form of thinking about measuring the impact of various elements on the environment, and is designed with the form:

$$IPAT \rightarrow Impact = Population * Affluence * Technology$$

Impact, in this case, is defined more broadly than just energy use and can measure almost any aspect of resource absorption or its effects and is often used to measure the environmental impact of society's industrial choices. It suggests that limiting the impact of energy resource choices (such as pollution or resource absorption) across a fixed population can only be achieved through either dramatically improved technology or an eventual loss of affluence.

This relationship is identical to the one that determines society's relationship to the amount of energy it consumes. Thinking about impact as the amount of energy needed, or the amount that can be procured, presents the same type of structural limit. Society's ability to get richer is limited by the amount of energy that can be captured and the efficiency (i.e., the technology) with which it is used to create economic output. If there is a limit to the amount of energy procured and technology does not improve to efficiently convert that into income generation or production, that limits the amount of affluence that can be expected.

Though conceptually useful as a starting point, these simple frameworks fail to provide the machinery to fully appreciate the complex relationship of the volume of energy required to meet growing needs and their economic and societal costs and benefits. Unpacking those questions will require a more complete analytic toolkit.

C. What Are the Key Questions to Examine in the Energy System?

Given how vital the human relationship to energy is, a comprehensive treatment of the energy sector remains elusive. At best, the models used to explain it are incomplete and insights only partially formed. At worst, what is presented as analysis can be unintentionally, or even intentionally, biased.

The problem of analysis and developing insights about energy is initially complicated by a lack of common language across sectors and stakeholders that influence aspects of the energy system. The tools of business are not the same as those of government, and consumers are rarely fully informed about the tools or processes of either. Poorly formed understanding and partial analysis are especially dangerous when so many different sectors of society must collaborate to achieve optimal outcomes of resource allocation and collective prosperity.

Additional issues arise from the existence of a combination of different goals and limited information of stakeholders. Categorically, and especially true in the energy system, prescriptive claims that argue for an optimal course of action or best choices in any given situation are based on limited analysis, sources of data, or previously held norms about what is right or wrong and so will tend to favor the perspectives of only some of the stakeholders in that system. Making good analytic conclusions and decisions about the energy system requires understanding the range of biases stakeholders may possess and appreciating that every decision individual actors make is subject to a lack of complete (or outright incorrect) information.

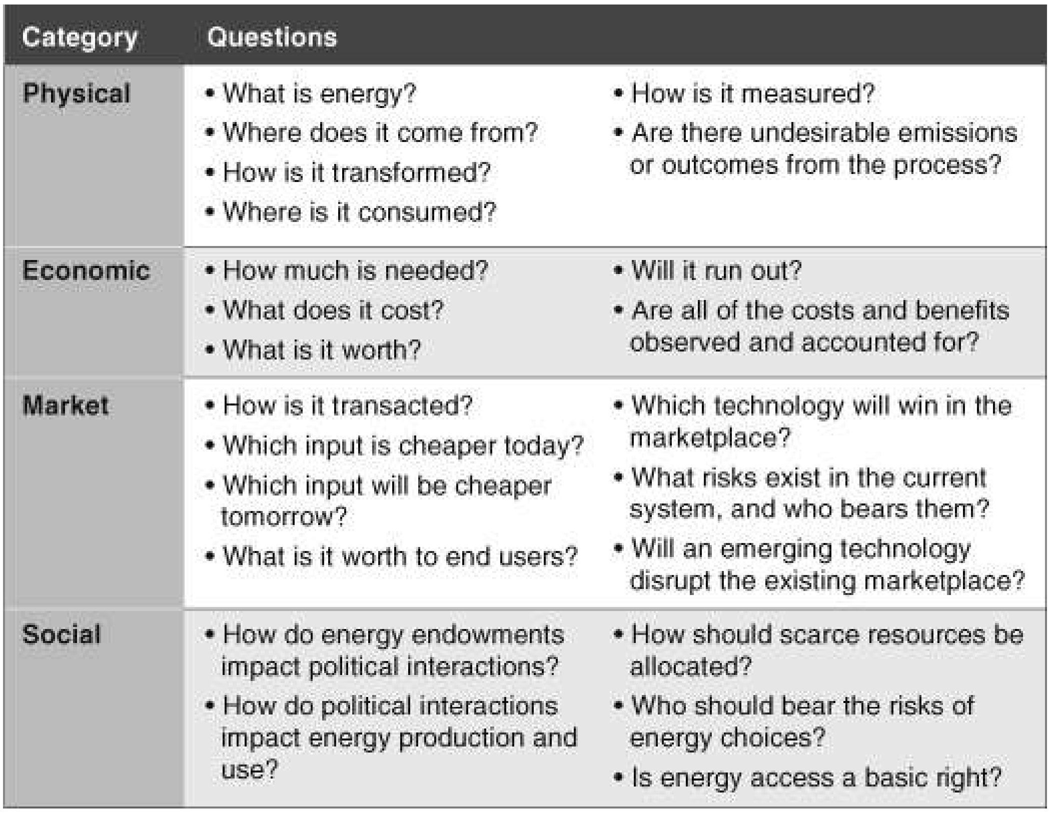
Unfortunately, such partial analysis and claims can sometimes settle into rules of thumb or **memes** that people use to understand events or predict outcomes—examples might include that the lowest energy cost solution will always win in the marketplace, that intermittent renewables can never power an economy's total energy needs, or that there is a natural tradeoff between clean energy and cheap energy. Even if occasionally locally true, these memes are rarely generalizable, and they can lead stakeholders to make unexamined or incorrect assessments of observed phenomena. Despite evidence to the contrary, memes can persist over large populations and through time until they can be overwhelmingly disputed—altering behavior and investment decisions. One of the primary goals of energy system analysts is to detect and correct these memes in others' work, as well as in their own.[17]

