4 Life-Cycle Properties of Engineering Systems: The Ilities

In the *epoch of great inventions and artifacts*, the implicit mandate of the engineer and inventor was to "design for first use." The aim was to design and build an artifact that would "work" and fulfill its primary function when first turned on or started up. If it did not, it was back to the drawing board. Immediate functionality was the main focus. Little or no attention was paid to side effects or other more subtle behaviors, especially those that might be far in the future.

In the *epoch of engineering systems*, the focus has changed. As we discussed in chapter 3, their evolution over long lifetimes is a significant aspect of large-scale complex systems. Understanding and working with engineering systems requires attention to properties that have long time exposure. Attention to side effects and the context that establishes ground rules and constraints within which systems operate is crucial, as these factors are part of the systems' very essence.

The Importance of Not Simply "Working"

The first automobiles were largely motorized versions of the horse-drawn carriages that preceded them. But as the artifact improved and began to work in more demanding operating environments—at higher speeds, at night, in adverse weather—new subfunctions, beyond the primary function of the car, became important. Over time, inventors responded by adding windshields to cars to protect the eyes and mouths of drivers from bugs, windshield wipers to ensure visibility in the rain, and headlights so drivers could see in the dark. Lots of other improvements were made over the years, perhaps more than most contemporary drivers know.

It wasn't long before it was became to address some side effects of driving automobiles. For example, the first cars were equipped with

brakes, but only on the rear wheels. Drivers of the time would swerve and skid when they applied the brakes, and stopping required a lot of distance.

In 1923, the relatively high-priced Buick appeared with brakes on all four wheels; these four-wheel brakes were invented by Charles F. Kettering (who was responsible for a lot of inventions that really changed the way people lived, including safety glass, the automatic transmission, incubators for premature infants—in fact, a list too long to include here).

By the time Henry Ford's Model A came out in 1927, four-wheel brakes were standard, and have remained standard ever since. Further improvements came in the 1930s, when hydraulic four-wheel brakes came into use, allowing for higher brake pressures and shorter stopping distances. Later, Europeans pioneered dual hydraulic brakes to address the problem of the original single hydraulics—namely, that a loss of hydraulics meant a loss of all braking ability. Power brakes that increase the amount of hydraulic pressure debuted in the 1950s. By 1961, rather rapidly, dual hydraulics became standard in U.S.-made cars, thanks to a competitive thrust by American Motors Corporation.

The early development of automotive braking and many early developments in airplanes are tales of *safety* becoming a consideration as the artifacts move beyond their first use—that is, the emergence of *ilities*. The ilities are central to any discussion of engineering systems, and require a very precise definition:

> The *ilities* are desired properties of systems, such as flexibility or maintainability (usually but not always ending in "ility"), that often manifest themselves after a system has been put to its initial use. These properties are not the primary functional requirements of a system's performance, but typically concern wider system impacts with respect to time and stakeholders than are embodied in those primary functional requirements. The ilities do not include factors that are always present, including size and weight (even if these are described using a word that ends in "ility").[1]

Over time, greater awareness of safety became characteristic of the *epoch of great inventions and artifacts*, although engineers concentrated primarily on making safety-related alterations and adjustments to artifacts (often products), they also participated in changing the underlying systems and operating environments within which they function. *Quality* was the other ility to emerge in this early epoch.

As background research for this chapter, we compiled a list of 20 ilities that we have frequently encountered in our work on engineering systems. For each of them, we collected data that would allow us to rank these life-cycle properties based on how frequently they are mentioned in the scientific literature and on the Internet.[2] Figure 4.1 shows the result of our analysis. The black vertical bars indicate the number of scientific papers (in thousands) that mention a particular ility in their title or abstract. The gray vertical bars show the number of Google hits (in millions) obtained for each ility.[3]

The results from the scientific database and the number of Internet hits are strikingly similar, with the notable exception of *sustainability*, which we discuss later in this chapter. The top four ilities are, in order, *quality*, *reliability, safety,* and *flexibility*.

Quality and safety are so important in part because they have received much attention since the beginning of the *epoch of great inventions and artifacts*. Note that figure 4.1 shows some ilities that are strongly related to quality as being of high importance (e.g., reliability, robustness,

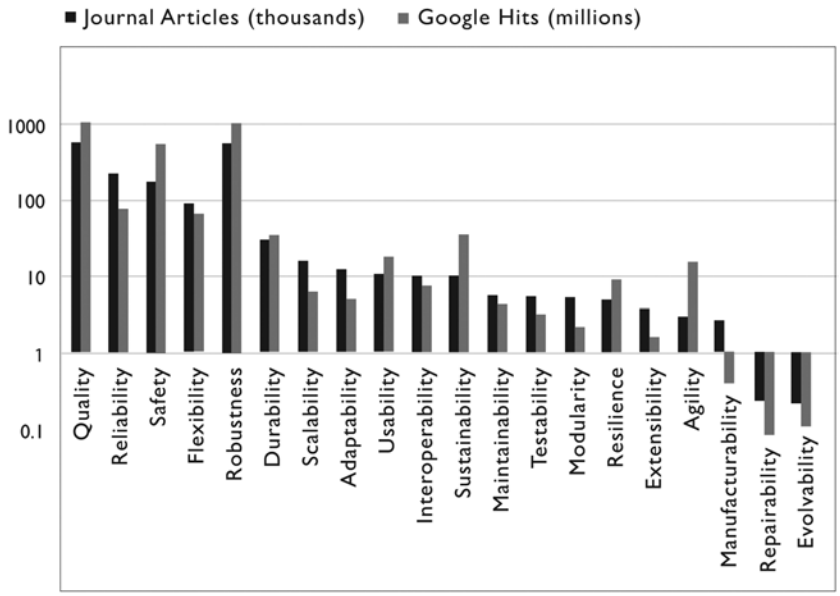

Figure 4.1
Ranking of the ilities in terms of frequency of occurrence: the black bars indicate scientific journal articles published from 1884–2010, in thousands (source: Compendex and Inspec databases); the gray bars indicate number of hits on the Internet, in millions (source: Google).

durability). We will consider such relationships a little later in the chapter.

As we entered the *epoch of complex systems*, usability—which, of course, had always been a significant concern of inventors and engineers—emerged as a specific ility, largely from how users (humans) perceived quality as well as from unanticipated difficulties in operating complex systems. Engineers also began to worry—to a greater or lesser degree—about the *maintainability* of the artifact(s) and, sometimes, the systems within which the artifact(s) function. This was driven in part by the growing realization that perfect *reliability* and *durability* were impossible to achieve and hence an unrealistic expectation, leading to focus on both preventive and corrective types of maintenance.

We think of these four aspects of artifacts and systems—safety, quality, usability, and reliability—as the classical ilities of engineering. In our present *epoch of engineering systems,* the list has grown much longer. This can be attributed partly to the fact that more attention to ilities led to more complex systems, and vice versa. More ilities emerged because growing complexity and scale of deployment led to more and more important side effects; the rapidly increasing rate of change in systems and concomitant social changes also spurred this expansion of the ilities (as we discussed in chapter 1). No one wanted to pay for things that did not contribute directly to the primary functionality of the artifact, but over time it became untenable to run systems without paying attention to characteristics—even if it sometimes took decades of use to realize this. Today, there is an increasing realization that much of the value that engineering systems generate depends on the degree to which they possess certain life-cycle properties, or ilities.[4]

The cumulative number of scientific articles published in the engineering literature on our set of 20 ilities from 1884 (the earliest date for which such data was available) to 2010 illustrates this point. Figure 4.2 shows only the top 15, to demonstrate more clearly the time dependence.

Indeed, quality and safety were given consideration early on, first in the building of national infrastructure such as railroads and bridges and later in the twentieth century when various electromechanical products became available to a wider population. Over time, during the *epoch of complex* systems and then in our current *epoch of engineering systems*, new ilities became the subject of intense interest and scientific research.

