nations presented similar orderings of their options, the fact is that French levelized costs looked clearly more sharpcut in real terms and displaying larger divergences between the respective life cycle costs of nuclear, coal, and oil stations. From here, it logically follows that the pressure to change the energy mix so as to take advantage of forthright differences in costs appears greater on the French side. The turn of events confirmed this outlook. When the oil crisis hit in 1973, already looming on the French horizon was an ambitious program based on the adaptation of the original Westinghouse-licensed 900 MW pressurized water nuclear reactor design along with a European partnership for the installation of uranium enrichment plant (to be located in *Tricasin*, in the *Rhone-Alpes* region). With the national resource base of competitive coal dwindling year by year, Prime Minister Pierre Messmer moved resolutely to announce a countervailing strategy based on the ordering of 13,000 MW reactors within just two years. The ensuing government headed by Valery Giscard D′Estaing not only confirmed this bet on nuclear but even amplified the vision of French energy independence and security through pro-active policies in the domain of transportation, energy conservation, and energy diplomacy. Unfortunately, not everything ran quite as expected and EDF and the Giscard government soon came across enhanced delays in construction and thus accruing additional interest and other escalations in nuclear power investment costs (Grubler 2010). Despite these burdensome circumstances, there was no retrenchment in the civil atomic program. The momentous turmoil of 1973 embodied the take-off of France as the world leader in nuclear energy and turning it into the country that currently generates the largest percentage share of nuclear electricity alongside Lithuania.

Nothing similar occurred in Great Britain. To put it simply, aside from the sharp attention to the boost in oil consumption and the concessions of North Sea oil fields and aside from their immediate exploitation through the installation of 70 MW sets of gas turbines, little seemed to change in British energy policy. There were some measures aimed at promoting energy conservation which attracted some public discussion but what is hard to find is either the kind of event-reaction sequence or the sense of urgency experienced by the French. Partly, this stemmed from the fact that Britain had itself become an oil producer and with perfect timing given the upward trend in prices which allowed the country to bypass the issues of supply security that beset other nations. Partly, this also resulted from the Labour government of Harold Wilson and his charismatic Secretary of State for Energy Tony Benn being far too occupied in fighting the fires then bursting out all around: wage demands, business failures, European integration, curtailing inflation, and the collapse of sterling (Benn 1990). However, the single most important reason was that the British energy system experienced excess capacity in terms of electricity generation and had consequently fewer degrees of freedom to alter the energy mix: from the 5,500 MW new power generators commissioned at the 1972 peak, the system went into steep decline until a minimum of 57 MW, commissioned in 1978. The slowing of energy consumption brought about by the oil shock further aggravated the braking of investments. Likewise, three of the four new nuclear stations under construction in 1973–1974 were plagued with technical problems and recursive delays amplifying naturally amplifying the misgivings about the real levelized costs of nuclear power. Changes were therefore correspondingly slow and no grand plan saw the light of day. Over time, natural gas and imported coal fulfilled a much larger role in the electricity system transforming Britain into a balanced "four-fuel" economy: gas, coal, nuclear, and oil.

The previous chapters suggest that economic engineering formulas, such as that of levelized costs, may be approached from different angles. Whereas in periods of relative price stability, all concerns centered on improvements that may prompt reductions in capital costs, in periods of higher volatility, particularly when affecting fuel prices, the focus switches to trend forecasting and the implementation of energy-saving procedures. Depending on the historical circumstances, the same formula may be viewed through the lens of "yesterday's costs"—the most updated estimates of existing equipment and installations—or through the lens of "tomorrow costs"—the likely evolution of fuels prices. When the present is deeply ingrained into the past, technology deserves much closer attention; when the present depends on the future, markets become much more important.

In any case, the levelized accounting cost methodology bears a clear resemblance to Bayesian probabilities in the sense that all hypotheses are based on conditional evidence and subject to review whenever new evidence is produced.

