Second Edition

HUMAN FACTORS ENGINEERING AND ERGONOMICS

A Systems Approach

Stephen J. Guastello

CRC Press
Taylor & Francis Group

Second Edition

HUMAN FACTORS ENGINEERING AND ERGONOMICS

A Systems Approach

Second Edition

HUMAN FACTORS ENGINEERING AND ERGONOMICS

A Systems Approach

Stephen J. Guastello

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20131017

International Standard Book Number-13: 978-1-4665-6010-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

This textbook is the outgrowth of teaching human factors engineering for 30 years to undergraduates. The course is an offering of the psychology department, just as it was decades ago when I was a student myself. The field of human factors psychology (or human factors engineering, or engineering psychology) has changed markedly during that time. Although it still stays true to its original concerns about the person–machine interface, it has expanded to include new developments in stress research, accident analysis and prevention, and nonlinear dynamic systems theory (how systems change over time), and some aspects of human group dynamics and environmental psychology. Computer technology has permeated every aspect of the human–machine system, and has only become more ubiquitous since the previous edition. The systems are becoming more complex, thus theories need to evolve to cope with the new sources of complexity.

It has been a challenge to find a textbook for the class under these conditions of technological change. At first, I found one that seemed just perfect with regard to the breadth and depth of coverage I was looking for. After a few years it only needed a supplementary reading or two to help out, but eventually it went out of print, never to return. The other textbook choices by that time had diverged greatly in how they characterized the scope of the field. One approach concentrated on tables and graphs for otherwise traditional topics. A second approach retrenched into the theories of cognitive psychology and focused less on the practical problems in human factors. Meanwhile, library shelves were filling up with books on human–computer interaction that were becoming progressively more dissociated from the core concepts of the human–machine interface. The fast pace of technological change did not help any textbook writer who had a mind to identify and extract the fundamental principles of the subject area.

In any case, I hereby present to you the new scope of the psychology of human–machine interaction. The typical roomful of students that I have in mind is usually composed of upper division students and a few graduate students. The class is typically composed of 60% engineering students of different sorts, 35% psychology students, and 5% sundry others. One implicit goal of the course is for the engineers to think more like psychologists, and the psychologists to think more like engineers. The sundry others usually show signs of thinking like both, and make the class situation more interesting for everyone.

I would like to take this opportunity to thank Joseph J. Jacobsen for helping to arrange some of the photographic opportunities that appear throughout this book.

New to This Edition

There are several new developments in this second edition, some of which span multiple chapters. Some of the more extensive areas of expansion include the following.

In earlier times, the allocation of functions between person and machine was a fairly straightforward separation of labor. Augmented cognition and automation are blurring this distinction, and the current thinking is that machines can adapt their functions when they detect that the operator is under stress. How does the machine make that determination? Is part of the stress related to the human not trusting the machine? The issues are first addressed in Chapter 2, but become more complex as the book unfolds.

The role of nonlinear dynamics in person–machine systems has grown as interest in how systems change over time has increased. The newest developments involve events that synchronize or emerge suddenly, apparently out of nowhere. The basic principles that are relevant to human factors have been expanded in Chapter 2, with specific applications to stress and human performance, occupational accidents and prevention, and complex systems in later chapters.

The design of visual displays has always been a staple topic in human factors. The latest innovations draw on evolving principles of iconography and synthetic vision systems. Although it was tempting to include them in the chapter on visual displays, they appear in the chapter on human–computer interaction where they build on some intervening material.

The chapter on cognition has changed substantially in light of developments that have consolidated in several theoretical areas. Look for resource competition theory, recognition-primed decision making, working memory, degrees of freedom in cognitive processes and psychomotor control, dynamic decisions, task switching and interruptions, subjective ratings of cognitive workload, and situation awareness. Navigation in real-world environments is also new to the chapter on environments.

New directions in psychomotor control are combining system control with the understanding of neural mechanisms that produce the behaviors. The practical objectives often surface in robotic components and designs for prostheses that can be controlled by human thought the way normal limbs are controlled.

There is a growing concern in society about cognitive workload and fatigue as more people find that they have too much to do, not enough time to do it in, and automation, which they do not necessarily trust, is making their work more mentally intensive. Long working hours can be hazardous. The effects of cognitive workload and fatigue have been historically difficult to separate because they both produce negative effects on performance over time, while other processes that transpire over time serve to improve performance. It is now possible to separate these effects with nonlinear models and experimental designs that capture all the ingredients of the models. Whereas systems have been traditionally designed to minimize the role of individual differences in cognition and personality, and with good reason, individual differences in cognitive responses to systems are now gaining importance.

New developments in accident analysis and prevention consist of a better understanding of the social context in which safety and risk occur. There has also been a recent shift in orientation from explaining *what happened* to a proactive orientation to anticipating *what could occur*. This is the principle of system resilience that takes on a few different forms and is informed by nonlinear dynamics and complex systems theory. Another distinctive feature of the accident chapter is that the catalog of interventions for occupational safety has been updated. Once again, a comprehensive ergonomics approach to intervention offers some of the largest improvements in accident reduction.

Artificial intelligence was once the label given to computer programs that emulated human thought processes. The programs have now expanded to include knowledge discovery that is more independent of human control than ever before and to manage and manipulate huge databases that never used to exist. It is left to the reader to ponder whether we are reaching a turning point where the machines will control humans more so than the other way around.

Complex systems applications in human factors and ergonomics are becoming more common. The broadest new developments here include network theory, human collective intelligence, and how to control your very own fleet of robots or drone surveillance aircraft. The chapter also considers scenarios in which multiple integrated autonomous systems can go out of control.

Author

Stephen J. Guastello is a professor of psychology at Marquette University, Milwaukee, Wisconsin, where he specializes in human factors engineering and organizational behavior. He was a pioneer in the applications of nonlinear dynamic systems theory to psychology and in the development of statistical analyses for behavioral science data. He is currently researching nonlinear processes in cognitive workload and fatigue and sees human factors and ergonomics as ripe with new possibilities for complex systems analyses. Professor Guastello earned his PhD in industrial/organizational psychology from the Illinois Institute of Technology and BA degree in psychology from the Johns Hopkins University.