To be effective at parsing these issues in energy system analytics requires a more fundamental and integrated understanding of related tools and methods. This text provides the tools necessary to analyze the energy sector in all of its complexity. It integrates the various disciplines that provide a critical perspective on the dynamics at work. However, the establishment of a critical-thinking toolkit must begin with a commitment to clarity and objectivity, including the following tools:

- A primary commitment to **positive analysis**, which is fact-based and objective, vs. **normative analysis**, which is subjective and value-based[18]
- A description of the system's current structure, including how and why it operates (its present)
- An explanation of how the system's status quo came to be that way (its past)
- A forecast of the future outcomes and scenarios for relevant questions (its future)
- Only once all of those have been established, an argument for normative outcomes to support or avoid forecast pathways by changing allocation of scarce resources

It is not the intention of this approach to avoid any of the normative questions that are so critical to moving industry and society forward. Instead, beginning with a positive approach is intended to ensure that the normative questions are informed by clear frameworks and correct facts, and that “facts” are not informed—intentionally or otherwise—by the normative conclusions.

Before embarking on any new study, it is useful to think about the questions that need answering. In the energy system, a number of questions arise that require continually improving methodology to address. These questions range from the physical to the economic to the societal, and a sampling of them is provided in Figure 1.7.



Category	Questions	
Physical	• What is energy? • Where does it come from? • How is it transformed? • Where is it consumed?	• How is it measured? • Are there undesirable emissions or outcomes from the process?
Economic	• How much is needed? • What does it cost? • What is it worth?	• Will it run out? • Are all of the costs and benefits observed and accounted for?
Market	• How is it transacted? • Which input is cheaper today? • Which input will be cheaper tomorrow? • What is it worth to end users?	• Which technology will win in the marketplace? • What risks exist in the current system, and who bears them? • Will an emerging technology disrupt the existing marketplace?
Social	• How do energy endowments impact political interactions? • How do political interactions impact energy production and use?	• How should scarce resources be allocated? • Who should bear the risks of energy choices? • Is energy access a basic right?

Figure 1.7 Framing Questions for Energy Systems Analysis
Source: Author.

A quick survey of these questions suggests that the answer to any might require insight into another. For example, understanding market dynamics will certainly require a basic grasp of the economic characteristics of the competing options. Determining the desirability of physical transformations of energy from one form to another would have to be informed by the economic opportunity of doing so.

This is one reason that the question “Will it run out?” is listed under the economic questions, as the integrated analysis conducted throughout this book will show that the question is one of economics and not physical processes. It is precisely these boundary relationships among various disciplines that this comprehensive view of the energy system is designed to explore.

D. Which Tools Are Most Useful?

A wealth of analytical capability can be brought to bear on questions about the energy system. Most of the tools needed already exist within the various disciplines that study some aspect of the questions in Figure 1.7. Once they are available as a full toolkit, they become a very powerful combination for developing insight into the answers to those questions.

1. A Bountiful Toolkit

The questions in Figure 1.7 are already divided into groups that track dominant disciplines or fields of study—including science, technology, social science, business, markets, policy, and society. Within those fields, however, many skills and disciplines might be necessary. These include but are not limited to:

- **Science and engineering**—Physics, chemistry, material sciences, thermodynamics, mechanical engineering, civil engineering, chemical engineering, information technology, monitoring and measurement
- **Environmental science**—Atmospheric science, hydrology, geology, mining, agriculture, ecology, carbon accounting
- **Business and economics**—Microeconomics, macroeconomics, behavioral economics, manufacturing, finance, management, capital markets, supply chains, commodities, currency markets, marketing, data analytics, corporate strategy, entrepreneurship
- **Law, policy, and society**—Law, trade, government, public policy, public administration, international affairs, development, welfare, poverty

Clearly, expertise in all of these disciplines and specializations is beyond the capabilities of any single person, but finding a way to integrate them into a comprehensive understanding remains vital to deeper understanding of the energy system.

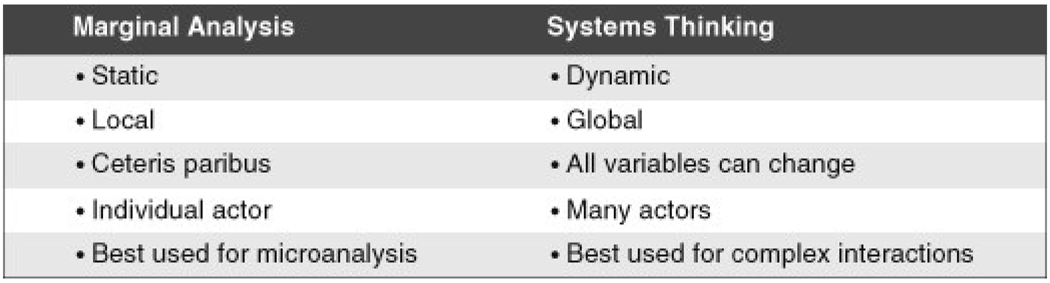
Fortunately, being an expert in each of the disciplines is not necessary to be able to understand how they all work together. Just as a family doctor can diagnose the whole body while relying on medical specialists to treat acute problems within their specialty, a well-informed analyst of the energy system, regardless of discipline, should be able to see how the individual parts of the whole complex system interact. Conversely, specialists in the medical field are required to have the same basic general training of the entire body as a family doctor has, without which the context of their specialty would be dangerously absent. Specialists in any of the disciplines listed above that touch the energy system would also be dangerously underinformed if they could not clearly describe the relationship between their discipline and the other related disciplines.

2. Systems Thinking

As the number of disciplines and the number of actors involved grows, the interrelationships and interdependencies can become complex. Therefore, analyzing larger and more complex systems will require new tools to integrate all of these elements into a robust and comprehensive framework. These tools must be useful in examining each of the parts and also show how they are interrelated and contextualized.