9.5 Levelized Costs in a Low-Carbon Economy

Institutional economics often deploys the concept "negative externality" to frame the problem of pollution. Within this framework, pollution represents a side effect of productive activities that becomes particularly relevant whenever the production of Q quantities of good K results in some harmful (i.e., external) effect, whose cost is borne by non-producers. What strikes the economist is the asymmetry of the situation: Pollution does not monetarily affect the individual who is responsible for its emission but still affects the standard of living and the economic position of society as a whole. According to the arguments laid down by the economist Ronald Coase and summarized by David Montgomery, in a well-functioning capitalist system, this is not a problem as institutional economics posits that externalities end up internalized by the transfer of the relevant property rights. The role played by environmental regulation is precisely to compel firms to "internalize the externality" by assigning property rights to the external costs of production (Coase 1960; Gorman and Solomon 2002). Pollution taxes, subsidies, and emissions trading are the policy instruments appropriate to carrying forward this assignment.

Currently, the most common is emissions trading. Under this scheme, the government sets the desired level of environmental quality and allocates only as many pollution permits or credits as is compatible with this predetermined level. No source can emit pollution without a permit or credit. The government then allows firms (or other actors such as individuals, institutions, non-governmental organizations.) to purchase these permits through market transactions. Firms with low costs of control thus reduce their pollution more than necessary so that they can sell their permits. For instance, in the classic European Union "cap-and-trade system," one allowance endows the right to emit one ton of CO_2 equivalent and an absolute quantity limit (or cap) on CO_2 emissions is placed on the 11,000 installations covered by the scheme. At the end of each year, these companies are required to ensure they have enough allowances to cover their plant and installation emissions. To meet this obligation, they may choose to reduce emissions through investments in cleaner technologies, change their output-mix or buy extra allowances on the market so as to cover their cap.

Launched on January 1, 2005, the European Union emissions trading system (EU-ETS) is the leading international carbon emissions abatement policy. Historically, it is also the first international trading system established for large industrial emitters approved in accordance with the goals set down by the 1997 Kyoto Protocol. Unlike regional regulations aimed at fighting problems related to local health and amenities in the 1970s, or unlike state-centered intergovernmental approaches to the damage caused by acid rain in the 1980s, decision-makers in the 1990s recognized that the challenges of global warning had to be tackled through multilevel governance. In effect, CO_2 concentrations in the atmosphere prompted delocalized and systemic changes that threatened Earth's temperature as well as many other chemical, climatological, and biological processes.

The gap between global awareness and state-centered interests has nonetheless widened and too many obstacles stand in the way of a global carbon market. What exists today is an array of emissions trading schemes operating across the world with different scopes, goals, and designs, although broadly linked up with the Kyoto protocol. Foremost among them are the aforementioned EU emissions trading system (ETS), the US Regional Greenhouse Gas Initiative, the New South Wales Greenhouse Gas Reduction Scheme, the New Zealand ETS, and the Tokyo Carbon Trading System (Perdan and Azapagic 2011; Skjærseth and Wettestad 2009).

The adoption of market-like tools for pollution control and the very definition of pollution as a "negative externality" is not without its own criticisms. Some authors claim that there is no inherent tradeoff between efficiency and equity because environmental regulations have not benefited all citizens equally and have instead led to poor and minority communities suffering disproportionate exposure to environmental contaminants and environmental risks; while still others condemned the practice of allocating allowances based on the firm's history of emissions; with still others asserting that these market-based schemes imply that economics, not the public will nor citizen participation would determine the level of air and water quality (Ringquist 2011; Skjærseth and Wettestad 2009). In spite of these criticisms, the bottom line is that even the most utmost capitalist societies seemed to display increasing reluctance for market-like pollution controls. Recent announcements with regard to the US cap-and-trade program indicate that carbon trading is beginning to lose ground within the U.S. Administration (Perdan and Azapagic 2011). In all likelihood, further advances in the next decade will result from the enforcement of the "cap-and-trade system" in the European Union and New Zealand through the inclusion of new sectors (particularly the aviation sector) and the phasing out of the free allocation of allowances.

In practical terms, the full monetization of allowances establishes a new parameter in the levelized cost formula. Since the estimate of the rights to emit tons of CO_2 is a function of both the equipment and technological choices and of the output produced, it makes sense to estimate this parameter along with the capital cost item. In other words, the amount of "allowances" should be levelized over the life cycle of the power plant to take into account the overall cost of CO_2 emissions, given a particular technology.