Let's look at some examples of the ilities in greater detail, more or less in the order in which they emerged.

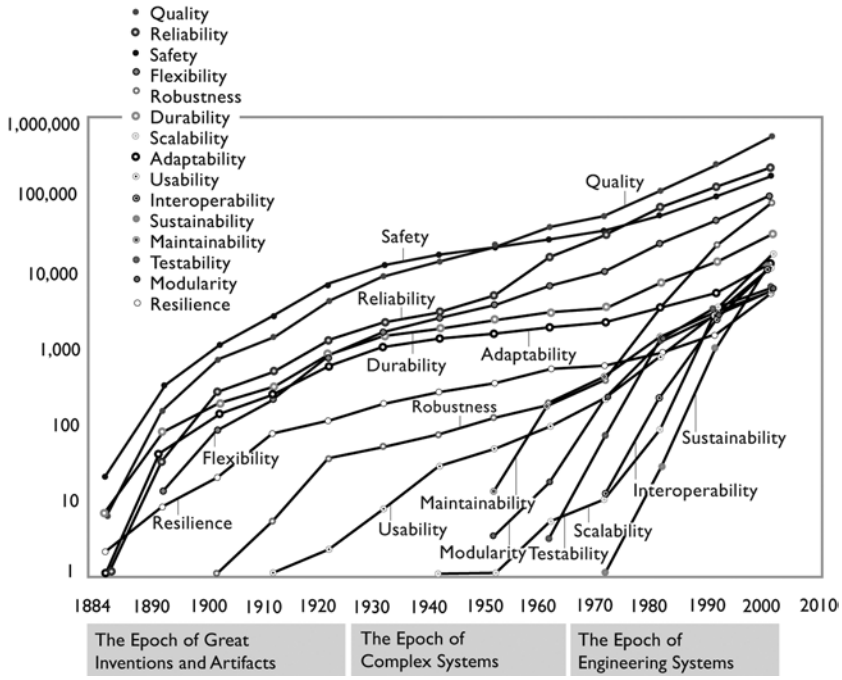

Figure 4.2
Cumulative number of journal articles in which an ility appears in the title or abstract of the paper (1884–2010). Source: Inspec and Compendex, accessed via Engineering Village (8 August 2010).

Quality

The first "ility" of traditional engineering to be discussed at length is *quality*. An extensive literature on quality exists, defining this ility from multiple perspectives. One conceptual framework categorizes quality as transcendent (some abstract philosophical, perceptual, moral, or religious entity), product based (fit for use, performance, safety, and dependability), user based (able to satisfy human needs), manufacturing based (conforming to engineering and design specifications), or value based (difference between conforming to specifications and monetary cost).[4] The latter category has a lot to do with perception; put simply, something that exhibits a high level of conformance and relatively low cost would be "high value."[5]

In our engineering systems context, quality means that the artifact or system is well made to achieve its function. In this respect, opening and

closing without squeaking is a sign of quality in a door. In strictly engineering terms, such quality is often a direct result of "tolerance"—the permissible limits of variation in a physical dimension or some measured value or physical property of an artifact or, for that matter, anything in a system. The story of Henry Martyn Leland illustrates how quality in engineering grew to become an important ility.

Leland was a machinist who made tools and micrometers using extremely tight tolerances of fractions of an inch. Settling in Detroit, he achieved tolerances as tight as 1/2,000th of an inch (astonishing for those days) and was recruited directly into automobile manufacturing. Later, he became the founding president of Cadillac Motor Company, which by 1905 was one of the world's leading automakers.

Cadillac automobiles were known, as one writer of the time put it, for being "free of temperament" because of their high levels of workmanship (or craftsmanship) and reliability (an ility that "supports" quality, as we will see later). Perfectionism in the pursuit of quality was the touchstone of Leland's approach, and given that Cadillac used standardized, machine-produced parts, the achievement was remarkable. Most automakers of the time bought into the notion that *hand-made* parts were more refined and precise.

Ideally, an artifact or system should work all the time and in the way intended, but that was usually an unrealistic expectation. Leland's story is about quality being associated with tolerances and translating into reliability and an assurance that the artifact is well made. Quality became important because its absence creates more side effects and exacerbates problems related to other ilities such as maintainability and reliability. The focus on maintainability and reliability gave rise to the need for service organizations. Car dealerships were never only in the business of selling cars, but offered an important service of making needed repairs to those cars. This may be a significant antecedent of the modern service economy.

No discussion of quality would be complete without a mention of W. Edwards Deming, an American statistician who is often referred to as the father of quality management. In the period after World War II, Deming worked as a census consultant to the Japanese government under General Douglas MacArthur. It was in Japan that he began to teach business leaders statistical process control methods. The rest is history, and Deming is thought to have had more impact on Japanese manufacturing and business than any other non-Japanese person.

Asked by the Japanese Union of Scientists and Engineers to teach statistical process control and concepts of quality, Deming gave a series

of eight lectures in the summer of 1950 in which he convinced top Japanese managers that improving quality would reduce their expenses while increasing their productivity and market share.[7] This flew in the face of the long-held conventional wisdom, which was that there was an inverse relationship between quality and productivity, and that improvements in the former would always lead to a decrease in the latter. Japanese manufacturers embraced Deming's ideas, much more so than those in the United States (that was to begin only two decades later), and the wide application of his techniques led to unprecedented quality and productivity levels, which lowered costs and boosted global demand for Japanese products.

Over time, the understanding of *quality* evolved to the point where quality became something engineers sought to achieve from the very beginning of the design process rather than at the end of the manufacturing process, by filtering out parts that did not meet some required tolerance threshold. In the *epoch of complex systems,* the objective of achieving "perfect first-time quality" was one of the prime motivators behind the Toyota Production System we mentioned in chapter 1 and cover more deeply in chapter 6. With a high level of quality from the outset, the artifact or system is far more likely to last a long time, thus giving it *durability* and requiring comparatively less preventive maintenance and repair, and hence generating fewer of the side effects (operating costs, etc.) associated with these problems. The longstanding reputation of Toyota cars as highly being reliable—the company's significant problems of early 2010 notwithstanding—speak to this very point.

Toyota was an early adopter of the ideas of Genichi Taguchi, a Japanese engineer and statistician who, beginning in the 1950s, developed a method for improving the quality of manufactured goods through the application of statistics. His work expanded Deming's ideas while also introducing some new concepts.

In Taguchi's philosophy of quality, design is used to obtain the minimum deviation from what customers desire from the outset. The goal of design and manufacture is then to minimize the "Taguchi loss function," which captures how far an artifact is from the ideal or desired target state. In addition, the design is optimized so that any unachievable or overly expensive tolerance does not affect the customer or overall quality goal by making the system's behavior relatively insensitive to such difficult-to-control parameters. Thus, the artifact is designed to be immune to uncontrollable environmental factors and internal variables that are difficult to control. This is the key concept of *robustness*, an ility whose

importance and emergence are captured in figures 4.1 and 4.2. Taguchi also emphasized that the cost of quality should be measured as a function of this deviation, and that it should be measured systemwide.[8]

One thing about quality that should be mentioned is that it tends to be relative. The Toyota story makes this point very clearly. Prior to the widespread availability of Toyotas in the United States and Europe, U.S. and European automakers performed at essentially the same level with respect to quality within their peer groups. That meant that for American and European consumers, the relative quality they perceived in cars was defined by their available suppliers. This changed rather suddenly with the introduction of competition from Japan.

Most of the preceding examples have been about manufacturing-based quality, but what about user-based quality? The early history of the telephone also tells us a lot about how quality was viewed from the user perspective and how it has evolved as an ility.