Some of his more extensive consultancies include human–computer interaction and expert systems for power plants and other applications, accident analysis and prevention, and the usability of prostheses. He has authored two previous books, *Chaos, Catastrophe, and Human Affairs* (1995, Erlbaum/Taylor & Francis) and *Managing Emergent Phenomena* (2002, Erlbaum/Taylor & Francis), and coedited *Chaos and Complexity in Psychology: The Theory of Nonlinear Dynamical Systems* (with M. Koopmans and D. Pincus, 2009, Cambridge University Press) and *Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data* (with R. Gregson, 2011, CRC Press). He is the founding editor-in-chief of the quarterly research journal, *Nonlinear Dynamics, Psychology, and Life Sciences*, published by the Society for Chaos Theory in Psychology & Life Sciences.

1

Introduction to Human Factors and Ergonomics

Ever since prehistoric times, *Homo sapiens* made tools to enhance their physical capabilities and to channel their physical energy. Eventually, they developed tools to enhance mental capabilities and channel mental energy as well. The first tools were handheld devices, such as hammers and cutting edges. The first machines such as notably wheels, levers, and pulleys channeled physical forces as well as human forces.

The industrial revolution, which started by some accounts in the late 17th century, benefited from major developments in the science of mechanical physics. Sir Isaac Newton is historically recognized for introducing the concept of a mechanical system. A system has parts that all work together to fulfill an objective. If we know how each of the parts work, we can know how the entire system works. This particular assumption produced clocks, clockwork-type machinery, and myriad devices that, once the human operator set them in motion, continued their function until completion. Machines such as the cotton gin and textile weavers started to replace the work previously done by people; we recognize the phenomenon as *automation*.

The advent of machine tools in the 19th century made it possible to produce hundreds or thousands of objects that looked and functioned exactly alike. If a product were to be assembled from many elementary machine-tooled objects, the product could now have interchangeable parts. Hence, there was now the potential for assembly line production systems.

Electricity was another landmark in tool making. It greatly expanded the number and type of tools and machines that could be made and reduced the amount of human energy required to make the production system work. The number of tools and products that depend on electricity today are uncountable. Before taking the ubiquity and inevitability of electricity for granted, however, it is important to remember that entire empires were built without it. There are still places in the inhabited world where electrical service is only available for 1 hr a day.

Communication devices have changed not only our way of life, but also our concept of who we are and how we are connected to other people. In primitive times we chiseled a piece of rock or scratched on a piece of papyrus. Writing and writing media were memory-enhancement devices at both the individual and collective levels of experience. Not only could the writers enhance their own memories, but they could also share them with anyone else who found the writing even after the original writer had died.

Books of memories were produced eventually, but until the invention of the printing press in the late 15th century each book had to be copied accurately by hand. The type for the print still had to be set by hand until the late 19th century, but the printing press was a major leap forward for automating the process and disseminating the information. In the same historical epoch, electricity was combined with communication goals to produce the telegraph, telephone, radio, and television later on; these devices allowed us to send messages over great distances (and eventually the entire world) in relatively short amounts of time. Other technical advances allowed the possibility of recorded sound and its automated production once the record master was available in 1916. Photography displaced the painter as the sole provider of realistic renditions of visual events. Photographs were combined with a hand-cranked machine to produce a movie; eventually, a motor replaced the crank. The communication devices were not only tools of communication, but they also became media of expression.

Transportation devices, especially the aircraft of the later 20th century, allowed us to move our entire bodies around the world in ways never before possible. Space travel has started to liberate us from one particular world, although the other nearby worlds tend to be cold and nasty. If people were meant to fly, they would need to build a really good set of wings and learn how to use them.

The earliest ancestor that resembles the device we know today as the computer was introduced in the late 1940s. The capabilities of the earliest versions of the machine were confined to rapid and accurate calculations, which were of course highly desirable. Other types of thinking processes were introduced into its programming in the years that followed, along with simplicity of use. Desktop models for individual owners appeared in the 1980s. The 1990s saw larger data storage and memory capacities, faster processing, and briefcase-sized models. Since then, computer-related technologies have been incorporated into virtually every machine of contemporary design. In some cases, a desktop computer acts as the controller for a complex electromechanical system.

Although the foundations of the shift were forming earlier, the prototype device we call "the computer" has become intertwined in every communications system available, short of talking, paper, pencil, and those who have kept the art of chiseling rocks alive. Actually, the ubiquity of e-mail, cell phones, and text messaging almost appear to preempt real face-to-face relationships and replace paper-based objects and activities with those that are plugged into an energy source and an information network. Can people do anything anymore without a computer-based device becoming involved? This question extends to simple arithmetic and walking across the street. The information technology industries have tacitly answered, "not if we can help it," and call what they do "ubiquitous computing."

Computers differ from conventional machines because they are programmable. Programming allows people to change the function of a machine in ways never before seen since the introduction of machines. Not only does the computer enhance human cognitive processes, it can substitute artificial skill for real skill, alter our concepts of ourselves as social beings, and create artificial psychological realities through particular combinations of programming capabilities; the more advanced combinations are called "virtual reality." All of these developments led Turkle (1984) to refer to the computer as *The Second Self*, to emphasize that this class of machines has affected our concept of who we are more so than any other technological advance. Thirty years later, the computer is more of the same.

Entry of Human Factors and Ergonomics

The complexity of machines is obviously growing, as are the demands on the user to operate them. Machines are only as effective as their operation by humans allows, and thus, there are opportunities and challenges that lie at the interface between the person and the machine. The setting in which the machine is used, the physical environment, other tasks underway, and other people and machines have a strong impact on the performance of the person–machine system (PMS).

Person–Machine System

Meister (1977) emphasized the concept of a *system*. A system has elements that interact to produce a result that cannot be reduced to, or understood as, the simple outcome of one of the system elements. The primary elements, of course, are one person and one machine, but person–machine systems can vary in size and complexity.

The person and the machine exert a close control over each other within the system. The point of control occurs at the interface, which is most often a control and display panel of some sort. Although Meister (1977) conceptualized the person–machine system as only containing those elements or people that have direct control over the machines, ergonomic science in later years recognized less direct sources of system control that do in fact affect the interactions at the interface.

The machine should be capable of carrying out functions automatically once the human has activated it. If not, then we have a tool rather than a machine. Both tools and machines are important to human factors engineering or ergonomics nonetheless.