Finally, global warning has furthermore fostered the expansion of new renewable energies. As a result of breakthrough technical achievements, cost reductions and consistent feed-in tariff policies and direct subsidies, the share of wind power and solar electric generation has been increasing and particularly so in Germany, Spain, Italy, Japan, the United States, and Denmark with India and China expected to follow on the heels of Europe. However, as the installed capacity of wind and sun power mounts, the concept of levelized costs runs into troubles. To a great extent, consumption is becoming covered by fluctuating renewable sources, which need rapid compensation whenever their outputs fluctuate. In this context, crucial questions are clarified neither by information on the "levelized cost" per kWh nor by figures on their average annual contributions. As trainer points out, "what matters most is the capital cost of the quantity of plant required to cope with (a) periods of minimal or zero energy availability, (b) periods of maximum demand, and (c) the required amount of plant redundancy to cope with variability which at times reduces or eliminates contributions from one or more components of the total system" (Trainer 2012: 480).

Even though the issue of the dynamic interaction of intermittent sources and dynamic loads is a key area of research from which further innovations may well be forthcoming, it seems misleading to restrict the accounting budgets to the levelized cost of wind and solar without considering estimations of total system costs. In periods when almost all demand would have to be met by energy sources other than solar and wind, it is necessary to add the additional backup plant costs (generally, flexible gas turbines) or the additional costs of energy storage (equipment and installations for pumped water storage, compressed air energy storage or batteries and capacitors) to the KWh cost of renewables (Arent et al. 2011; Trainer 2012). In terms of future availability, the pace of diffusion for wind and solar technologies is critically interrelated to the invention and deployment of costless storage equipment.

Bibliography

Archival Sources

National Archives (NA), London

NA.- Cabinet Papers (1955) A programme for nuclear power. Minister of Fuel and Power—Geoffrey Lloyd, Feb 1955

NA. EG-1-60 (1955) White paper correspondence and reports, 31 Jan 1955

NA. EG-20-3 (1974) Iain K. Forster, Fuel policy, June 1974

NA. POWE 14-740 (1954–1955) Documents and correspondence on the trend Report. 17 Jan 1955

NA. POWE-14-741 (1953) Atomic energy for civil purposes: note by the Ministry of Power. Note of meeting in BEA offices. Letters to Lord Cherwell. 4–11 May 1953

NA. POWE-14-1407 (1963–1965) Estimated costs of power stations. 18 Mar–3 Apr 1963

NA. POWE-14-2744 (1972) Construction costs of CEGB fossil-fired power stations 1950–1972. 20 Dec 1972; Note of a meeting to discuss energy policy in Thames House South, Millbank, Wednesday 21 June 1972

Archives Nationales de France (ANF), Fontainebleau

ANF. Box 19910737 Dossier 3 (1959) Gibrat, R. La filiére des reacteurs de puissance a uranium naturelle, graphite et gaz sous pression, allocution faite a Essen le, 25 Sep 1959

ANF. Box 19910737 Dossier 1 (1970) VI plan, Groupe prospective de la commision de l'enérgie, L'offre de pétrole, 14 Jan 1970

Archives Eléctricité de France (AEDF), Blois

AEDF. 800545 File 27243 (1970) Centrales françaises en service et en construction, 25 Mar 1970

AEDF. 800621 File 27319 (1971) M. Autesserre, Le probléme central de l' environnement pour EDF. 18 Aug 1971

AEDF. 800629 File 27327 (1956) D. Jung. Intervention des centrales nucléaires

AEDF. 800624 File 27322 (1971) Resumé des principales parametres economiques relatives aux tranches nucléaires, 14 Avril 1971

AEDF. 943408 File 39726 (1958) Étude technique des centrales thermiques. EDF—Direction de l'Équipement

References

Ailleret P (1954) Les perspectives d'intégration de l'énergie nucléaire dans les moyens de production d'énergie électrique. Revue Française de l'Energie 51:203–211

Arent DJ et al (2011) The status and prospects of renewable energy for combating global warming. Energy Econ 33(4):584–593

Benn T (1990) Against the tide. Diaries 1973–1976. Arrow Books, London

Brussière E (ed) (2003) Georges Pompidou face à la mutation économique de l' Occident 1969–1964. PUF, Paris

Clapp BW (1994) The environmental history of Britain since the industrial revolution. Longman, London

CEGB (1973–1974) Statistical yearbook. Central Electricity Generating Board, London

Coase RH (1960) The problem of social cost. J Law Econ 3:1–44

Cochran TB et al (2010) Fast breeder reactor programs: history and status. Available via http://fissilematerials.org/library/rr08.pdf. Accessed 22 May 2012