In 1910, some 10 million telephones were in use around the world; 7 million were in the United States, and 5 million of those were part of what came to be known as the Bell System or AT&T.[9] The earliest telephone systems had a limit of about 20 miles, but in 1910, voice could be transmitted from Boston to Denver, and the expectation of coast-to-coast transmission soon was high. It took an operator about a minute to find another user.

For the earliest telephone users, the amount of time it took to place a call was likely secondary to the primary quality issue: the sound, and hence the understandability, of the voice. The first transmitters had problems, and horrible noise accompanied speakers' voices with the early grounded wire system. The invention and use of "doubled wire" did a lot to eliminate ground and induction effects. Over time, more improvements in cables—mostly changes in insulation—were also critical to improving voice quality (and also to reducing costs).

Among many important innovations, the Pupin Coil stands out in the history of the telephone. In electronics, loading coils are used to increase a circuit's inductance. Oliver Heaviside, a self-taught English physicist, mathematician, and electrical engineer, had theorized in 1881 about transmission lines in studying the slow speed of the trans-Atlantic telegraph cable.[10] Representing the line as a network of infinitesimally small circuit elements, Heaviside concluded that distortion of the signal transmitted on the line could be mitigated by adding inductance to prevent amplitude decay and time delay. The mathematical condition for distortionless transmission came to be known as the Heaviside condition.

In 1894, Mihajlo (Michael) Pupin, a Serbian immigrant to the United States, patented a type of coil that "loads" the line with capacitors rather than inductors—an approach that was largely dismissed by others. AT&T fought a patent battle with Pupin. The short version of the story is that Pupin's approach ultimately prevailed. It greatly reduced the amount of copper required (at that time, expensive copper could account for half of the capital investment required to set up long-distance telephone lines) and made longer distances feasible.

Improved versions of the Pupin Coil developed by AT&T were called repeaters. Basically, a repeater is a device that amplifies the signal so it can be "regenerated" and passed along without diminishing its quality. This is, of course, the realm of analog signal processing, well before information theory and digital communications were invented.[11] Research began in 1912, and by 1915 success had been achieved to the point where long-distance phones were working between New York and San Francisco.

Telephone quality in the early days revolved around very important concepts that would emerge much more strongly in the *epoch of complex systems* and even more so in the *epoch of engineering systems*: human factors and ergonomics.[12] These are discussed later in this chapter.

Safety

The story of the automated traffic signal illustrates both *safety*, discussed here, and *maintainability* (detailed later). Police officers had long been stationed at busy intersections to direct traffic in cities, even before the introduction of the automobile. In 1868 in London, a revolving gas-illuminated lantern with red and green lights—indicating "stop" and "caution," respectively—was installed at one busy intersection; it was turned by a policeman who operated a lever at its base to have the appropriate light facing traffic. Then the light exploded on January 2, 1869, injuring the policeman. As a result, the policeman was replaced by an automated traffic light. At least that's the story that has been handed down for nearly a century and a half; it may be an early example of an urban legend.

Decades later, as automobiles increased the problem of safe traffic flow and unanticipated traffic jams became more commonplace (as described in chapter 1), inventors and engineers rose to the challenge of how to ensure greater safety. Advances in traffic signaling came fast and furiously. A policeman in Salt Lake City, Lester Wire, invented a

red-green electric traffic light in 1912. Cleveland, Ohio, saw the installation of James Hoge's two-color traffic signal in 1914, with a buzzer to warn of color changes and the ability of police and fire stations to control the signals in the case of an emergency. Back in Salt Lake City, in 1917, six intersections were linked so that their signals could be manually controlled simultaneously. Some of these milestones are better documented than are others, but one thing is certain: A lot of engineers were busy at work on this aspect of safety.

Our modern-day, three-color, four-way traffic light was the idea of William L. Potts, a police officer in Detroit, Michigan. Inspired by railroad signals, but adapted to the right-angle nature of street traffic, Potts used red, amber, and green. His light was installed in 1920 at the corner of Woodward and Michigan Avenues, and within the next year some 14 others were installed throughout the city.[13]

Automatic control came first to Houston, Texas, in March 1922, at least according to the Federal Highway Administration.[14] However, this is disputed by the story of Garrett A. Morgan, an inventor in Cleveland who laid claim to inventing the "electric automatic traffic light." While others had obtained U.S. patents for traffic signals, it was Morgan's patent that the General Electric Corporation purchased for $40,000. GE began manufacturing the signals, and soon had a monopoly in the United States.

The early successes of the traffic light and other features to improve safety cemented in our minds the idea that safety can always be improved by simply integrating a clever device into an existing system. Barriers at rail crossings, the grounding pin in an electrical connector, or—dare we mention it—a blowout preventer atop a deepwater oil well are all examples. However, "systems safety"—as with all the other ilities—requires a much deeper understanding of dysfunctional or detrimental interactions among technical components, user behaviors, and the operating environment.

Notably, some ilities tend to be very important in some kinds of systems but relatively unimportant in others. As the preceding examples suggest, safety matters a lot in transportation-related systems. It doesn't come up nearly as often when we look at communication systems.

Usability/Operability

The achievements of the *epoch of great artifacts and inventions* were great only because they could be put to use, whether it was the automobile or the telephone or any other artifact. For instance, telephone quality

in the early days (and still today) revolved around a very important concept that would emerge much more strongly in the *epoch of complex systems* and even more so in the *epoch of engineering systems*: human factors. Put simply, human factors are the properties of human capability and the cognitive needs and limitations of humans—in our telephone story, the capability of humans to hear and *understand* the voice at the other end of the line. Ergonomics, which came to prominence later, tends to be concerned primarily with biomechanical usability—as in the case of a computer keyboard that works well with the human hand, or a control screen in a nuclear reactor that works well with human visual and cognitive processes.

In the *epoch of complex systems*, this understanding of human factors emerged as *usability* or *operability*, one of the classic, or traditional, ilities of engineering. Although sometimes considered synonymous, they are slightly different. Usability most closely corresponds to human factors and ergonomics-type issues, whereas operability more clearly denotes institutional concerns beyond single humans. As we saw in figure 4.2, research into usability began as early as 1910 but really took off during and after World War II, when the ability to operate military equipment efficiently became a matter of survival.

Amateur radio operators took human factors into consideration very early, linking quality and usability as they expanded a system that had been developed by the British government around 1909 to facilitate communication between ships and coast stations using Morse Code. The "Q codes" embodied a list of abbreviations that could be used to ask simple questions. "QSL" was used to ask and answer whether a transmission was received. As early as 1916, ham radio operators began to send postcards that would verify receipt of a station. A standardized QSL card—with callsign, frequency, date, and other information—emerged a few years later (figure 4.3).

Over time, the QSL employed human judgment to help radio operators—from small amateur stations to huge, government-funded shortwave megastations—determine the quality of their signals. Someone in the United States receiving a shortwave broadcast from, say, Radio Nederlands in Hilversum, Holland, might send a QSL in the form of a letter, requesting a QSL card (which became a collectible) in return for grading the station's "SINPO"—an acronym for signal strength, interference, noise, propagation (whether the signal is steady or fades from time to time), and overall—on a 1-to-5 quality scale. By indicating the date of the transmission and a brief description of what was heard

Figure 4.3
QSL card from Radio Nederlands, Hilversum (1969).

(e.g., "a concert by a Dutch folk ensemble"), the requester helped the station gauge its quality. After all, what better way to determine whether humans, for whom the broadcast is intended, are receiving a quality broadcast (even if the SINPO ratings are subjective)? Like the telephone, radio couldn't just *work*; it had to work well enough for its users.

The standardization of electric power is also a story of how usability/operability came to be an important ility. Thomas Edison established his electricity generating station on Pearl Street in New York City, which opened for business in 1882, featuring what have been called "the four key elements of a modern electric utility system: reliable central generation, efficient distribution, a successful end use—in 1882, the light bulb—and a competitive price."[15] As demand for electricity grew, though, electricity was provided to end users primarily through small central stations, often many in one city, each limited to supplying electricity for a few city blocks. These stations were owned by any number of competing power companies, and it wasn't unusual for people in the same apartment building to get their electricity from completely separate providers. This competition, in a natural monopoly like electric power, did not drive down prices because an operating problem remained: The generating capacity was very much underused and thus the investment cost to serve outlying regions was much larger than users desired.