Ultimately, energy is best understood as a set of interconnected systems, which are collectively referred to as the **energy system**. The object of analysis becomes these system elements, including many parts, subsystems, and interactions. Being able to view this aggregated and interactive system requires developing a **systems thinking** approach that is distinct from the traditional marginal analysis that populates much of economics and social sciences.

Marginal analysis, which simplifies a relationship to a few variables that can be analyzed by holding all other variables constant (ceteris paribus) has been a bedrock of analytics in these social science fields and is an incredibly useful tool. It explains individual behaviors very well, can be used for allocation decisions of producers, and defines the rate of change at a specific point under local conditions. Marginal analysis remains a very useful tool for understanding the energy system. However, when multiple actors are involved and all variables are susceptible to change simultaneously, it is also necessary to think in terms of systems. (Figure 1.8 describes some basic differences in these approaches.)



Marginal Analysis	Systems Thinking
• Static	• Dynamic
• Local	• Global
• Ceteris paribus	• All variables can change
• Individual actor	• Many actors
• Best used for microanalysis	• Best used for complex interactions

Figure 1.8 Marginal Analysis vs. Systems Thinking
Source: Author.

a) The Tools of Systems Thinking

To begin, systems thinking requires establishing an integrated model of how the observed phenomena interact. **Models** are constructed representations of how some elements of the world operate in relation to each other. As such, a model is an abstraction of reality, though hopefully a useful one. Being useful in this analytic context suggests that it can explain the correlation or causality of observations and might even be used to help predict outcomes.

The simplest idea of a model is the kind found in marginal analysis that relates two things to each other. For example, it might be the relationship between a customer and a seller or perhaps the relationship between the amount of raw material available and the amount required to achieve a desired amount of final product. Each of these individual relationships links a small number of elements rigorously and, ideally, in a well-defined way.

A system is a model constructed by weaving together a number of these individual relationships into a more complex set of relationships. It develops a more comprehensive set of relationships, feedback mechanisms, and transformations that must be mutually satisfied. These relationships are dynamic and changing and can even alter the nature of the system itself over time. Systems hold very little constant, instead allowing all actors (including individuals, firms, policy makers, etc.) to simultaneously react and optimize their decisions based on their individual goals and preferences.

So, a system is just a model, though typically a more complex model than that used in marginal analysis, but with rules for integration that allow it to be consistent and useful. A basic open system will encompass some **inputs**, some internal transformations and processes, and some **outputs**. The internal transformations and processes will relate some inputs to some outputs and specify under what conditions the transformations will take place. These relationships can get complex as outputs of one transformation become inputs to another. These can split or combine and even turn back on themselves to add texture and complexity to a given system. Grouping or subdividing elements in a logical and consistent manner describes the relevant parts of the system, and enhances the ability of a system model to explain or predict behavior.

However, complex systems are much more than the collection of parts. When viewed as a whole, systems can exhibit dramatic and unexpected behavior. Just as organs do not describe how the various bodily systems—such as circulatory or nervous or respiratory systems—work, and understanding only these systems fails to adequately capture how they integrate into a whole human organism, higher-level systems take on forms and properties that must be understood on their own merits.

A closer examination of systems leads to a great appreciation for how they behave and respond to stimuli. These **system dynamics** and the often-surprising outcomes that emerge reveal that systems can grow or shrink, they can learn and adapt; causal relationships can comprise many steps; and the outcomes can seem counterintuitive. Examining system dynamics opens very interesting avenues for exploration, regardless of which system is being observed. These include:

- **System structure**—The system can be viewed as a collection of components at any given moment in time. These components have natural groupings and relationships and can provide a geographic "map" of the system structure.
- **Transformations**—Once the structure is established, it is useful to observe and understand the transformations within that structure as time passes or elements change. The strength of these relationships and the direction in which they flow can explain dynamic behaviors. Systems are best understood not in how they are but in how they change.
- **Leverage points**—Because systems are interconnected, any point can be affected by many others. Not all of these will have an equal effect, as the strength of the transformations may vary, particularly across a number of relationships or structural elements. Identifying where small efforts in one part

of the system can create major change in other parts of the system allows the identification of leverage points.

- **Nonlinearities**—Sometimes, dramatic change can occur but only after a delay and not linearly. Systems often maintain themselves until certain thresholds are reached, and then system dynamics can radically alter their behavior. Observing and predicting these nonlinearities reveals much about the system itself.
- **Root causes**—Explanations of certain observed behaviors may have many levels. Sometimes there is an immediately obvious reason, but the behavior is usually motivated by deeper relationships in a system. An analogy would be evaluating a patient's symptoms vs. the disease, and uncovering the root cause of the observed phenomenon can be enabled using system dynamics.

These are the tools of systems analysis, though they represent as much a shift in mindset and approach as a simple set of tools. Correctly using these tools reveals a set of system-level dynamics and behavior beyond the observed outcomes for any single part of the system.

b) Defining the Energy System

To apply systems thinking to the energy system, or any system, it is important to define system boundaries. There is no scientific method for determining what these are, but the approach should be to try to encompass the scope of the phenomena that are of interest, and no more. Figuring out what is of interest remains an art, but ideally it will include, either within the systems model or as a clear input to the system, anything that might have a material impact on its behavior.

A rigorous formulation of the energy system is developed in Chapter 2, but it is useful to examine the elements that make it work and those that influence it. The main components of the energy system include:

- **The energy supply chain**—This represents all of the energy in the human-industrial system, from total energy supply (inputs) to final energy consumption and energy services (outputs), and is the basis of energy system analysis. It also includes the physical delivery system (**infrastructure**) to move and transform the energy from its origin to its final disposition.