Cognitive Core

The field of human factors engineering (HFE) started in the 1940s as a joint effort between psychologists and engineers (Chapanis, 1975). Their overarching goal was to study the interactions that occur at the human–machine interface. They studied practical questions such as, "Why did a seemingly competent cockpit crew fly their aircraft into a mountain?," "How could the chemical tragedy in Bhopol, India, ever happen?," or perhaps more mundanely, "Why are people having such a hard time operating our software?"

Figure 1.1 depicts the traditional scope of HFE. The human interacted closely with the machine. The information provided by the machine to the human is a *display*. The information given to the machine by the human is a *control*. Three groups of cognitive psychology topics that were immediately invoked were perception, psychomotor skill, and principles of cognition. (In the early days, cognitive science was not yet defined as a discipline.) Theories of perception explained how the human assimilated the display. Psychomotor studies explained how the person managed to control the machine effectively. Cognition theories explained what transpired between the human's ears between the moment of

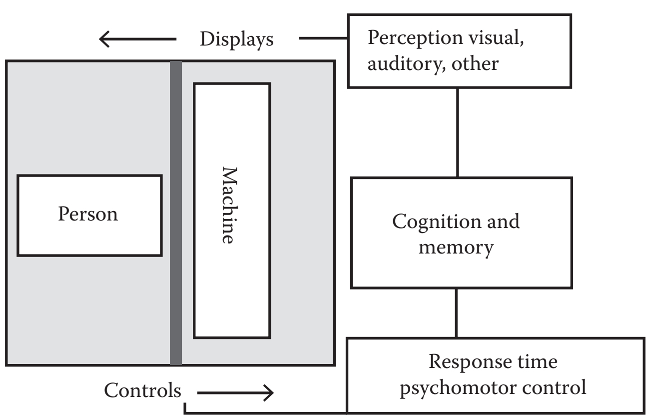

FIGURE 1.1
Traditional range of HFE topics.

display and the moment of control. The overriding objectives were, and continue to be, the optimization of system performance and the minimization of human or system errors.

Chapter 2 covers a set of concepts that are pertinent to most any human factors initiative. Who is the user of a system? What composes the system? How do we allocate functions between people and machines? When are redundancies in function relevant? What is information and how does it flow? How do system events change over time? How do we assess the functionality of the system?

Chapter 3, on psychophysics, is actually about the interface between people and their real worlds. It explains how people detect a physical stimulus—sound, sight, vibration, and so on—and how to characterize simple decisions that are made on the basis of the presence or absence of stimuli. There are stimulus detection problems that are not nearly so simple, however.

Chapter 4 pertains to the processes of visual sensation and perception and their impact on the usability of visual displays. Chapter 5 pertains to auditory sensation, perception, and displays, and also includes the human interpretation of tactile information. These core topics have a long history in the field, and there are innovations to keep them exciting.

Chapter 6 delves into the cognitive processes that occur between the perception of a display and the moment of a control action. How did the human make the evaluation and decide what to do? The principles in this chapter are foundational to Chapter 12 on artificial intelligence. Chapter 7 describes controls, human response times, and related issues of human performance when operating machines.

Work-Space Shell

The subject matter of HFE expanded in the 1980s. The word *ergonomics* arrived from Europe. Initially it denoted the interaction between the human and the nonliving work environment, but today, *ergonomics* and *HFE* are interchangeable in meaning. Contemporary ergonomics is now conceptualized as "the design and engineering of human–machine systems for the purpose of enhancing human performance" (Karwowski, 2000, p. 1).

The expanded range of topics integrated well with another classic area in Chapter 8, *anthropometry*. Anthropometry, or human measurement, affected the design of hand tools, workspaces, driver, and passenger compartments of vehicles of all types, and requirements for human movement within workspaces. In this chapter we consider topics in human strength because of their proximity to other human-measurement concepts. Two regular topics of concern are the control of back injuries and carpal tunnel syndrome. Both are considered in Chapter 8, although the broader range of topics in biomechanics is beyond the scope of this text.

Figure 1.2 depicts the scope of HFE as it exists today. The immediate work environment surrounds the person–machine interface, which is in turn surrounded by environments of broader physical scale, known as microenvironments and macroenvironments. Kantowitz and Sorkin (1983) were among the first authors to expand the boundaries of HFE to these broader physical ranges.

Environments are sources of many forms of stress. Although the study of some types of stress and human performance is part of classical HFE and nearby areas of work psychology, the major thrust of the subject matter was not introduced until the 1980s. In a departure from the traditional impression that heat, cold, and noise are the primary stressors, some attention to the social origins of stress are given as well in Chapter 9 along with shift work, cognitive workload, and time pressure. Workload and fatigue effects have been historically very difficult to separate because they both occur simultaneously along with

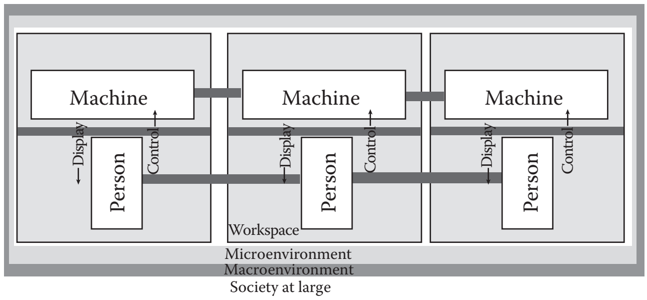

FIGURE 1.2
An expanded view of HFE.

other dynamics that improve performance over time. Progress has been made to separate them, however, using a mathematical modeling approach and experimental designs that are sufficiently complex.

We now arrive at the topic of accident analysis and prevention in Chapter 10. Accident analysis and prevention has an independent literature, which made use of ergonomics thinking only relatively recently. Traditional HFE appeared to regard the topic as another reason to control human error from the perspectives that it has always taken. Systems thinking promoted an integration of ergonomics, social dynamics, and management into new approaches to accident analysis and prevention in the workplace. Although this HFE text is *not* about business management in its usual scope, some aspects of management do play a serious role in accident analysis and prevention and ergonomics policy in work organizations, however. Also, cultures differ on their perceived importance of accident analysis and prevention relative to other economics objectives.

Computer

Human–computer interaction became an active area of study in the 1980s. The area requires a good deal of HFE and ergonomics thinking, especially since the influence of computer-based products has become widespread. Basic HFE concepts were still relevant to these new applications, but computer-based systems raised a number of new questions and issues. The development of computers in the next decades was the result of interplay between developments in programming capability and developments in cognitive science. Books on human–computer interaction and computer chapters in HFE books appeared but became obsolete quickly.