Not only was there competition for customers, though—there was also technological competition for which type of electricity would be used: alternating current (AC) or direct current (DC). In fact, historians have dubbed what unfolded in the late 1880s the "War of the Currents." It was most definitely a war over usability and operability, as this technological choice profoundly affected both customers and producers.

Thomas Edison and George Westinghouse were the major adversaries. Edison promoted DC for electric power distribution, and Westinghouse (and his ally, Nikola Tesla), were the AC proponents. Edison's Pearl Street Station was a DC-generating plant, and there was no reliable AC-generating system until Tesla devised one and partnered with Westinghouse to commercialize it. Meanwhile, Edison went on the warpath, mounting a massive public campaign against AC that included spreading disinformation about fatal accidents linked to AC, speaking out in public hearings, and even having his technicians preside over several deliberate killings of stray cats and dogs with AC electricity to "demonstrate" the alleged danger. When the first electric chair was constructed for the state of New York to run on AC power, Edison tried to popularize the term "Westinghoused" for being electrocuted.

Technologically, direct current had all sorts of system limitations related to usability and operability. One was that DC power could not be transmitted very far (hence the many stations and their limited service areas in cities), so Edison's solution was to generate power close to where it is consumed—a significant usability problem as rural residents desired electrification. Another limitation of DC is that it could not easily be changed to lower or higher voltage, requiring installation of separate lines to supply electricity to anything that used different voltages. Lots of extra wires were ugly, expensive, and hazardous. Even when Edison devised an innovation that used a three-wire distribution system at +110, 0, and –110 volts relative potential, the voltage drops from the resistance of system conductors was so bad that generating plants had to be no more than a mile away from the end user (called the "load").

Alternating current, though, used transformers between the relatively high-voltage distribution system and the customer loads. This allowed much larger transmission distances, which meant an AC-based system required fewer generating plants to serve the load in a given area, and hence these plants could be larger and more efficient due to the economies of scale they could achieve. Westinghouse and Tesla set out to prove the superiority of their AC system. They were awarded a contract to harness Niagara Falls for generating electricity, and began work in 1893 to produce power that could be transmitted as AC all the way to Buffalo—a distance of about 25 miles. In mid-November 1896, they succeeded, and it wasn't long before AC replaced DC for central station power generation and power distribution across the United States. The roots of the architecture and structure of our current centralized electrical power system can thus be traced back to a fierce battle of technologies and personalities more than a century ago.[16]

Of course, this left a lot of DC systems still in place. Some cities kept small DC networks running long after AC had essentially won the war; notably, Boston was still using 110-volt DC in a small area near Boston University into the 1960s, and there were always stories of BU students who had destroyed their hair dryers or phonographs because they hadn't heeded the DC-related warnings in their dormitory building. Most DC systems that remained, though, were for electric railways; that famous third-rail typically employs DC power between 500 and 750 volts, and the overhead catenary lines often use high-current DC. The choice of type of power is an example of structure or architecture concerns in systems—a concept introduced in the previous chapter.

As more and more power came to be generated by AC stations, the needs of these large DC applications were met thanks to the rotary converter, one of the most important inventions that you may never have heard of. It acts as a mechanical rectifier or inverter that could convert power from AC to DC (and vice versa). The rotary converter, which has since been largely supplanted by solid-state power rectification (although some railway systems still use the old technology), created increased usability and operability on the growing electric grid.

Maintainability/Reliability

The fourth ility of traditional engineering is *maintainability*. With its counterpart *reliability*, both are intimately related to quality and usability/operability.[17]

In the case of the automatic traffic lights, it is notable that cities did not initially have people who could maintain these systems, so eventually they were simply shut off in many places. In the *epoch of great inventions and artifacts*, reliability and maintainability had often been largely ignored.

The story of one of the fighter jets used by the U.S. Navy, Marines, and Air Force during the Vietnam War—the McDonnell Douglas F-4 Phantom II—elucidates the issue of maintainability. The F-4, which first entered service in 1960, was used extensively during that war. It was the principal fighter in the air and, over time, came to play important roles in ground attacks and reconnaissance. The plane, though, was not without its problems. For instance, the early aircraft had leaks in wing fuel tanks that required them to be resealed after each flight.[18] Problems of this sort began to attract attention. The Department of Defense (DoD), during the late 1960s and early 1970s, began to recognize an alarming trend across all military systems, where rising operating and support costs were using up much of the military budget and impeding the ability to achieve readiness goals. In the case of the F-4, its lack of *reliability* and *maintainability* was beginning to eclipse its value as a fighter jet.

In the case of an airplane, be it commercial or military, maintainability is a central concern. Airplanes are subjected to significant mechanical and thermal loads and varying weather conditions, and they have to fly and land thousands of times. A high level of maintainability helps in minimizing downtime and making the plane available to fly. The key metric is this: How many person-hours of maintenance are required for each hour of flight? In the case of the more than 5,000 F-4s that were

deployed in Vietnam, the cost of maintenance in person-hours, parts, and so on, had gotten out of control.

There are, in essence, two types of maintainability, as mentioned earlier. One is *preventive*, in which care is taken to ensure than an artifact or system doesn't break down. For mechanical parts, this might mean regular lubrication or periodic, scheduled oil changes—like you do for your car. It might mean replacing some parts *before* damage from wear and tear gets too bad. The other type is *corrective*, which involves repairing things that break to restore the artifact or system to its fully functioning state.

The jet fighter story highlights the connection between maintainability and reliability. Highly reliable systems often require less maintenance overall. They may require preventive maintenance (to prevent failures), but they typically require much less corrective maintenance in the form of repairs.[19]

The DoD's realization about the lack of maintainability of the F-4 and other military systems was a turning point. When it came time to procure a new fighter jet, action was taken. "Prompted by rising operating and support costs, shrinking procurement budgets, and deteriorating levels of operational readiness, the U.S. Navy challenged the F/A-18 Hornet program. … New high levels for reliability, maintainability, and operational readiness were specified."[20]

With the newer plane, the direct maintenance person-hours per flight-hour were reduced from 56.13 for the F-4 to 27.97 for the F/A-18, and eventually to 21.05 for later iterations of the F/A-18.[21] These figures speak directly to the jets' availability for use and the ability of designers to impart desirable ilities to a system as long as the importance of such properties is recognized early, the system ilities are anchored in the set of requirements, and sufficient resources are allocated during the design phase to achieve desired levels of safety, maintainability, or other ilities. As we will see for ilities such as flexibility and resilience that have emerged in the *epoch of engineering systems,* the degree to which they are achieved may also depend on the more fundamental choice of system architecture.

An Expanding View

As the complexity of artifacts grew along with a greater recognition of their importance in a systems context, the ilities of traditional engineering began to be viewed more broadly. New functions beyond the most

basic core functions of artifacts and systems continued to be added, as we discussed in chapter 1, and with that growing complexity designers and engineers worked to suppress undesired behaviors such as unexpected failures, difficulties with user interactions, excessive emissions, and so forth.

Increased complexity was not the only factor spurring an expansion of the ilities. The ilities were also strongly influenced by changing social values. The book *Silent Spring*, which we discussed briefly in chapter 1, brought issues such as environmental pollution into the consciousness of more people than ever before. Indeed, figure 4.2 shows that sustainability has been the fastest growing ility over the period since that book's publication. It is also notable that the number of Internet hits related to sustainability outpaced the scientific literature in this area (as we illustrated in figure 4.1).