- **The economy (i.e., the economic system)**—The energy system does not operate independently from the rest of society but rather is part of it and the overall economy. Understanding how economics, markets, and the policies that regulate them play a role in initial resource allocation, endowments, capital investments, and consumption will provide information about how the energy system behaves. This system can also be referred to as the macroeconomy.
- **The ecosystem (or natural resource system)**—Both the energy supply chain and the economy sit within the broader ecosystem, on which they rely to provide vital resource inputs and into which the energy supply chain sinks the wastes and outputs of its transformations.

Each of these elements can be examined as systems on their own merits, each with its own inputs and outputs and dynamics. In this book, describing and analyzing the energy system will proceed using energy as the primary focus of observation, defining its relationship to other systems in order to explain energy system dynamics.

c) Subsystems Within the Energy System

An important distinction when examining systems is the degree of openness that they exhibit. A system can be characterized as an **open system** or a **closed system**. The technical distinction between open and closed systems is that an open system is continually influenced, informed, or constrained by the activities of elements outside the system, whereas a closed system receives its endowments when it is set up and then remains largely isolated from outside influences.[19]

By this definition, nearly all of the systems listed above are essentially open systems, which means that they are, at times, both enabled and constrained by activities in other systems. The energy supply chain takes inputs (resources and capital) from the natural resource system and the economy and sends its outputs (economic productivity and waste products) back into those systems. These systems can more specifically be described as

nested systems, meaning that one fits easily inside another, both of which fit inside a third. Figure 1.9 shows how these systems—the energy supply chain, the economic system, and the ecosystem—are nested. A number of acronyms and relationships depicted in this figure will require further explanation in subsequent sections, but it is the first step in building an integrated map that links these basic systems into an analytic framework.

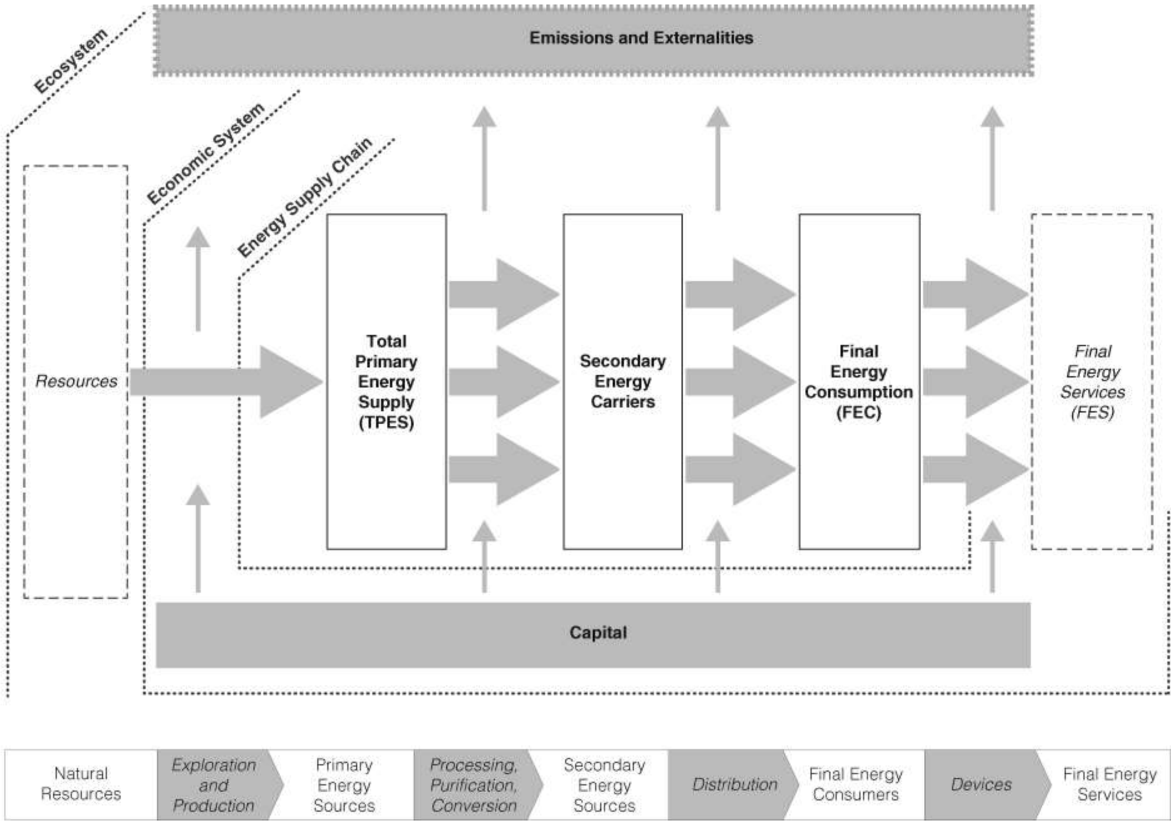


Figure 1.9 Energy Supply Chain Nested within the Economic System and Ecosystem
Source: Adapted from *Energy Efficiency Indicators: Fundamentals on Statistics* (Paris: OECD/IEA, 2014).

Of the three systems making up the energy system, only the ecosystem could potentially be considered a closed system with no new inputs or influences. There are some reasonably hard planetary boundaries when thinking about the entire endowments available from the earth and all of its natural resources and sinks. But even this characterization fails to include the substantial solar resource input that falls onto the earth every day or future space exploration, for example.

Thankfully, not every analysis must begin with the hard boundaries established by closed systems. Examining open systems can also be very useful because they allow examining pieces of a larger and comprehensive closed system. Sometimes these can be thought of as **subsystems**, or smaller open systems that combine into a larger one, and restricting analysis to these subsystems can be useful for granular examination. In doing so, great care must be taken in appropriately dissembling the system so that information is not lost and confusion is not created in the observed subsystem. Finding natural breaks in the system where they are relatively discrete with clear linkages is vital. Also, finding the locations where limited numbers of inputs or outputs occur can simplify the complications when examining subsystems.

Within the energy supply chain, there are three major subsystems, and these three will make up the major subcomponents that this book explores in the following sections.