In this writer's opinion, the developments in computer science have stabilized sufficiently, or at least have left a long enough trail, to prepare two chapters on computer-relevant themes. Chapter 11 covers input–output devices and virtual reality. Chapter 12 covers artificial intelligence, expert systems, and systems requiring massive computation components.

Chapter 13 pertains to complex systems that include but are not limited to complex human–computer systems. Two of its themes address the questions, "Why do attempts to improve or fix a system sometimes backfire and produce a situation that is worse than the previous one?" and "How do events in the system emerge that are not readily explained by events in their elementary parts?" Workgroup coordination, synchronization, and collective intelligence are also included in that chapter.

Beyond the Great Outdoors

Chapter 14 considers environments that are larger in scale than the immediate work environment. Microenvironments would include private homes, offices containing several workspaces, or factory departments. Macroenvironments are substantially larger and more complex, such as entire factories, housing developments, airports, and even cities and the great outdoors. Where should we put the nuclear waste dump? How do we find our way out of the nuclear waste dump?

Human factors in outer space comprise the final saga of the text. This trajectory through the history of the U.S. space program will highlight HFE issues that emerged over the decades. Themes from prior chapters will converge here along with space-specific problems.

Broader Themes in the Text

Karwowski (2000) emphasized that ergonomics today requires a greater use of general theories than what used to be the case in previous eras. The subject matter of traditional HFE, when presented in its most traditional fashion, was composed of elementary facts and reference tables. Although the data- and event-driven approaches placed a strong emphasis on practicality, theory had to emerge from the bottom up to account for facts and allow for equally useful generalizations. As technologies evolve, one must approach new and untested technologies with a deeper understanding of the principles that might be involved. Some of the broader themes of this book are:

1. The human–machine interface is changing alongside the changes in the allocation of functions between person and machine. The tool that was under the continuous direct control of the human hand evolved into a machine. The machines evolved in their functions so that the primary point of contact with the human was at the interface where information was sent and received. Then, the interface became physically separated from the rest of the machine in many cases, and the boundaries between the control and the display merged together. Machines evolved beyond sending and receiving information; they started to pick up some of the thinking tasks as well. With virtual reality, the interface became an environment that surrounds the human; in other words, the human is in the machine. But why stop there? We can have machines carrying out their function from within the humans' shirt pocket or from inside their bodies.
2. The needs of the system user affect optimal design in both broad and specific ways. Just because it is possible to make a machine, it does not follow that the machine is optimal in any particular way. Historically, it has been more often the case than not that the machine came first and the finesse of usability came much later. The introduction of a new display or control may have been a large step forward without much regard for how well the new feature was designed or positioned. Usability *did* matter eventually, however.
3. Systems behave over time. Thus, concepts from nonlinear dynamics, such as chaos, catastrophe, and self-organization, are relevant to the understanding of system performance over time, especially the complex systems. Nonlinear dynamic systems theory, which is first introduced in Chapter 2, clarifies that events do not

change smoothly over time, as the thinking behind common linear statistics might suggest. Rather, a small change in a system parameter could have a dramatic effect if it occurred at the right place and the right time. Similarly, a large change in a system parameter could have no discernible or permanent impact on the system's output. Nonlinear dynamics also underlie *emergent phenomena*: system states that bear no visible similarity to previous states and that would appear to have come out of nowhere to the untrained eye. Within the nonlinear dynamics framework, it is processes that lead to emergent phenomena that put the "complex" into a complex system. Complex systems require more specialized analyses of system performance and human error potential compared with simpler systems.

4. HFE concepts are applicable anywhere, and nearly everywhere. Although elaborate commercial systems such as power plants and aircraft offer many opportunities for human factors investigations and thus receive a lot of attention, sophisticated technologies are inherent in contemporary household appliances. Do you know how to program your VCR or satellite download service to record a broadcast?

Criteria of Human Factors

Karwowski's (2000, p. 1) definition logically applies to the entirety of the field: "the design and engineering of human–machine systems for the purpose of enhancing human performance." The central criterion plays out in three forms: human performance itself, industry standards, and legal responsibility.

Performance Criteria

Performance criteria include maximizing output, minimizing error, ensuring safety, and devising training programs to meet these standards. These criteria are often interrelated. It is often possible to introduce a design idea that satisfies more than one objective. On the other hand, gains in speed might result in compromises in error rates or safety. Thus, it is well advised to consider multiple aspects of system performance when evaluating a design.

Output, from the standpoint of most human factors issues, is calibrated in terms of the amount of time required to produce a unit of work or a particular decision. Here we evaluate the ease of garnering relevant information from a display, making a decision, and executing the response.

Errors can be trivial or costly, depending on the situation. Managerial decisions often toggle between speeding production and compromising the quality of the product or the quality of work life (Borges & Guastello, 1998). The impact of errors can be a matter of degree—some of the more poignant examples considered in this text have resulted in deadly tragedies. Chapter 7 examines the speed-accuracy trade-off explicitly. Issues in quality of work life that translate in some of the sources of stress in the workplace are examined in Chapter 9.

Industry Standards

Various industries have at least some standards for equipment design. Human factors issues often fall outside the boundaries of many official regulations, but not in every case (Karwowski, 2006). Standards have different levels of officiousness. A *state of the art* standard represents the most progressive idea that is publicly known. State of the art ideas might not have been thoroughly tested, and could easily be upstaged by an equally experimental new idea before the evaluation of the human factors is reasonably sufficient. *State of the science* standards reflect scientifically supported ideas that have a reasonable amount of generalizability. The *state of the industry*, however, is often slower than the state of the science to uptake the scientifically sound principles into practical systems. Old equipment that is still functional is not going to be discarded just for the sake of keeping up with trends. New purchase decisions, however, are likely to reflect the more progressive systems designs. Thus, within a particular industry one can anticipate a cascade of technologies.

Industry standards reflect the minimum acceptable design and performance standards in many cases. They usually adopt requirements that are least ambiguous in terms of their scientific support for their efficacy in system performance. Although industry standards per se do not have the force of law behind them, they often serve as reference points in civil liability litigation where the goal is to determine whether a design flaw was responsible for an injury or another form of damage. Representative of the industries themselves are usually part of the process of developing regulatory standards.