The sheer scale of new technologies being adopted, coupled with these changing social values, helped fuel the need to pay more attention to the "old" ilities and consider "new" ones other than sustainability. Again, figure 4.2 shows that between 1950 and 1980 a new set of considerations—maintainability, scalability, modularity, and particularly interoperability—became increasingly important, and that much of this was fueled by population growth and the broader scope of deployment of complex systems for transportation, telecommunications, and energy. The preceding F-4 story shows how this happened in one sector with respect to *maintainability*. For *safety*, the automobile again provides a telling example.

For some time, car designs were including more and more safety features, from the rear-view mirror mounted on a racecar for the Indy 500 in 1911 and later adapted for street cars, to the turn signal first widely offered on 1939 model cars, to standardized sealed-beam headlamps introduced in 1940 and soon required for all vehicles sold in the United States, to Chrysler's 1971 introduction of a computerized three-channel, four-sensor, all-wheel antilock brake system on the 1971 Imperial—among many, many others. Beyond the artifact car, there was a recognition that safety was just as important an issue for the system as a whole. A good law governing traffic safety, and enforcement of that law, can be as critical—or even more critical—to the system as a car built with embedded safety features. Imagine if safety in any highly complex sociotechnical system addressed only the artifacts within the system rather than the wider system and environment in which those artifacts function.

The *epoch of complex systems* saw huge progress in terms of the four traditional ilities—quality, safety, usability/operability, and reliability/maintainability—especially as the wider, system-related view of them became prevalent. In some areas, better and stronger rules and regulations emerged, along with enhanced technology, in many sectors. Human factors (such as the problem of drunk driving in the context of safety in the transportation system) were given more attention. Notably, though, improvements have not been equally distributed in the world. For example, traffic fatalities are considerably worse in some countries than in others. Prior to the 1960s, the United States had the world's safest traffic, and by 2002 it had dropped from first to sixteenth place in deaths per registered vehicle and from first to tenth place in deaths for the same distance of travel.[22]

Some might argue that this deterioration in U.S. standing reflects an emphasis on regulating the vehicle and the auto industry, rather than focusing on consumer behavior. Both the U.S. federal and state governments have been reluctant to impose restrictions on the voting public. Thus, although seat belt laws were implemented, many states were slow to do so, and the laws requiring their use vary from state to state. Punishment for drunk driving, a major cause of traffic accidents, was only recently strengthened as a result of advances forced by organizations such as Mothers against Drunk Driving (MADD).

One thing is clear, though: Today, in the *epoch of engineering systems*, an engineer must be concerned with a whole host of things that in earlier epochs were not typically given much thought, or in some cases even a first thought. The expanded list of ilities speaks directly to the combination of multidimensional social and technical complexity that characterizes systems in our times.

How the Ilities Are Related

Based on our search results for all the ilities shown in figure 4.1, we see that some ilities are much more prominent than others. We also learn that different ilities became more important over time (see figure 4.2) and some, such as sustainability and interoperability, are still in a nascent state. But what are the relationships of the ilities to each other? We conducted a more detailed search on the World Wide Web, looking for instances where two illities (e.g., safety and reliability) are mentioned together. From this search, we constructed a 20–by-20 matrix showing which ilities are most strongly connected and whether these connections

are symmetric.[23] For example, we found that of the 69.9 million pages containing the word "reliability" as the first keyword, 15.3 million also contain the word "safety." In other words, 22 percent of hits about reliability mention its relevance to safety. On a scale of 0 to 10, this represents a strength of relationship of about 2 out of 10. Figure 4.4 shows a hierarchical network of ilities; their strength of relationship to each other is depicted by the strength of edges between the ilities based on the weighting just described.

The size of the nodes scales with the number of pages found on the Web containing a particular ility (see figure 4.1). The thickness of lines indicates the strength of relationships. This view tells us that quality, safety, and reliability—the classical ilities—indeed play a central role and that they are highly connected to each other and to some of the newer ilities. Another ility that features prominently at the center of the network is flexibility.

Figure 4.4 implicitly shows hierarchical levels: The most important and more independent ilities are shown near the center, whereas the

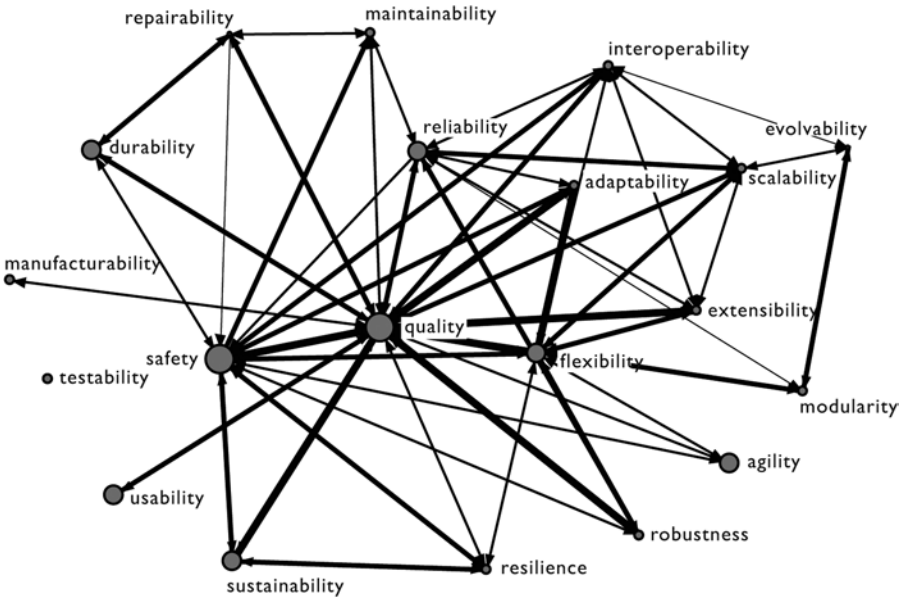

Figure 4.4
Correlation network of ilities based on a normalized 2-tubel keyword analysis. Node sizes correspond to the prevalence of each ility shown in figure 4.1, while the line thicknesses indicate the strength of relationship between two ilities.

periphery contains some ilities that, though important, are essentially supporting other ilities (e.g., reliability, durability, and robustness strongly support quality). Also, some ilities such as sustainability, resilience, interoperability, and evolvability are shown at the perimeter because they are relatively new and may not yet have developed their own set of supporting ilities.

Reliability, durability, and robustness have the expected strong relationship to quality and are shown to support quality directly with strong ties. Safety is also shown as a strong ility, with inward-pointing supporting properties such as durability, maintainability, reliability, and resilience, among others. Flexibility emerges as a strong cluster that includes robustness, modularity, extensibility, scalability, and adaptability. It is interesting to note that modularity appears to be an important enabler of both flexibility and evolvability. We note that sustainability in the lower left (our fastest-growing ility) has, as yet, no clear second-level supporting ilities, which may reflect its relative immaturity. The figure reveals a number of ilities that are all closely related to the concept of flexibility—the ability to change or adapt to new circumstances.

Other ilities *influence* but do not subsume each other (e.g., higher reliability will have a beneficial impact on safety, but itself does not guarantee safe operation[24]); and some are essentially orthogonal to each other and *have little interaction*.[25]

Figures 4.2 and 4.4 show that the ility explosion is part of the *epoch of engineering systems* and many strong connections exist among the various ilities. Let's look in greater detail at the most important ilities emerging in the current epoch.