- **Electricity system**—The modern electricity delivery architecture
- **Transportation system**—Moving people and goods by land, sea, and air
- **Thermal system**—Using heat for everything else, such as cooking, buildings, and industrial applications

Even these subsystems could be further subdivided. For example, the transportation system could be subdivided into the following categories and their components:

- *Geographic subsystems*—Continental, national, urban, island
- *Supply chain subsystems*—Upstream, downstream, refined products, distribution
- *Modalities*—People vs. goods, land vs. sea vs. air, vehicle types (cars vs. buses vs. trains), fuel types (diesel vs. gasoline, natural gas vs. electric)

Such subdivisions allow interesting texture to emerge from systems analysis. In the transportation example above, thinking about land-based infrastructure suggests that appropriate subsystems would probably be based on the limits of travel across a single continental or geographical area, which usually remains relatively distinct (and thus closed off) from land transportation and infrastructure in other continental geographies. Conversely, oil supply chains exhibit a great deal of portability or openness around the world due to the robust oil tanker transportation infrastructure and high value-to-weight ratio for oil products. Therefore, the appropriate range of observation for oil markets would probably be global.

As mentioned, this book examines the three major energy subsystems of electricity, transportation, and thermal systems separately in the following sections. Today, these subsystems remain relatively isolated from each other from an economic and physical infrastructure point of view. But because they are not completely independent either, they must be integrated into the overall energy system, particularly when aggregating their impact on either the economy or the ecosystem in which they are situated. Both subsystem and aggregate system analysis provide necessary insight into the overall energy system dynamics.

d) Supply Chains: A Special Type of System

As the energy supply chain will be at the heart of this book's exploration of the energy system, it is useful to understand what a supply chain is and how it behaves as a system. Systems can behave in a variety of ways, but one useful distinction is whether they are **circular systems** or **directional systems**. Circular systems, as the macroeconomy is often modeled (see Chapter 19 for a description and diagram of this), have many interrelated elements that can exhibit a balance with feedback keeping the various elements in check. It is often difficult to discern the beginning and the end of a circular system process, just like the old chicken and egg problem. In contrast, directional systems tend to have a distinct beginning and a distinct end. They start with some inputs and go through a series of transformations resulting in outputs, but the outputs don't stay in the system or recycle in any significant way.

The energy supply chain, like nearly all supply chains, is an open and directional system. Supply chains are generally constructed to take raw materials and inputs, most often obtained from the environment or ecosystem, and transform them through the application of capital into goods and services of higher value. Ultimately, this results in a product that is of value to an end user, whether for investment or consumption purposes. For example, in an agricultural supply chain, corn is grown and harvested, and then transformed through many processes and distribution channels into food products, and finally delivered to and consumed by a customer.

In the energy system, the supply chain proceeds similarly from raw energy resources to final energy services. This supply chain constantly transforms these energy sources into more refined or more useful forms for energy needs, which are eventually consumed in pursuit of economic activity. In the language of systems, then, the industrial energy supply chain is a subset of the overall energy system, with inputs of both energy resources and capital, and outputs of energy services provided (which are wanted) and waste or emissions (which are not).

Another critical benefit of examining the energy supply chain is that its industrial nature means it is driven by a set of rules, markets, economics, and interactions well developed in other fields of study and analysis. The tools available include the concepts of supply and demand, costs and values, market structures, consumer behavior, and investment, which can be brought to bear on the problem and create a great deal of transparency and granularity in the analysis of system behavior.

e) The Limits of Systems Thinking

Despite its many benefits, systems thinking cannot answer all questions. Its primacy in the energy discourse comes from its ability to integrate different disciplines in the large, important, complex, and global energy system. However, being integrative and complex can sometimes lead to messy or subjective errors. Systems thinking is often conceptual and always an abstraction of reality, as are all models, though it can be rigorously pursued with mathematical modeling or scenario building, if more precision is useful. However, even the best models will rely on simplifications of the underlying relationships within the system.

Systems thinking is most useful as an alternative or complement to ceteris paribus and marginal analysis thinking. It is more useful than marginal analysis to explain the existence of nonlinear relationships and to develop intuition about the complicated and unexplained interactions that persistently appear in the energy system. However, systems thinking is best used as part of a complete set of analytical tools, including marginal analysis, to promote insight and develop fully informed perspectives on the complex human relationship to energy.

E. What Constrains the Energy System?

Studying systems implicitly requires studying the constraints under which those systems operate. Systems tend to create their own constraints by having limiting factors or regulating elements built into their dynamics. They also may be limited by external factors, such as inputs available or outputs required at any time. Revealing and relieving these constraints is a fundamental part of any systems analysis and intervention.

1. Scarcity

Some of the most basic and fundamental constraints in systems arise from what economists term **scarcity**. Scarcity implies that human needs and wants will always be greater than the ability to procure them from the resources at hand. Basically, people constantly suffer from a lack of income or assets to meet their material needs or wants. Individuals want more satisfaction, businesses need more capital, governments want to provide more services for their citizens, but all of them are limited by the endowments available to them.

This scarcity forces people to make allocation decisions with their limited resources. Generally, people choose to allocate their resources first to uses that have the highest relative benefit, and sequentially to those of less relative benefit—that is, the one deemed best first, and the next best next. This benefit determination is more than a fact of its inherent desirability; it must be related to its cost. Whether using a cost-benefit relationship, a return on investment, or the net benefit of benefits minus costs, people will try to procure the most benefit for each unit of the scarce resources they have.

This seems simple from the point of view of a single actor, but it gets much more complicated when trying to imagine that all actors are engaging in the same behavior at the same time. It is precisely this perennial gain-seeking behavior by all stakeholders simultaneously that creates what Adam Smith termed the **invisible hand** in his book *The Wealth of Nations*. The invisible hand describes the optimal societal result of the dynamic where all people are trying to optimize their individual outcomes, even without external governance or intervention.[20] Though the strength of this effect and the definition of “optimal” can be disputed, there is no doubt that society is made up of a bunch of individual actors each trying to better their position, creating a collective allocation of scarce resources.