Cockcroft J (1953) Nuclear reactors and their applications. J Inst Electr Eng 100(3):117–120

Cockcroft J (1957) The development of nuclear power. J Br Inst Radio Eng 17(5):245–252

Cooke S (2009) In mortal hands—a cautionary history of the nuclear age. Bloomsbury, New York

Davies R (1984) The sizewell B nuclear inquiry: an analysis of public participation in decision making about nuclear power. Sci Technol Human Values 9(3):21–32

Dürr M (1996) Le tournant nucleaire d´Eléctricité de France. In: Morsel H (ed) Histoire de l´électricité de France, vol 3. Fayard, Paris, pp 693–783

Economic Statistics Database—ECONSTATS (2010) Economic statistics for France and Great Brirtain. Available via http://www.econstats.com/. Accessed 3 Oct 2012

Frenken K, Leydesdorff L (2000) Scaling trajectories in civil aircraft (1913–1997). Res Policy 29(3):331–348

Foasso MC (2003) Histoire de la sûreté de l’énergie nucléaire civile en France (1945–2000). Dissertation, Université Lumière–Lyon II

Gorman HS, Solomon BD (2002) The origins and practice of emissions trading. J Policy Hist 14(3):293–320

Gowing M (1978) Reflections on atomic energy history: the rede lecture. Cambridge University Press, Cambridge

Grubler A (2010) The costs of the French nuclear scale-up: a case of negative learning by doing. Energy Policy 38(9):5174–5188

Hall P (1986) Governing the economy: the politics of state intervention in Britain and France. Oxford University Press, New York

Hecht G (1998) The radiance of France—nuclear power and national identity after World War II. MIT Press, Cambridge

Hinton C (1957) The axel axson Johnson lecture. Ingeniörs Vetenskaps Akademien, Stockholm

Hinton C, Brown FHS, Rotherham L (1960) The economics of nuclear power in Great Britain. In: World power conference, Sucs. Rivadenevra, Madrid, pp 1–24

Howland N, Daigle C (ed) (2011) Foreign relations of the United States, 1969–1976, V. 25, Arab-Israeli crisis and war. State Department—Bureau of Public Affairs, Washington. Available via http://history.state.gov/historicaldocuments/frus1969-76v25. Accessed 15 Oct 2012

Jukes JA (1955) The cost of power and the value of plutonium from early nuclear power stations. In: Proceedings of the U.N. international conference on the peaceful uses of atomic energy, Geneve, vol 1, pp 322–326

Komanoff C (1981) Power plant cost escalation: nuclear and coal capital costs, regulation and economics. Komanoff Energy Associates, New York

Lamiral G (1988) Chronique de trente années d’équipement nucléaire à Electricité de France. AHEF, Paris, tomb I

McCabe MJ (1996) Principals, agents, and the learning curve: the case of steam-electric power plant design and construction. J Ind Econ 44(4):357–375

Perdan S, Azapagic A (2011) Carbon trading: current schemes and future developments. Energy Policy 39(10):6040–6054

Picard J-F et al (1985) Histoire de l’EDF. Comment se sont prises les décisions de 1946 à nos jours. Dunod, Paris

Robens L (1972) Ten year stint. Cassell, London

Ringquist EJ (2011) Trading equity for efficiency in environmental protection? Environmental justice effects from the SO2 allowance trading program. Soc Sci Q 92(2):297–323

Roberts F (1999) 60 years of nuclear history-Britain’s hiden agenda. Jon Carpenter Publishing, Oxfordshire

Sheail J (1991) Power in trust: the environmental history of the Central Electricity Generating Board. Clarendon press, Oxford

Skjærseth JB, Wettestad J (2009) The origin, evolution and consequences of the EU emissions trading system. Glob Environ Polit 9(2):101–122

Trainer T (2012) A critique of Jacobson and Delucchi's proposals for a world renewable energy supply. Energy Policy 44:476–481

UKAEA (1957) The nuclear power programme. Atom 5:5–17

Williams R (1980) The nuclear power decisions: British policies, 1953–1978. Croom Helm, London

Wilson C (2012) Up-scaling, formative phases, and learning in the historical diffusion of energy technologies. Energy Policy 50:81–94

Yvon J (1955) La Pile de Saclay, expérience acquise en deux ans sur le transfert de chaleur par gaz comprimé, par J. Yvon communication du C.E.A. à la conférence de Genève, août 1955. Centre d'Études Nucléaires, Geneva

Chapter 10
Overview

Abstract The overview provides a revision of the main concepts presented in this book from the generally held perspective and understanding of technological, scientific, and policy issues.