Flexibility

Anyone who has ever used a Swiss Army knife appreciates the artifact for its *flexibility.* When Karl Eisner invented this knife in 1891, he was motivated to create a tool for his nation's army that was not manufactured by Germans. Combining a cutting blade, screwdriver, bottle opener, and all manner of other tools as the knife increases in size and complexity, the Swiss Army knife has become, for many, the very epitome of operational flexibility. More specifically, this type of flexibility is known as *reconfigurability,*[26] that is, the ability to change into different configurations that allow the system to perform multiple functions (in the case of the Swiss Army knife, cutting, filing, opening wine bottles, cleaning teeth, tightening screws, and so on) but not at the same time.[27]

No wonder Apache calls the "mod_rewrite" module of its HTTP server product "the Swiss Army Knife of URL manipulation,"[28] or that when Mercedes unveiled its BlueZero concept vehicle, built with the flexibility to insert electric, plug-in hybrid, and fuel cell technologies into the same vehicle design, the company was hailed for taking a "Swiss Army knife approach" to electric cars.[29] These are two of hundreds and hundreds of examples.

Eisner's invention speaks to one way in which flexibility is manifested, and that clearly predates engineering systems: the regime of *operation*. In use, the single "knife" can be operated in multiple ways to serve multiple functions. On your bike ride to the countryside for a picnic, your Swiss Army knife could come in handy for tightening the screws that secure your bike rack, opening the bottle of wine you brought, and slicing some salami. Similarly, in the regime of operation, your automobile transmission is flexible enough to allow you to operate optimally under different road conditions.

The other way flexibility is manifested relates much more closely to systems than to artifacts. In the regime of *redesign*, flexibility is about the relative ease with which a system can be changed to embrace a new function or engage with another system. How much redesign is required? Can it be done at reasonable cost?

In this regime of redesign, flexibility also functions as a kind of umbrella term for a number of other related ilities. *Evolvability* is about fundamental change to what might be called the "DNA" of the system—that is, the system's very purpose. This ility tends to be something that is manifested over the long term, and involves deliberate initiatives to enact. The term is clearly inspired by biological evolution in the Darwinian sense. It sits squarely in the regime of redesign and is a major theme of chapter 6. As we have already seen, *modularity* appears to be a major promoter or perhaps even a prerequisite for various aspects of flexibility. *Adaptability*, by contrast, is more like the classical Darwinian concept in the sense that changes in the system are driven by changing external environments. Striding both regimes of flexibility, operation and redesign, an adaptable system is one that can be reconfigured in response to external stimuli (such as a change in the environment of the system, like an organism adapting to the unfolding Ice Age, for example). Related to both is *agility*, the ability of a system to change quickly.

Two other ilities under the flexibility umbrella are scalability and extensibility. They, too, are closely related. *Scalability* is the ability to grow the size of a system to support a greater number of something. This could

be how many users the system supports, or how many daily transactions can be completed by the system, and so on. Scalability is about volume, and involves both of flexibility's regimes (operation and redesign). *Extensibility* is about extending the way a system works so that it can fulfill its original function *and* a different function or set of functions.[30] It defines a system's ability to add new functions over time. For example, a lawnmower might be extended so that it not only cuts your grass, but gauges the health of your lawn as you use it, or perhaps it can also be configured to cut a logo into the grass (like at a ballpark). An irrigation system may have been designed to cover only a couple of acres, but an extensible irrigation system can enlarge its footprint to cover a square mile.

All of these ilities fall under the flexibility umbrella. Notably, a system can be flexible in some dimensions and inflexible in others. For instance, a stamping press line in a factory has flexibility in that it can accommodate different dies to stamp out different shapes, but it is inflexible with respect to factors such as tonnage limits, size limits, and so on. Flexibility is a relative, not an absolute property of a system, and it can be deliberately designed into systems or exist fortuitously. Its increasing importance in numerous dimensions that affect systems is a direct consequence of the rate of technical change and complexity these systems are experiencing.

It is interesting to note that a large number of articles on "flexibility" were published even during the early days of the *epoch of great inventions and artifacts* and the *epoch of complex systems* (pre-1960), as can be seen in figure 4.2. A closer examination of these records, however, shows that flexibility was then almost exclusively understood as the property of materials and structures to bend under an oblique load and not in the more abstract sense of "ease of change" we think of today. Similar observations hold true for the term "extensibility," which was related primarily to the ability of materials to stretch under the influence of axial loads.

Few systems—if any—can match the Internet as an example of flexibility in both an operating and evolvability sense. It is a global system of interconnected computer networks that serves billions of users worldwide. The Internet's *scalability* is shown by how the system has grown to accommodate this enormous growth in what it today "runs" and the number of users. During the 1990s, for instance, the Internet handled an estimated 100-percent annual growth rate!

The interconnected networks comprising the Internet run the gamut from relatively small local networks to ones that span the globe, and are

based in public and private settings from government and academia to business and personal. The system links these smaller networks into the larger Internet with an array of networking technologies, outlined in chapter 3. The sheer volume and array of information resources and services this system carries is staggering: It hosts the World Wide Web,[31] provides the infrastructure for e-mail, and its technologies are increasingly used for telephone and television services. An estimated quarter of Earth's population use the Internet in some way, some 1.8 billion people as of July 10, 2009.[32]

The Internet system has many of the ilities that fall under the flexibility umbrella, too. Its e*xtensibility* can be seen in that it originally carried only text but has been consistently extended to photographs, sound, movies, and other material. Its extensibility can also be seen in the story of the Internet's origins. It began as an experiment funded by the U.S. DoD to link some of its research locations. The experiment was run through DoD's Advanced Research Projects Agency (ARPA), and by 1967 ARPANET was launched, using packet switching—a method for digital networking that groups data, regardless of type, into suitably sized blocks called packets for transmission (packet switching was also discovered in England at roughly the same time). From four domestic connections in 1969, ARPANET grew to provide a number of connections to Europe by 1973.

ARPANET's rapid growth required some protocol for host-to-host communications, and very effective protocol designs were developed relative to *scalability*. The first could not keep up with the growing load of traffic on the network, so researchers began to develop new protocols. Eventually, what emerged was TCP/IP (Transmission Control Protocol/Internet Protocol), along with the layered architecture discussed in chapter 3. The term "Internet" to describe a single, global TCP/IP network was first used in December 1974 by Stanford University researchers in their first full specification of TCP.[33]

When the U.S. National Science Foundation commissioned creation of its NSFNET in 1985, TCP/IP was chosen as the new system's core protocol. Three years later, the network was opened to commercial interests, first for e-mail services (some readers may remember MCI Mail, or Compuserve). In the same period, separate networks that had been created, such as Usenet and BITNET, merged with what was now called the Internet. Various commercial and educational networks interconnected with the Internet, too—all made possible by the amazing flexibility of the standardized TCP/IP and the layered architecture, which would

work over almost any existing communication network, along with standardized commercial routers and other equipment. Over time, nearly every existing public computer network was merged into the Internet.

TCP/IP helped give the Internet its tremendous *scalability*. The fact that TCP/IP is nonproprietary makes it easy for the network to expand, encourages interoperability of the various devices the Internet requires, and keeps any one company from exerting undue control (at least so far). Internet *adaptability*—the system's ability to be quickly reconfigured in response to external stimuli—unfolds in several ways. One is through the routing flexibility inherent in the protocol suite that now defines the Internet. Another is through Internet's Integrated Addressing System, which discovers and directs devices on both large and small networks within the larger Internet and permits them to be directed regardless of the lower-level details of each component network's construction. The way the protocol suite is configured to facilitate the routing of data and manage the efficient flow of information from one network to another is a sign of the Internet's *agility*.

The Internet demonstrates *extensibility* in how ARPANET, enabling its original function, extended to embrace what the Internet has become. Still, for some time the Internet was largely used by the academic, technical, and government communities. Public use of e-mail was beginning to expand, but most users tended to have e-mail at their places of work and not at home. What changed all this was the invention of the World Wide Web.

The Web was a project of Tim Berners-Lee[34] and others at CERN, the pan-European organization for subatomic particle research. It is a system of interlinked hypertext documents that can be viewed using a Web browser. These documents exist in the form of Web pages that may contain text, videos, images, and other multimedia. They are navigable using hyperlinks. In essence, the World Wide Web is a massive application, and Berners-Lee's genius was not only in the application he created but his decision to marry hypertext to the Internet. It is this marriage of the Internet and the World Wide Web that enables many aspects of the Internet's *evolvability*.