Regulatory standards have the force of law. As examples, the Federal Communications Commission sets standards for the quality of audio and video broadcasts and the technical systems by which the broadcasts will be made. The U.S. Department of Defense (DOD) sets standards for the human factors attributes of equipment purchased for the military. The standards are strongly influenced by the results of scientific contracts issued by the DOD. Any manufacturer desiring a military contract must meet those standards. Although DOD standards do not pertain to civilian products that are not purchased by the military, their information base might contribute some influential information to a civil litigation.

Another governmental nexus surrounds heath and safety. The Occupational Safety and Health Administration (OSHA) regulates safety in the workplace. The majority of its work since its inception in 1971 was restricted to employers' use of equipment and supplies—electrical configurations, chemical handling safety, proper ventilation and exhaust, personal protection equipment, and so forth. OSHA also has standards for exposure to noise and excessive heat, which are two clearly ergonomic matters. In late 2000, Congress passed a law extending OSHA's purview to several other ergonomic matters, but rescinded the law in early 2001. Since that time OSHA has served an informational and advisory role to employers on ergonomics matters. Although its primary function is setting standards, auditing, and enforcement, it now positions itself as a consulting resource to employers who wish to know more and take care of situations themselves. The National Institute of Occupational Safety and Health (NIOSH) is a research branch of OSHA. NIOSH commissions studies of situations and investigates the scientific side of new and emerging issues. NIOSH is currently part of the Centers for Disease Control and Prevention, which is concerned with public health issues outside the workplace as well as within it.

Civil Liability

The 50 states vary somewhat regarding their criterion for proving a case against a manufacturer or other defendant in a civil liability trial, but there are some common concepts nonetheless. They are strict liability, negligence, and contributory negligence.

Strict Liability

A legal theory of strict liability states that a manufacturer of a product is liable for damages connected to the product because the product is produced for profit. Here one might examine whether a human factors flaw could be detected from the product design, and whether such a flaw could have reasonably induced an error that led to the accident.

Dangerous products such as chainsaws are made and sold regularly of course. One suitable defense for the manufacturer is the obviousness of the danger. Is there any question about the function a chainsaw is supposed to serve and what the blades are for? Some dangers on other products are less obvious, however, and the burden of responsibility is on the manufacturer to provide suitable warnings and instructions for safe operation of the product. Appropriate warnings constitute another viable form of defense. One might then debate whether the warnings were in fact clear, informative, and relevant to the danger in question.

Contributory Negligence

In a defense of contributory negligence, the defendant argues that the plaintiff took actions that were irresponsible in some way. One example would be to use the product for a purpose for which it was not intended. A tale from the insurance industry captures the essence of improper use of a product. Several decades ago, a manufacturer of home washing machines designed a model with a glass top. The idea was that the happy housewife could watch the clothes go around without opening the lid. One proud owner, however, decided to dance on the lid of the washing machine. Unfortunately, she pounded her foot down too hard, it broke through the glass, and her leg was torn up by the washing mechanism. Although the manufacturer prevailed in this instance, they now had evidence that the glass lid might not be a good idea and pulled it off the market.

Another group of examples of contributory negligence would charge that the plaintiff ignored a warning. A third group might involve the plaintiff not attending to information on displays, taking an incorrect control action, or taking the action too soon or too late for the purpose. Human factors experts might be called in to determine whether it was reasonable to expect the operator of the device to see, hear, or do something differently that would have prevented the accident. All told, there are reasonable and unreasonable expectations from the human operator. Even a successful legal defense is costly, and designers are advised to idiot-proof their products to the extent humanly possible.

Negligence

A negligence suit is generally one that is aimed not at the manufacturer, but against a third party that might have provided a service or a piece of equipment for the plaintiff or contrary to the interests of the plaintiff. In these cases, the defendant might have operated a piece of equipment improperly, or rented it to the plaintiff in a state of disrepair. The outcome of such cases depends greatly on whether the defendant was providing a service for hire or in some way profiting using the equipment in question.

Another type of negligence suit might occur in an employment context where the employer disregarded safety appointments for the equipment or instituted some other work procedure that compromised safety. Under current law, employees who are injured on the job are entitled to compensation for medical expenses and lost wages from the

Workman's Compensation agency. According to the regulations, however, the employer cannot be sued, although OSHA will investigate and levy fines as necessary. Injured employees who believe their compensation from Workman's Compensation is inadequate, however, might file a suit against the equipment manufacturer for an alleged design flaw; the defendant manufacturer then has the burden of responsibility to shift the blame onto the employer or employee.

Proving the Case

States differ in their statutes, but Wisconsin serves as a representative example for present purposes. The plaintiff's case consists of two parts: the contributions of the parties to the events that led to the liability, and the monetary valuation of the damages. The jury's task is to determine the percentage of fault associated with each party's negligence or liability in the matter. If the plaintiff is not successful in convincing the jury that the defendant was responsible for at least 50% of the fault, the plaintiff receives no award from the court. If the plaintiff can establish that the defendant's fault is greater than 50%, the award is equal to the amount of damages times the percentage for the defendant's fault. Admittedly, there is some subjectivity as to what constitutes a percentage in these types of decisions.

DISCUSSION QUESTION

Consider the following traffic accident. The defendant was driving her car at approximately the speed limit (35 mph) down a two-lane city street. No cars were parked on the driver's side of the street at the time the accident occurred. The car proceeded through a lighted intersection and struck a 12-year-old boy at a point approximately 150 feet past the intersection. The car dragged the boy 41 feet before stopping. Once the car stopped the driver backed up and drove over the boy again. Most of the boy's injuries were to his leg. Damages were set at $100,000. What human factors issues should be considered to determine whether the driver was able to stop the car prior to striking the boy? Reconsider your answer to this question after reading a few more chapters.

2

Elements of Human Factors Analysis

This chapter considers elements of HFE that are fundamental to the field. They are the allocation of function between person and machine, human error and system reliability, communications and information, usability testing, and some principles of nonlinear dynamic systems theory (NDS) that describe different ways in which system events can change over time.

Allocation of Function

The first design question after establishing the goals of the system itself is to determine what functions should be assigned to the machine, and what functions should be assigned to the human. Often, a new person–machine system is designed to replace a strictly human action by a combination of human and machine actions. The importance of the goals should not be underestimated. Designers should have a clear idea of what the intended users are trying to do, how many varieties of work they need to do, and what types of discretionary controls the users wish to have. It often helps greatly if some of the people contributing to the system design were actually experts at the task themselves.