10.1 As Seen from Below

It was the possibility of converting heat into mechanical effect (mechanical work) that paved the way for the overarching concept of energy. Initially, this was perceived as the conservation of living force in the passage from one source into another. Later, thermal transformations were explained by resorting both to the principle of conservation and to the principle of dissipation since part of the original heat was neither destroyed nor conserved but simply lost in the process. By the mid-nineteenth century, physicists posited that if motion, heat, light, electricity, gravitational force, and chemical energy could be reciprocally converted, it was because they were not separate phenomenon but instead different physical manifestations of the "ability to do work." The bond between motion, heat, light, electricity, gravitational force, and chemical energy was consequently abridged into a single word: "energy." Once taken, this step changed the perceptions of science, policy, and business. Materials and productive processes hitherto separated were brought together, converted into a common standard, and added, to produce new categorizations of human activity. The very existence of things now depended on their being fueled either on the microscale (the dietary standard of ingested "calories") or on the macroscale (the total primary energy consumed at the entry of energy systems).

In the daily life of the common man, however, the all-encompassing energy concept remained senseless. Each activity seemed drawn from its own separate source, depending on its respective trade dealers, and there was little practical chance to convert or reshuffle one energy source into another. Aside from the

N. L. Madureira, *Key Concepts in Energy*, DOI: 10.1007/978-3-319-04978-6_10,

possibilities of burning wood, charcoal, or coal to produce heat for cooking and warming the house, all other inter-fuel substitutions were rare. What drove ordinary life was not abstract “energy”, but kerosene, whale or colza oil acquired in grocery stores, charcoal or coal bought in hardware stores, and an increasing amount of muscular force deployed by man, horses, mules, and oxen. For sure, railways and steamships had in the meantime replaced the work of stallions with wind force, giving way to the mechanical power of steam engines even while sailing ships still continued to carry on their trades and railway companies were the largest world recruiters of teamsters for loading and transportation within the cities: complementarity ruled out substitution. In this context, “energy” was a catchword for physics treatises and for poetry but much less so a token of the workman’s experience. Only with the installation of the first electrical power stations, near the center of large towns, did people foresee that the same thermal source could be used simultaneously to produce light and industrial power and, sometimes, electric traction. Certainly not by accident did many companies adopt exactly the label “Light and Power Co” to publicize their services. For decades, their scopes of activity were demarcated to the commercial districts of downtown centers and a few enterprises.

Less widely recognized but of equal importance, oil ceased being a specific source for lighting and became the hotbed of light, power, and motion. Experiences with internal combustion engines led to the progressive discarding of fuels like alcohol, ether made from alcohol and sulfuric acid, benzol obtained from coal tar, kerosene, and crude wood naphtha mixed with kerosene, in favor of gasoline as the best propelling source for “horseless carriages.” The new market for automobile gas-filling grocery stores grew quickly and soon surpassed traditional sales of illuminants produced from the lighter fractions of crude oil. On the other hand, the very invention of the first successful applications to pulverize raw oil or fuel oil and blow it in spray form into a furnace introduced a new power source into the economy of steam engines, prompting the inter-fuel substitution of coal and oil. This hybridization took place at the moment of consolidation of the dominant steam design technology but only reached significant importance in oil-producing countries where it did prove a real alternative to coal fueling. In spite of this, by the dawn of the twentieth century, petroleum usage had undergone important changes, surfacing not only as the most flexible energy source but also that technologically best suited for conversion into light, industrial power, and traction. For the common man, oil unveiled the principle of energy conservation in completely different final usages just as much as electricity did. And the more inter-fuel substitution progressed, the more flexible, cost-effective and integrated the market for these commodities became. The idea that a higher level of conceptualization could account for the seamless branching of powering activities accrued to the practical recognition of the “energy” concept.