Sometimes, this invisible hand compels people to share gains that can only be achieved through collaboration, leading to organizational development, resource and risk pooling, and legal infrastructure.[21] For instance, it takes more than one person to construct and operate large capital projects. However, the collective benefits of doing so can be substantial, and the sharing of these gains, also called **non-zero-sum gains**, is a way to improve overall system efficiency, thereby expanding the pool of resources and outcomes for those willing to work together and fairly allocate the additional benefits.

Growth within the economic system emerges in precisely this way. Individuals specializing in some task for which they may be relatively well suited (technically, a **comparative advantage**) create additional productivity that can be shared with others who specialize in different outputs, thereby raising the aggregate pool of outputs available for all. Architects can focus on design, masons can pursue stonework, and accountants can dedicate themselves to accounting—each gaining efficiency from their focus. Again, this situation can only exist where trade in goods and services can occur, people are made better off by it, and the rights and obligations of the various parties can be clearly established and are generally considered enforceable.

Sometimes, though, the simultaneous optimization of people in pursuit of their own gains leads to distributional tensions among stakeholders where one side's gain is another side's loss. A high price might benefit a seller, while a low price may benefit the buyer. Under such a **zero-sum game**, both cannot be made better off at the same time. Understanding how these distributional outcomes resolve themselves will require understanding more about market dynamics and power relationships among stakeholders.

While one particular stakeholder may have tremendous influence and excessive proportions of the available gains, in a system all stakeholders sooner or later have a say in determining outcomes. Eventually, someone constantly paying too much for something will find a cheaper alternative. Forced or undercompensated labor will be met with declining productivity or even active resistance. Everyone who participates in a system has some degree of say over the system outcome and provides a check or constraint on the unregulated desires of other stakeholders.

A simple way to visualize this mutual dynamic of interlocking system participants is in the context of sustainability. While the notion of sustainable development will be revisited in detail in Chapter 20, Figure 1.10 shows the intersection of three distinct classes of stakeholders that might have competing claims on defining outcomes. In this example, the stakeholders could include the forces of social equity, environmental stewardship, and economic growth. Observing tensions among the stakeholders is easy. Environmental stewardship might be perceived as costly, at least in the short term, and accommodating it would reduce economic growth. It is also possible that unrestricted access to environmental and natural resources might be good social policy; however, it might be very bad for long-term environmental health.

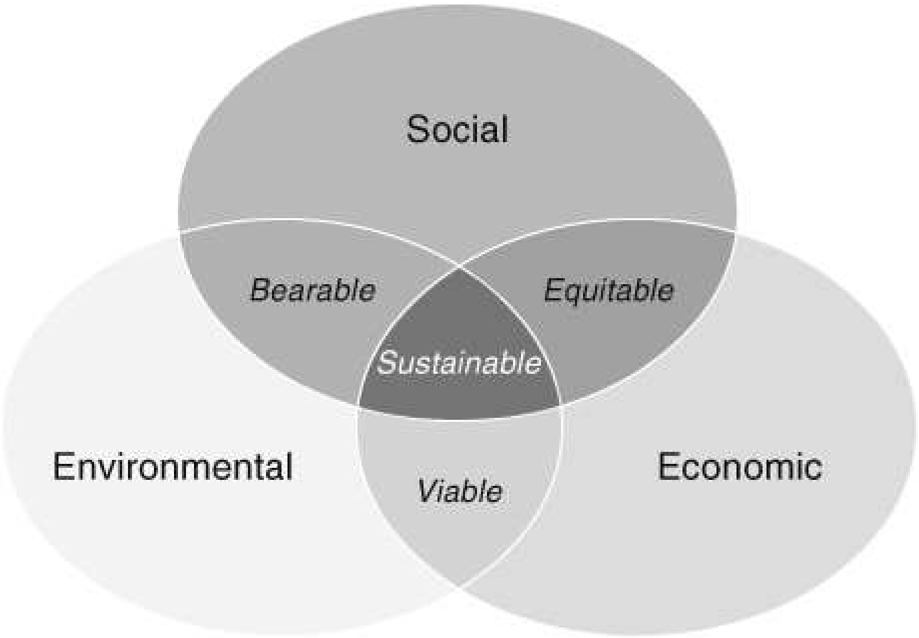


Figure 1.10 Stakeholders in the Sustainability Debate
Source: *Wikimedia Commons*, http://en.wikipedia.org/wiki/File:Sustainable_development.svg.

A study of history shows a relentless series of tensions among these three forces as the various stakeholders bargain within the constraints of a global scarcity of resources to meet perceived needs. Despite the fact that economic growth tends to be considered important, times in which social equity or environmental stewardship were ignored have led to both gradual and sudden shifts of priorities: the French Revolution against social inequities, the antitrust movements of the early twentieth century favoring social welfare over pure economic growth, and the environmental crises in the US in the 1960s and 1970s are examples of when such imbalances have been brought into focus and scarce resources were reallocated in pursuit of improved sustainability.

Since the dawn of human activity, scarcity has provided basic constraints on human behavior and welfare expansion, and insatiable human appetites suggest that it always will. Within a world of perpetual scarcity, stakeholders will continue having to make allocations that are limited by society's overall endowments, even with continued expansion of opportunities and reduction of scarcity through innovation and efficiency. (The Economics Box on constrained optimization provides additional insight on this topic.)

Economics Box: Constrained Optimization

In economics, the concept of **constrained optimization** is used to demonstrate the relationship between **objectives** (representing a mix of goals or desires) and **constraints** (representing limited resources). Often, a person has the objective to maximize or minimize some outcome (conceptually, find the best solution or get the best or most that they can in a situation) but must always do so within the limits imposed by their basic resource constraints. As a tool, constrained optimization can also be treated both conceptually and quite rigorously with linear programming, modeling, or even theoretical mathematics. These approaches accommodate substantially more variables and many more complex constraints and objectives.