User Population

Satisfactory definitions of system goals depend on a clear knowledge of the user population. The same can be said for virtually every aspect of the interface design. Typical questions the designers should ask are:

Is this intended for an occupationally specific population, or is it intended for a broader slice of the general population?

Do the users have normal-range physical capabilities, or do they have specific limitations or handicaps?

What is the typical occupational training or educational level of the user?

Are the users already skilled in tasks relevant to the new person–machine system, or are the users seeking to acquire an artificial skill through the new system?

Given that they have normal-range physical and mental abilities, what abilities do they wish to extend through use of the new system?

Does the user population vary with respect to its mix of experts and novices? If so, how do experts and novices differ in their expectations for the system or their approach to using it? The differences between experts and novices are often relative to the system or group of tasks that the system is supposed to facilitate. They tend to fall into at least five categories, nonetheless. First, at the level of commands and operations, experts may have developed "population stereotypes" concerning how the system should operate. "I should be able to find the X control [here], and it should allow me to do Y if I select Z."

Population stereotypes are considered further in Chapter 7. Designers of relatively new systems should thus ask: "Can the users be expected to have experience with similar but different person–machine systems?"

Second, some features of systems could frustrate experts in ways that would not faze novices. As an example, some of the early versions of popular statistics programs did not offer some of the more advanced calculation options that experts might want to use. (Repeated measures analysis of variance was once one of them.) Novices might be content with default settings that might be reasonable to use in many situations, but ill-advised in others. In situations like these the system did not have enough capability to meet some of the operators' goals or subgoals.

Third, experts and novices also differ in the way they solve problems. The expert is likely to have developed some sense of probable cause-and-effect relationships. Given a problem, they would have some reasonable guesses as to the most likely causes, which aspects of the situation should be checked in what order, and which solution options should be eliminated and which ones are most likely to work.

Fourth, experts do not confine themselves to solving problems after they are identified. Experts can anticipate problems and look for clues that signify a critical situation in the making (Vicente, Mumaw, & Roth, 2004).

Fifth, experts can manage their workloads better than novices can do (Vicente, Mumaw, & Roth, 2004). Workload issues are pervasive in human factors as multiple tasks and different demands levels are routinely involved. Some tasks are better done sequentially, while others can be done simultaneously. The expert in any case develops strategies to get better results with the least amount of wasted effort.

What other tasks are the users performing in their daily work routine other than work that involves the new system? This is essentially a question of workload. A system might be the center of the operator's activity or it might pertain to tasks that are done less often or with less priority. A system might not appear to absorb too much mental workload if it is one that is done in isolation from other tasks, but if other things are going on that require a lot of attention, the secondary system should keep its demands simple and convenient.

Automation is often effective at reducing mental workload, but there are times when the pre-automation workload is comfortable, and the automation reduces the workload to an uncomfortably low level leading to boredom and reduced attention capacity (Young & Stanton, 2002). On the one hand, reducing the demand on one task frees up mental attention for a secondary task, but if secondary tasks are not available, the overall effort and arousal levels can decline to a critical point where human error increases. A related issue, however, is that the secondary task might not be a wise choice. There have been numerous initiatives during the last decade to automate aspects of automobile driving, such as monitoring the distance between the following car and the leading car in traffic, or the lateral position of a car in a traffic lane. Although the automatic function appears to work as intended, the mental capacity of the driver, which is often underutilized in normal driving, is diverted to nondriving tasks that are usually regarded as driver distractions (Carsten, Lai, Barnard, Jamson, & Merat, 2012).

The measurement of workload is, unfortunately, not straightforward, and something that is again relative to the task environment. The point is expanded further in Chapter 6. The U-shaped relationship between arousal or load and errors is considered further in Chapter 9.

Benefits of Machines

Machines are good for tasks or task components that involve repetitious actions that must be produced exactly the same way each time they occur. Machines are also preferred when large mechanical forces are involved. Humans typically get the job of setting the machine into motion, and choosing machine settings for specific outcomes, especially when the system allows some choices and flexibility in the matter. Humans are also good at making decisions when imprecise or incomplete information is available. A machine may be able to process information, but it can only do so if all the necessary starting information is supplied. Humans also repair the machines, and repair tasks typically require additional tools and machines.

Figure 2.1 depicts one of many automation success stories. Figure 2.1a shows a lathe (1960s design) that is used to make one-of-a-kind machine parts. The operator guides and controls the sequence of cutting motions through completion, a process that may take hours for each rendition of the manufactured object. Figure 2.1b shows the control panel for a contemporary version of a lathe that has the same basic system goals. This time, the operator can program the whole sequence of cutting motions, set the zero-point in the

(a)

(b)

(c)

FIGURE 2.1
(a) Semi-manual machine lathe, (b) controller for a programmable lathe, and (c) blueprint and code sheet.

cutting chamber (not shown), and then just push the start button. The information can be stored to make more copies of the same object if desired, and the cutting process is much faster overall. Figure 2.1c shows what the operator still needs to do to translate a blueprint into a sequence of machine codes that will be programmed into the machine.

After sketching a prototype of the new person–machine system, designers should consider the impact of the system on the operators' overall mental and physical workload. The operators should be left with meaningful tasks and a sufficient stimulation level. Similarly, mental overload should be avoided. Unfortunately, many technologies have failed in this regard.

Both lathes in Figure 2.1 require a substantial amount of skill to operate, although the skill composition is different. Some forms of automation, however, have the effect of deskilling the task overall and are deliberately introduced as a means of cutting some of the human costs of production (Hammer & Champy, 1986). In those cases, the liability to the system overall is that there might not be anyone around who knows what to do when the automation fails or recognizes the situation while it is happening. Thus, training programs that accompany automated systems should teach operators much more than what buttons to push when things go well.

Flexible Allocation of Function

Scerbo (2001) distinguished two classes of flexibility in automated systems, adaptable, and adaptive. In *adaptable* systems, the operator makes some selection for what functions will fall under user control and which ones will be carried out automatically. There is typically a dialog sequence at the start of an operation where the choices are registered. It is usually possible to change the choices when desired.