Just as old technologies were adapted to new power sources, making possible the hybridization of steam engines, they furthermore benefited from the spurt of innovation brought about by more advanced competing technologies. The case of the charcoal iron industry, whose resilience lasted throughout the nineteenth century, is an example of the persistence of traditional wood fuelling side by side with the "modernized" coal industry sector. As mentioned before, the course of actions that lead to destruction through first-mover innovation and cost decreases also unleashed creative market mechanisms and the opportunities for incumbent manufactures to stage fight backs. In the first place, the imitation of leading coal innovations introduced both in puddling and rolling, and in blast furnaces opportunities for the charcoal industries to enhance their scales of economy and energy saving efficiencies. Secondly, the market expansion caused by lower prices gave charcoal producers a boom in demand from which they could take advantage to sell high-quality and differentiated products. The crucial factor was nonetheless the positive effect of resource endowments upon prices and upon the choice of technology. This was a somehow unique historical circumstance because the abundance of woodlands relied on the displacement, whether further north or further west, of the natural resources frontier. Owing to this momentous opportunity, nineteenth century entrepreneurs could seize raw materials still untapped and benefit from copious local resources.

The expression "age of coal", often employed to characterize the nineteenth century energy transition, roughly translates the idea that coal's edge was in fact punctuated by technological hybridizations and by complementarity with other power sources. It proceeds along a chain of adjustments, checks, and balances rather than by some overwhelming dominance. During the entire period, the set of technical possibilities remained widely open, blossoming into an atmosphere filled with hope and buoyancy but also with apprehension and uncertainty.

Societies feared, first and foremost, that the dislocation of the core engine of growth toward exhaustible fossil fuels might push humanity toward a trapped deadlock: With fossil fuel consumption mounting to unprecedented levels and with no clear idea of the remaining stocks underground, there was a growing sense of alarm as to whether the resources might run out in the near future. How long could steam engines and internal combustion engines run on the foreseeable reserves? The commissioning of the first national geological surveys consequently became a key event for the awareness of the environmental path taken by industrial societies. Undertaken by specialized public agencies, the 1861 coal survey in the United Kingdom and the 1909 oil survey in the United States represented milestones in environmental history. The public disclosure of the results of both assessments for the first time revealed that, at current production levels, reserve stocks would last 10 years. It is important to note that dependent upon the amount of physical reserves and the time lag before depletion, the systematic character of the geological survey exposed the unavoidable reality of scarcity. People interpreted the path brought about by industrialization as a dead end: Sooner or later, countries would have to face the end of their inventoried stock and thus also lose out on the riches on which their economies had become increasingly dependent. Moreover, the explanation was biased by an overvaluation of what was geologically known relative to that unknown and neglecting the potential for untapped resources. By this means, the systematic inventorying of natural resources became a source of uncertainty: Geologic knowledge resulted in more doubts about the future with all concerns converging over the guesswork involved in estimating the quantity of yet-to-find fossil fuels.

Political, economic, and social anxiety changed the awareness over the finiteness of fossil fuels and, more fundamentally, changed the very concept of energy reserves. In the United Kingdom, the looming uncertainty of the 1860s and 1870s paved the way for new probabilistic assessments of mineral patrimony. Geologic surveys began reckoning in the coal left behind in open pits; still undiscovered coal and the coal that might be possibly extracted were new technologies to be invented in the future. All these parcels added up to different categories, namely existent, probable, and possible reserves. In turn, more parcels, more categories meant extra cushions likely to stretch the life span through to depletion. From this perspective, the postulate of any given fixed asset of coal reserves opened up major controversy as this concept was directly at odds with the ongoing developments in geologic knowledge. What is more, the threat of depletion, set forth by the theoretical mechanism of the rebound effect, had the practical consequence of stepping up the geologic categorization and quantification of uncertain energy reserves. Whereas economic science posited presumptive evidence, geologic practice was already utilizing probabilistic evidence (evidence about the circumstances and consequences of geological events). However, perhaps the most important point to stress is that the outstanding progress made in the domain of gathering such information remained almost unnoticed: In the face of contradictory information, and different estimates for depletion, fears about scarcity seemed to have gained in strength rather than fading. In sharp contrast with the categorization of coal reserves, oil reserves in the United States refused the probabilistic path since the major aim of the leading institution—the American Petroleum Institute—was to replace geologic uncertainties by a narrow but accurate appraisal of oil reserves. Between 1925 and 1935, all open possibilities were locked into the concept of "proven reserves", grounded on technical and economic feasibility. Such closure ran against the grain of current technological improvements as it excluded probabilistic methods of oil finding by the geophysical sciences and neglected the still novel enhanced oil-recovery practices.