It is worth mentioning that the Internet has become *the* global communication system of choice, beating out at least two competitors, and this can be attributed directly to its flexibility. One of these competitors is the worldwide telephone system, generally, which managed to incorporate some data-transfer capabilities over time, but basically was not as evolvable or extensible as the Internet. The circuit switching (instead of

packet switching) was almost impossible for telephone designers to forgo because it was an inherent part of their solution to their overriding earlier concern with voice quality. However, the increased bandwidth of ever-improving optical cables has overcome this voice quality issue for packet switching.

The second, related system is Minitel, a service launched in France in 1982 by what are now France Télécom and La Poste.[35] The Minitel technology was "considered cutting-edge when Ronald Reagan was in the White House and Pac-Man and Asteroids ruled the arcade game roost."[36] This phone-line–based system was one of the most successful online services in the world before the World Wide Web was introduced. It allowed users to make train reservations, check stock prices, make various online purchases, search a telephone directory, and even chat with other users. It even showed some of the promise of the Internet with respect to social networking; a 1986 nationwide university student strike in France was largely coordinated through Minitel terminals and online connectivity.

Minitel is still in limited use in France. It has been eclipsed by the Internet and the World Wide Web by orders of magnitude, and the main reason is *flexibility*: The French system simply lacked the ability to accommodate all of what the Internet has become.

Resilience

Resilience is the degree to which a system can recover quickly from a major disruption while regaining—or even exceeding—its original level of performance.[37] That recovery may mean adjusting during the disruption or soon thereafter, so that the system can sustain its required operations under all conditions, whether expected or unexpected. Where designing for flexibility involves more proactive planning for possibilities, designing for resilience is about creating a system that can bounce back from something no one ever thought would happen. Early papers using the term resilience (see figure 4.2) referred to the ability of materials and objects to handle sudden drops or impacts, but now we also think of resilience as the ability of complex systems to respond to unanticipated shocks or events such as 9/11 or natural disasters such as earthquakes and tsunamis.

Like flexibility, resilience is an umbrella term under which other ilities can be found. We already mentioned that agility falls under both umbrellas. Elements of adaptability speak to a system's resilience, and

robustness is definitely a relevant sub-ility here. Robustness is the ability of a system to work as intended even when conditions change. However, resilience involves an aspect not as strong in other major ilities. It clearly reflects response to a major disruption, so the artifact aspect of design is not as involved as is the enterprise aspect.

The U.S. electricity grid is often mentioned as a good example of a system that *lacks* resilience. Clearly, the repair of local problems resulting from storms, for example, can often be very slow. However, it is rare that a large-scale breakdown, no matter how memorable, is not repaired immediately.

On August 14, 2003, North America experienced its worst blackout in history. It left more than 50 million people across eight U.S. states and Ontario, Canada, without power, and resulted in about $6 billion in business losses.

The detailed sequence of events contained in a subsequent investigative report by a special task force reveals the system's level of vulnerability. Although the full story is too lengthy to include here, a few elements of the sequence of events will illustrate the errors and limitations of the system, both technical and human.

Shortly after noon, an operator in Ohio failed to restart a monitoring tool after a problem had been corrected. At 1:31, a generating plant in Eastlake, Ohio, shut down, and about a half-hour later the first of several 345-kV overhead AC transmission lines in northeast Ohio failed because of a falling tree in Walton Hills, Ohio. About an hour later, another tree took down a 345-kV line in Parma, Ohio. When the voltage dipped temporarily on the Ohio portion of the grid, controllers took no action.

The line failures shifted power to another 345-kV line, and it sagged into a tree, taking it offline. The Ohio-based controllers dealing with these failures didn't bother to notify anyone in nearby states. Over the next 90 minutes or so, problems cascaded throughout northern Ohio as overloaded lines were tripped off, along with circuit breakers. A couple of minutes before the blackout, with Ohio drawing some 2 gigawatts of power from Michigan, the simultaneous overcurrent and undervoltage conditions caused the system to attempt to rebalance the system's voltage, causing new problems. In under 2 minutes, subsystems within the larger grid began to disconnect from each other: Eastern and western Michigan disconnected; Cleveland separated from Pennsylvania's grid; and the international connection between the United States and Canada began to fail. Ontario power plants began going offline, New York separated from the New England grid, and northern New Jersey separated

from New York and the Philadelphia area, which caused a cascade of failing secondary generation plants along the New Jersey shore and inland heading west to Ohio. By 4:13 that afternoon, 256 power plants were offline, 85 percent of which had gone offline after the grid separations as a result of automatic protective controls. Had it been nighttime, the Northeast United States and part of Canada would have appeared, from space, to have dropped off the face of the Earth.

Power had been restored by that evening to parts of the blackout area. However, some parts of New York City did not have power again until early the next morning. It took more than 24 hours for the Cleveland and Toronto airports to resume service, and Toronto's subway and streetcars had to be shut down until August 18 because of concerns that equipment might be stuck in awkward locations if another interruption in power occurred.

What might a resilient electricity grid look like? It certainly would not be one where the failure of a single part can cascade to bring the entire system down. In the years since the Northeast Blackout of 2003, some steps have been taken to increase the reliability of the system. The North American Reliability Corporation insists that events like those that led to the 2003 blackout are much less likely to occur. New standards and fees imposed on utilities that fail to meet them should help. But even with investment in new transmission lines, the system still has a very long way to go to correct the underlying technological deficiencies and architectural shortcomings as well as the human errors that exacerbated those problems. Part of the problem is the simple fact that so many power lines are so close to vegetation. And even if the reliability of individual components of the electrical system has now improved, it is not clear that the resilience of the system as a whole is any better now than it was in 2003. Meanwhile, the demand for power in North America continues to grow.

The electricity grid's lack of resilience is counterposed to the U.S. air traffic management, or air traffic control system. This has proven to be a highly resilient system—and it's a good thing, since aviation is so critical to the economy of a country as large in area as the United States. Planes move goods and services and support tourism, and the convenience of air travel coupled with the fact that long-distance travel by plane takes so much less time than other modes of transportation makes it an increasingly common choice.

The air transport system, though, is highly susceptible to problems and failures, both large and small. Anyone who lives on the U.S. East Coast

has experienced how weather-related delays in one hub city in North Carolina can cascade to bring the entire system, from Maine to Florida, nearly to a halt. A small maintenance problem on a plane on the ground in Chicago can cause other flights to be canceled or delayed as passengers wait for replacement equipment or simply for that plane to make it to, say, Atlanta for a flight to Boston.

In 1981, more than 13,000 U.S. air traffic controllers went on strike for better working conditions, better pay, and a 32-hour workweek. These controllers were organized in the Professional Air Traffic Controllers Organization (PATCO), a trade union that had operated since 1968. On August 5, President Ronald Reagan fired more than 11,000 of them after they refused to follow his order to return to work under the terms of the Taft-Hartley Act of 1947. Within a matter of days, U.S. Secretary of Transportation Drew Lewis had to organize and arrange for training of replacements, many of whom had little or no experience. Military air traffic controllers also stepped in to fill vacant positions, and overall flight schedules had to be reduced sharply. Despite predictions that it would break down completely, the air traffic system never came to a standstill and continued to operate despite this major disruption. It took the terrorist attacks of September 11, 2001, in the United States for a complete shutdown to occur, but even then the system bounced back very quickly.

On that day, the U.S. air traffic control system began a process that showed its tremendous resilience. In the minutes after the second plane flew into New York City's World Trade Center and the authorities realized that these crashes were not accidental, some planes in the sky began to receive messages on their cockpit printers that read: "SHUT DOWN ALL ACCESS TO FLIGHT DECK." Shortly thereafter, just minutes after a jet smashed into the Pentagon in Washington, D.C., managers at the command center of the Federal Aviation Administration (FAA) in Herndon, Virginia, issued an unprecedented order to every air traffic controller in the United States: "Empty the skies. Land every flight. Fast."