In *adaptive* systems, however, the machine makes the choice as to whether to put a function under automation or returning it to manual control. The system senses elevated workload by using EEG signals or other biometric information and goes into automated mode, thereby relieving the operator of some of the demand (Schmorrow & Stanney, 2008). The automated mode turns off when workload returns to normal. The switching between automated and normal modes makes sense if the human performs better than the machine under normal circumstances. One design challenge is to determine the point where the mode switches are most advantageous, which has to be worked out in laboratory studies that resemble the real-world task involvement as closely as possible. Another challenge is to properly alert the operator that the system has gone into automatic mode, which means changing the operator's expectations of what the machine will accomplish during a particular sequence of activities. Changing the operator's mental model of the work environment involves mental switching costs that need to be taken into consideration as well (Reinerman-Jones, Taylor, Sprouse, Barber, & Houston, 2011). Brain-based adaptive systems have made significant improvements on the performance of primary and secondary tasks (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006).

Trust in Automation

The classic contributions of machines have been to reduce the amount of physical force required, to enhance speeded precision in production, and to extend human sensory capabilities. The advent of artificially intelligent software (Chapter 12) led to new vistas in automation whereby human thought processes were taken over by the machine; the need for rapid, accurate, and repeated calculations led to numerous successes before artificial intelligence per se was introduced. Other types of decisions and psychomotor control are becoming targets for automation as well. There is a tendency among machine-centered designers to reason that if 80% or more of accidents or unwanted events are the result of human error, all those errors will go away if the human in the system can be automated. Unfortunately, automation might only move the errors around the system rather than eradicate them. Automation can change the nature of the task to make the task more programmable, but in doing so could make the task incompatible with the operator's expectations for proper functioning (Papantonopoulos & Salvendy, 2008) and the means of monitoring it (Weiner & Curry, 1980).

Automation can intervene between the human and system operation at four different stages of information processing: information acquisition, information analysis, decision making, and action (Parasuraman, Sheridan, & Wickens, 2000; Parasuraman & Wickens, 2008). Each of the four types of function intrudes on the operator's decision schemata to different extents. Automation that maximizes convenience and organizing information into a form that the operator can use right away to pursue system goals is relatively welcome so long as the information is sufficiently accurate. Otherwise the operator needs to invoke new rules for when to believe the information and when to take it with the proverbial grain of salt. Automation that sorts information by priority conditions can also be a work-saver, so long as the system is using the same rules that the human would use. Systems that use decision rules that are not clear to the operator defeat the operator's ability to qualify the results and produce conflict with the operator's better judgment are likely to be very frustrating and possibly more trouble than they are worth.

Any of the foregoing sources of inaccuracy in automated process undermine trust in automation and keep the operator oriented toward maintaining system control and delegating less to the automatic process. There have been several new initiatives to introduce new kinds of automation into high stakes conditions with high workload, such as air traffic control. Controllers who evaluated one particular new system were divided between those who felt they benefited from the new automation components so that they could attend to the biggest problems that required their judgment, and those who thought the overall risks were too high to trust the machine too far (Cho & Histon, 2012).

Knowledge of the unreliability of a system is a factor in the operator's response to the incoming information. For example, firefighters worked through a computer simulation of a fire response scenario wherein they received information that changed over time and made decisions similar to their real-world decisions (Omodei, McLennan, & Wearing, 2005). In one experimental condition, the participants were informed that the information they were receiving was highly reliable. In another condition, the participants were informed that they could not always be sure of the reliability of the information they were receiving. The participants in the high reliability condition spent more time analyzing the information and lost time making optimal responses. Those in the less reliable condition moved more quickly to their own better judgment and ended up making better decisions overall.

In contrast, however, less expert users of automated systems tend to trust the computer program even if it is wrong and a human advisor tells them so (Madhavan & Wiegman, 2007). This *complacency* effect, which is the tendency not to question the automatic system, arises from the user's self-assurance or inexperience with the task, excessive workload, fatigue, or insufficient communication (Bailey et al., 2006; Singh, Molloy, & Parasuraman, 1993). To complicate the matter further, if a system has multiple automated components and some of the components are more reliable than others, operators apply a global evaluation to the reliability of the system, and thereby ignore the specific sources of unreliability (Rice & Geels, 2010). This phenomenon can be reasonably traced to a workload problem: How many different evaluations and responses to a system's unreliability can an operator handle and still do the actual work?

Another source of complacency arising from the biasing of errors toward misses and false alarms plays an important role. False alarms are more often tolerated if the cost is light compared with a disaster caused by a miss (Parasuraman & Wickens, 2008). The trade-off between the two types of errors is considered further in Chapter 3 in strictly human contexts but it carries over to the automation and complacency when *other* humans—the designers—make the same trade-off.

Other hidden humans could be contributing to the automation-complacency dilemma as well. Trains are highly automated, although humans can override the automatic process. The automatic process has its limitations, however, as it is not programmed to respond to unusual circumstances such as obstructions or ice on the tracks; the operator would need to override the automation to avoid an incident. Sheridan (2002) recounted a situation where two commuter trains crashed with fatalities and serious injuries resulting because the operator did not override the automation when he saw his train was not slowing as expected when approaching a station. Ice on the tracks foiled the automation. When asked why he did not use the manual override, the operator reported that, according to his understanding, he was working under strict orders from management to maintain automated control (p. 30).

Human Error and System Reliability

There are five common types of human error. Errors of *commission* occur when the operator intends to take an action that needed to be taken, but selected the wrong action or pushed the wrong button. Errors of *omission* occur when the operator fails to take a needed action ("I just put your engine back together, but I had a couple parts left over. Do you want them for anything?"). In the same grouping, *extraneous acts* are those actions that the operator takes when doing nothing would have been the desirable response ("If it ain't broke, don't fix it").

Sometimes *when* is as important as *what. Sequential errors* are actions that are taken in the wrong order. *Timing errors* are those that are taken too soon or too late.

Mode error is the latest addition to the common types of error. Complex control systems are often designed so that a set of buttons works one way in one mode, but the same set takes on other functions in a different mode. Engaging the correct mode is a separate control action and another opportunity for error. Mode changes often result from automated functions that turn themselves on and off. In the case of airline pilots, they are usually absorbed in monitoring the actual behaviors of the aircraft such as speed and altitude that they overlook indicators of mode (Sarter, 2008). This is only one example of automation attempting to reduce workload but only succeeding in moving it around, or shifting it to other people.