One famous example of constrained optimization for a nation allocating resources is represented by the guns-vs.-butter tradeoff, shown in Figure 1.11. This optimization posits a situation in which a society's economy can produce only two goods, guns for war or butter for consumption, both of which are desirable, and it has a fixed amount of resources to divide between their production. The combination of the two goods that can be produced is represented by the **production possibility frontier**. How much of each good is desired is defined by a set of parallel **indifference curves**, showing the tradeoff in desirability of these two products. The goal in this case is to maximize—that is, achieve—the highest possible indifference curve allowed by the constraints imposed by the production possibility frontier.

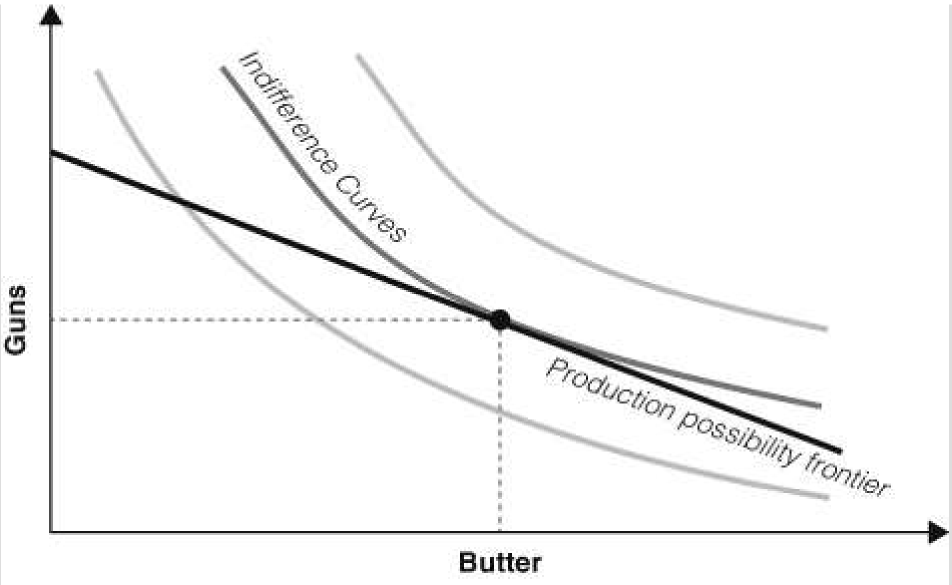


Figure 1.11 Constrained Optimization for Two Goods with Indifference Curves
Source: *Wikimedia Commons*, https://commons.wikimedia.org/wiki/File:Indifference_curves_showing_budget_line.svg.

When applied to individual behavior in economics, the assumed goal of individuals is generally thought of as one of maximizing utility in a world of scarcity. **Utility** in this context means a level of happiness or satisfaction, and scarcity suggests that it must be maximized subject to some budgetary constraint. Though easy to conceptualize, the notions of utility and preferences can be difficult to measure, and great economics debates are fought over the type, persistence, and amount of utility obtained by individual actors from their choices. Once objectives are established, ensuring efficiency in resource allocation is important to achieve maximum utility within a given set of constraints. Not doing so would cause an unnecessary loss of utility; that is, it would result in an inefficient allocation of resources.

In all the constraints described in this section, their presence will (by definition) limit the potential outcomes of the system and will direct the growth and dynamics of system behavior. Hard constraints of the various types will limit progress in that direction and signal investments and resource allocations to less constrained opportunities and avenues for progress. In the context of sustainability, the simultaneous constraints imposed by the various forces must be considered in forecasting and intervening in the system outcomes.

2. Input (Supply) Constraints

In the energy system, scarcity appears in many ways. One way is as a limit on primary inputs or supplies of resources, such as a lack of basic energy inputs to meet all human needs and wants. These concerns about energy supply constraints dominate the energy literature and invoke concepts like peak oil and resource depletion, which will be explored in detail throughout this book. One reason that supply constraints dominate the literature has much to do with the original problem that the energy system has been trying to solve, namely getting more energy. Another reason is that the energy literature has historically been dominated by the technical disciplines required to obtain, transform, and move these energy resources. The scientists and engineers who have historically written the most about the energy system are specifically trained to think about the physical characteristics of obtaining the necessary supply, and the limits they face in doing so.

In addition to the basic scarcity of acquiring physical volumes of energy (which, as will be seen, is both a physical and an economic limit), each of the energy supply options available must enter the energy system from the ecosystem, and each of these primary resources has additional limits constraining the ability to procure more. John Holdren has provided a useful characterization of the specific cause of resource scarcity for each of the primary resources as follows:[22]

- **Conventional oil and gas**—Not enough physically and economically available resources
- **Coal, tar sands, and oil shale**—Not enough atmospheric sinks for the carbon implications of using these sources
- **Biofuels**—Not enough land to grow meaningful energy volumes
- **Wind and hydropower**—Not enough acceptable sites
- **Solar photovoltaics**—Too expensive

- **Ocean energy**—Too much capital and too ecologically disruptive
- **Nuclear fission**—Too risky and unforgiving
- **Nuclear fusion**—Too technically difficult
- **End-use efficiency**—Not easily understood or adopted by users

Though they vary in character, each of these constraints could be relieved under the right circumstances or with application of sufficient invention and investment. However, such investments require resources, and until resources are obtained and constraints reduced, the scarcity will persist. Some of these constraints are physical, some are technical, some are economic, and some are behavioral. Further work is required to determine which might be temporary or permanent constraints, or which might be cheaper or more expensive to relieve. The elements constraining the ability to acquire sufficient energy supplies can come from any dimension, which underscores the need for an interdisciplinary approach to describe the energy system dynamics.