All in all, it seems that collective fears regarding fossil fuels depletion were more of a matter of elite-based speculation rather than of concern to the common man. From the street point of view, the market gauge was the best way of assessing the state of natural resources. The credibility of geologic warnings about forthcoming scarcity could be called into question when the prices of basic fuels experienced depreciation in real terms or whenever in news broke about the discovery of sizable reservoirs/pits. How could exhaustion be a tangible threat when oil was flowing from several oilfields at an unprecedented pace? How could the conservationist appeal catch the average man's attention when coal prices remained as low as ever driving an increase in consumption? This was precisely the situation that followed both the release of the British coal survey in 1861 and the US Oil survey in 1909: In constant prices, the ensuing years witnessed a downward trend in the price of coal and oil. For the common man, this could only mean that regardless of the debates, whether in Parliament or in Congress, business "as usual" continued. There were no impending signs of scarcity.

In the end, the optimists prevailed. Particularly between 1930 and the mid-1960s, an era of irredeemable optimism dawned in which everything seemed finally achievable. This was the "free lunch" time in energy demand. This free lunch was made possible by the combination of huge discoveries (US 1920s discoveries; 1940s tapping of Middle East oil; nuclear take-off; the commodification of natural gas), lower prices, and cost-oriented technological agendas (networking; up-scaling). One way or another, the contemporary energy world as we know it was shaped and moulded by this historical framework.

Established principles for recovering costs and efficient pricing, discussed within the electricity generation industry since its very earliest days, began to wane. Instead of the transposition of cost accounting classifications into the bills of customers so as to make them pay for the "standing" and "running" costs covering constant and variable capital, other billing schemes came to the fore aiming rather to promote consumption, to reduce regional variations and to democratize citizen access to modern energy. The drift away from competitive pricing was thoroughly accompanied by the commitment to commercially orientated rates. Moreover, European managers of electrical undertakings were caught up in a tide of pressure from public opinion and from governments that stepped into demand fair energy supply prices and the extension of the benefits of progress to the least well-off segments of society. Regardless of national diversity, the same purpose loomed everywhere: boost consumption, even at the cost of further raising peak demand. At the policy level, the "free lunch" view imbued the strategy of accelerated industrialization pushing the economy into sectors with intensive energy demands per unit of output. Metallurgy, chemicals, and machinery production were key sectors to heavy industrialization strategies. In all these visions, the scope for stepping up capital accumulation mattered much more than energy, which was relegated to a secondary and supportive role. By the 1950s, modernization by import substitution seemed further justified by the pursuit of gigantic enterprises on the assumption that any increase in the size or the performance capacity of a given technology yielded a downward slope in average costs. Economies of scale and economies of learning placed increased emphasis on technical expertise so as to accomplish accurate cost forecasts and accurate consumption planning. To account for these challenges, the methodology of levelized electricity cost inaugurated a cycle of comparative investment decision monitoring. Super technologies were becoming difficult to control from the logistical point of view and politicians learned to keep an open eye to accounting reports. Tighter and more accurate forecasts were nonetheless hard to come by, particularly in the case of nuclear plant construction assignments that involved many different suppliers and faced unforeseen drawbacks behind each door.

The free lunch view got painfully swamped by a cascade of events: stalemate in the advantages of up-scaling, the emergence of diseconomies of scale, far-reaching environmental awareness, the public debate of energy projects, rising fuel costs, rising capital costs, the disclosure about pollution and climatic impacts, and catastrophic nuclear accidents—everything seemed to contribute toward dismissing the idea that energy represented some second order factor in human life. The call for public participation in energy decisions surfaced slowly in the 1960s, sponsored by counter-culture movements and by environmentally minded actions undertaken by local communities that fought back against the intrusion of central power stations or, at the worst, nuclear stations in their backyards. Local protests driven by amenity considerations began to coalesce with nationally structured political institutions in the 1970s, a trend reinforced by the increasing media coverage of scientific environmental debates. Little by little, marginal green thought began being assimilated by mainstream politics. Soaring prices also led to the first widely publicized energy conservation campaigns. Altogether, these factors unveiled the rough fact that energy usage was not inconsequential: The concepts of energy demand their own reflection and action.