When air traffic controllers stop controlling, they call it "ATC zero"—a situation usually reserved for when radio transmitters go silent or radars fail in one part of the system. Planes stay in the air, and the redundancy in the overall system allows for controllers in other centers to take over and compensate for failure in one part of the system. This event, though, would give new meaning to "ATC zero."

The nation's air traffic control system facilitated the safe landing of almost 5,000 planes in under 2 hours, according to FAA radar records,

many of them at airports nowhere near their destinations. According to John Carr, president of the National Air Traffic Controllers Association, they "achieved the impossible."[38]

"It was something that had never been contemplated, something that had never been practiced. And yet they did it with professionalism and skill." Generally good weather across the country helped, and the early hour meant that few West Coast flights had yet taken off. But it was not the weather, nor the hour, that made this possible. It was the *system*, including its human component.

This highly resilient system resumed operations a week later, when the authorities decided to allow traffic in the skies over the United States once again. Technically, it could have been done on September 11, but flights did not resume until Friday, September 14, and regular service not until September 18. The system worked quickly to get all planes where they needed to be.

Of course, what the air traffic control system demonstrates in resilience it lacks in evolvability. For several decades, the FAA has attempted to modernize the air traffic control system in the United States, but most of these efforts—except for smaller local and regional changes—failed because it was not clear how the current successful system would continue to operate while the new system would be phased in. The newest initiative, called Automatic Dependent Surveillance-Broadcast (or ADS-B), is a technology with both ground equipment and on-plane equipment that has the potential to reduce the workload of controllers and give pilots more freedom in choosing altitudes, while maintaining the highest levels of system safety. The FAA has mandated full introduction of ADS-B in the United States by 2020, but many obstacles remain.[39]

Interoperability

The last of our highest-level ilities is *interoperability*, which characterizes systems that can function independently in their own right but can also work together as a larger whole, even if they are owned and operated by different entities and were not designed originally to work together. The related concepts of collaborative systems and systems of systems is covered in some depth in chapter 6.

Related to interoperability are *compatibility* and *modularity* (sometimes called decomposability). Compatibility tends to relate to consumer products and systems, although not exclusively, and describes how well components of the system can be connected and work together.

Modularity has two primary aspects. One is functional decomposition and encapsulation: that the subfunctions of the larger artifact or system, which have their own larger function, can be decomposed and assigned or "encapsulated" into particular smaller units or "modules" of the overall systems. The second is that the artifact or system can be pulled apart and can be put back together with relative ease. The presence of both aspects gets you the greatest modularity, which is a powerful enabler of interoperability and other ilities.

One of the best examples of interoperability is the system in which your personal computer functions. This example also shows us how much more important both aspects of interoperability—modularity and compatibility—have become in the past 30 years.

From the introduction of computers, IBM used the marketing strategy of product bundling to gain market dominance. When personal computers first came into widespread use in the early 1980s, most manufacturers also bundled, that is, they sold the hardware—central processing unit, keyboard, monitor, printer, and so on—as a combined product, typically under the same brand name. But the bundling did not end with hardware; typically, the product bundle also included preloaded software on the computer, including the operating system, a word processor, a spreadsheet program, and perhaps some type of database application. If you bought an IBM PC, it was IBM peripheral equipment you had to use. If you bought a Tandy computer, your printer would be a Tandy printer.

Gradually, this began to open up, and later generations of personal computers offered options. You might be able to use a printer made by a company other than IBM with your IBM personal computer, so long as the connections ("ports") matched, you had the correct driver installed, and the components of the system could be hooked together. This was the beginning of compatibility in personal computer systems.

Concurrently, the modularity of the IBM personal computer's architecture created opportunities for other companies. Columbia Data Products was the first company to capitalize on this by creating a clone of an IBM personal computer—an almost exact duplicate of all the architecture's significant features. Compaq became the early leader, creating an IBM AT clone in the early 1980s. Later, other companies—notably Gateway and Dell—joined Compaq to take modularity in personal computers even further, and began to manufacture personal computers that integrated a small number of modules and made it possible to reduce considerably the time and cost of assembly, thus enabling mass customization.[40] Oddly enough, the clones and later variations came to be

known as the "IBM PC compatible" computers, thus fusing the two primary sub-ility terms of interoperability. The standardization of interfaces and architecture in the PC industry also created vast opportunities for chip manufacturers such as Intel, which gained substantial strength and demonstrated the importance of control of the supply chain in an interoperable environment (the "Intel inside" advertising campaign illustrating Intel's great success in this regard). Interoperability changes the environment not only for the original equipment manufacturers (OEMs) but also for powerful suppliers.

Over time, interoperability was greatly expanded, until today we have one of the most useful devices ever invented for the personal computer: the Universal Serial Bus (USB) that has largely replaced serial and parallel ports in most personal computer peripherals. The USB makes all manner of devices easily compatible, from mice and keyboards to digital cameras and video game consoles.

Despite their somewhat awkward collective name, the ilities nevertheless capture and express the subtle and important behavior of systems beyond their primary intended function and use. During the early *epoch of great inventions and artifacts* (about 1880–1920), the classical properties of systems were born: safety, quality, and reliability. It was not enough to launch a new product; one also had to ensure that the product would not kill or injure people (at least not too often), that it was well made, and that it would work for longer periods of time without breaking down. During the *epoch of complex systems*, as highways were built, telephone networks expanded, and the electrical grid reached into nearly every household of an increasingly industrialized world, new properties such as usability, extensibility, and robustness became increasingly important.

Today, as we have entered the *epoch of engineering systems*, the complexity and density of connections between previously separate systems keep surprising us. Unexpected shocks and the finiteness of our resources become more apparent. So, we grasp at yet another set of illities such as resilience, flexibility, and sustainability. Unlike the classical illities, these new ones cannot be directly associated with individual components or artifacts. They result from the collective structure and behavior of the various technological, human, and natural components and subsystems that are woven together in complex ways. Thus, the kind of thinking stressed in chapter 3 demands that these ilities be given due consideration.

It must also be said that culture has a large effect on the ilities and, in particular, on how ilities are prioritized, how they are implemented, and how trade-offs among ilities are resolved. Take auto safety. In the German national culture, the absence of speed limits on the Autobahn is a given. In Japan, the culture does not tolerate drinking and driving, and so there are strict requirements for designated drivers. Until recently, mitigation of noise pollution was of far greater concern in Europe than in the United States. Urban air pollution was of more concern in the United States, which led to some of the strictest emissions standards in the world, often spearheaded by the state of California. Culture is a big factor when it comes to the life-cycle properties of engineering systems. Thus, understanding these life-cycle properties of engineering systems requires not only a mathematics- or physics-based perspective but a deep appreciation of social factors.

Given these complexities, how can engineering systems be modeled and analyzed in more depth? That is the subject of chapter 5.

Suggestions for Supplemental Reading

Richard de Neufville and Stefan Scholtes (2011), *Flexibility in Design*, Cambridge, MA: The MIT Press.

Leonard Evans (2004), *Traffic Safety*, Bloomfield Hills, MI: Science Serving Society.

Nancy Leveson (2011), *Engineering a Safer World*, Cambridge, MA: The MIT Press.

Joel Moses (2004), "Foundational Issues in Engineering Systems: A Framing Paper," at http://esd.mit.edu/symposium/pdfs/day1-2/moses-slides.pdf.

Yossi Sheffi (2005), *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*, Cambridge, MA: The MIT Press.

Genichi Taguchi (1986), *Introduction to Quality Engineering*. White Plains, NY: Asian Productivity Organization, UNIPUB.