3. Conversion, or Capital and Infrastructure, Constraints

Beyond energy itself, a nuanced understanding of supply constraints should include the limits imposed by the infrastructure available to procure and transform it through the supply chain. When the energy supply comes into the energy system as a primary source, it must generally be transformed (often many times) before becoming useful for delivering energy services. Infrastructure collectively defines the mechanics of making these transformations of location and form and function through the supply chain.

Infrastructure is not a constant element, and it can change in many ways. If a lack of energy infrastructure to transform energy constrains behavior, it is often possible to simply invest more capital and create additional infrastructure. The decision to do so may be constrained by access to capital or expected return on investment, but relieving the constraint by adding infrastructure usually remains a possibility.

Conversely, energy infrastructure that currently exists is subject to deterioration and diminishment of its productive capacity.

Eventually, capital reaches the end of its useful life or is no longer useful for its original purpose and must be taken out of service. This depletion of productive capacity can be a constraint on behavior in the energy system and must be considered, certainly until additional capital investment can be made to compensate for it.

4. Output (Demand) Constraints

Because the energy supply chain proceeds in a directional fashion toward the desire for energy services, constraints on the behavior of the system may come from the ability or willingness of consumers to pay for the energy services that they desire—that is, demand scarcity. The infrastructure may exist, the resources may be available, but the consumers simply fail to demand more energy services.

This suggests that the relationship of the energy system to the macroeconomy, income, and consumer behavior is very important for determining the energy system's performance. It also suggests that the conditions affecting the energy system are dependent on fluctuations in the macroeconomy and end-user preferences that can occur over short and long time cycles. A favorable economy and increasing budgets grow the demand for energy services (reduce demand constraints), while recessions and disruptive events can reduce demand (increase these constraints).

Strategies aimed at increasing the value of the energy services can help to relieve these constraints, as can reducing the costs the customer has to pay. Both can improve the cost-benefit ratio to buyers. Conversely, the value of any particular supply of energy can be reduced by substitution of cheaper energy supply options to meet the same energy service. Such a demand constraint creates ripple effects and has repercussions throughout the supply chain.

5. The Fundamental Tension between Innovation and Depletion

As a whole, the energy system is seen as a complex system of relationships, each of which is subject to some constraints. In aggregate, the system appears complex but balanced by tension between the forces of supply and demand, which proceed as if by Adam Smith's invisible hand to direct primary energy inputs to become energy service outputs.

Within the system, many incentives and opportunities exist to procure more energy inputs and to use them more efficiently to create outputs. Constraints compel invention and creativity in trying to create additional advantages in the form of reduced costs or increased profits. This **innovation** occurs everywhere in the system—supply, efficiency, demand, cost, and benefit—and is a permanent fixture of the energy system, which is driven by the very basics of ingenuity and greed.

At the same time, people tend to procure the cheapest and easiest resources first, leaving the more expensive ones for later. The infrastructure investments made to transform energy through the system begin to deteriorate as soon as they are installed. Competitors are constantly trying to take away market share, which keeps prices in check. These are just some of the manifestations of **depletion** (of the relevant resources or capacity or value or market opportunity) and result from normal economic behavior of people and firms minimizing costs or pursuing the easiest alternatives first. But capturing the very best opportunities first uses them up, leaving only more expensive resources or less ideal opportunities available for the next time. When those opportunities are then captured, even less optimal ones are available for capture, and so on. This is depletion.

One interpretation of the natural balance in the energy system suggests that this entire system remains together by a fundamental tension between these forces of innovation and depletion. On the one hand, cleverness, entrepreneurship, and ingenuity are creating opportunities and driving efficiency in resource use, a process that has resulted in hundreds of years of global economic growth through industrialization. On the other hand, depletion is constantly making things more difficult and costly, ultimately threatening a collapse of wealth and welfare if the resource base is damaged or exhausted before innovation leads to an alternative, effective path. The very notion of sustainability (explored in Chapter 20) tries to reconcile these competing forces of innovation and depletion.

Understanding how this tension resolves itself (i.e., how innovation and depletion will interact and what the outcomes are) is the very root of energy system analysis. Thinking of the energy system and all of the mechanisms that define it—bounded by the context of the macroeconomy and natural resource systems in which it operates—explains where the system is headed. It will also provide the tools and the leverage to change that outcome, should the current trajectory prove untenable.

Key Terms

- energy services
- gross domestic product (GDP)
- ceteris paribus
- population momentum effect
- energy intensity (E/GDP)
- energy productivity (GDP/E)
- IPAT framework
- memes
- positive analysis
- normative analysis
- energy system
- marginal analysis
- models
- leverage points
- nonlinearities
- root causes
- open vs. closed systems
- nested systems
- circular vs. directional systems
- scarcity
- invisible hand
- constrained optimization
- utility

innovation
depletion

Appendix 1: Compound Growth

Many of the calculations conducted throughout this book will require understanding how growth rates change the volume or amount of things over time. The easiest way to conceptualize the impact of growth rates is to understand how the value of anything today (**present value**) increases by a certain periodic growth rate (**compound growth rate**, or ***r***) over a number of periods (**time**, or ***t***), to determine its value at the end of those periods (**future value**).

Depending on the question being asked, it is also possible to calculate the imputed annual growth rate by knowing the present value and future value and applying the **compound annual growth rate** (**CAGR**) formula. It is simply a rearrangement of the future value formula to isolate the imputed interest rate. (See Metrics Sidebar below.) This creates a metric that is suitable for comparing relative growth rates across similar types of growth and similar periods. Calculating CAGR can get more complicated with more inflows and outflows, and advanced methods of calculating CAGR can be applied using advanced modeling tools such as Microsoft Excel formulas. Investment evaluation tools such as internal rate of return (IRR) and net present value (NPV) used throughout this book will rely on these basic compound growth formulas.