Some advisements concerning types of error have come to the surface. One is that the taxonomy of errors found in one industry probably do not transfer well to other industries (O'Connor, O'Dea, & Melton, 2007). Furthermore, in a complex system, attempts to minimize a type of error might only have the effect of changing it from one form into another. Thus, the *context* of the errors is as important as the errors themselves for defining effective control solutions (Renshaw & Wiggins, 2007, p. 201). For instance, in the context of aviation mishaps involving night vision glasses, Renshaw and Wiggins (2007) included several questions in their incident reports: Describe the preconditions and outcomes of the incident along with a checklist of common features; a checklist of organizational factors such as command and training; and a ranking of the relative importance of risk present, uncertainty or unpredictability of the situation, time pressure, and multiple tasks or workload at the time. In the context of naval underwater diving incidents, O'Connor et al. (2007) developed an extensive list of incident features in the categories of situation awareness, appropriateness of decision making, communications, stress and fatigue, and supervision or leadership contributions. Several of the foregoing categories of events are considered later in this book.

Error Probability

The human error probability (HEP) is simply the ratio of actual errors to the number of opportunities for error:

$$\text{HEP} = \frac{\text{Number of Errors}}{\text{Opportunities}} \tag{2.1}$$

Reliability is the additive opposite, or 1 – HEP.

Some examples of HEPs from Adams (1985) would include the following: The HEP for selecting a wrong control from a group of identical controls could range from .001 to .01. The HEP for turning a poorly designed control in the wrong direction could range from .1 to .9. The HEP for operating a sequence of valves 10 times in a row could range from .001 to .05. The HEP for not recognizing a status indicator that is located right under the operator's nose could range from .005 to .05. Some of these HEPs look large at face value whereas others may look very small. But if an action is taken 20 times a day or a 100 times a week, the odds of something bad happening become very strong.

Redundancy

One way to improve the reliability of a system is to introduce redundancy in the operation. For instance, two or more operators may have to agree on an action before it is taken. In another type of example, two or more indicators from machine displays must be present before the operator takes the action. The effect of redundancy in a system can be stated as

$$R_{xx} = 1 - (1 - r)^c \tag{2.2}$$

where R_{xx} is the reliability of the total system, r is the reliability of each system component, and c is the number of redundancies in the system (Kantowitz & Sorkin, 1983). Equation 2.2 assumes for convenience that all redundant components have the same individual reliability, but there is no need to require that real-world system components be equally reliable.

The disadvantage of building redundancy into the system is that redundancy slows the system down. It might also require more personnel, which is a labor cost. Some operations consultants (e.g., Hammer & Champy, 1986) argued in favor of removing redundancies in work systems as part of their "reengineering of the organization." They did not address the positive effect of redundancy outright, but a wise systems engineer should consider whether some of the redundancies are really necessary. Sometimes a new machine system can be introduced that is more reliable than two or more humans in an old system together.

Usability Testing

Product manufacturers are strongly advised to incorporate HFE principles in the early stages of product development. To paraphrase an old saying, an ounce of problem prevention is worth a pound of retrofit redesign. There are three basic stages to usability analysis and testing: preparation, iterative laboratory testing, and field testing.

Preparation

At this stage, the designers should establish the goals of the product. Much of the information that guides them in these decisions is based on marketing analysis. Too often, the person–machine interface looks good, but is not as functional as it appears.

Defining the user populations and the allocation of function are also parts of the preparation stage. These particular points were covered earlier in the chapter.

Iterative Laboratory Testing

At this stage, the designers make up a prototype interface with the controls and displays organized in a sensible fashion. If there is no clear answer in the HFE literature about the design of a feature for an interface, the best response is to think through some options, make up a few different prototypes, and conduct an experiment. The experiment would contain a small but suitable number of probable users who are given a task to perform using one or more prototypes. Participants would be assigned to prototypes in a randomized or counterbalanced fashion. They would perform a standard task in which one or more aspects of performance are measured. Indicators of speed and accuracy are the common choices in this phase of the research. The ensuing statistical analysis should show that one prototype is better than the others.

The design of the keypad on a touch-tone telephone was a case of using experimental analysis to determine optimal design (Deininger, 1960). Previously there was no known rule for a good keypad layout, and designers could only guess what the possibilities might be, so they designed several of them. The experimental task required users to enter phone numbers accurately. The results of the experiment are on everyone's desk, wall, or pocket now.

There are other issues beyond raw speed and accuracy. If operation speed of the person–machine system appears slow, there may be a way to revise the workload of the human operator. Speed is relative, of course, to particular system requirements. Advantages in speed could be determined simply by comparing the new device with other devices that have similar purposes or designs.

Safety issues may be involved. Can the operator possibly activate the wrong control to produce a disaster? An investigation into the nature of the operators' errors can produce a great deal of useful information.

This stage of the process is called *iterative* because it may occur more than once before a product design is ready for field testing. An adaptation of the iterative design and testing process was known as *rapid prototyping* in the software industry a couple decades ago. Part of what made the process rapid is that software is often more readily changed than interfaces made out of hard materials. For instance, some control-display sequences in computer programs can be designed, improved, and finalized in a matter of days as long as there are no conflicting demands on the programmer's time and the human participants are standing by ready to try the new design. Hard systems, in contrast, may require months to design and test an interface. Thus, greater forethought in the initial design will shorten the sequence of iterative tests.

Field Testing

Laboratory studies are explicitly designed to vary some variables and control others. The real world is replete with uncontrolled variables and tasks that might not have been anticipated in the laboratory phase.

The field-testing phase is also known as *beta testing* in the software industry. Often enough, beta testing is intended to see if the programming actually works; the more basic HFE aspects of control and display design are secondary goals because the main issues should have worked out before reaching this stage of the process. Nonetheless, some controlled monitoring of users and their experiences in their natural work settings should occur at this phase before the final product is released.

The iterative design and field-testing processes are regrettably subject to some abuse. Because of the ease with which user input can be collected and assimilated in software design, we end up with software products being released for sale, and beta tested at the same time. The market then sees a myriad of versions and upgrades leading one to question, "Why couldn't you get it right the first time?" Sometimes, it is difficult for a user to distinguish a visible change between an upgrade and a downgrade. Sometimes the changes are not visible at all; one can legitimately question whether the upgrade is actually installing something necessary or useful, or whether it is just an invasive probe.

Technical Manuals

The information about user interactions can serve additional purposes beyond system design and upgrades. The user manual is one purpose. Cost–benefit analysis and planning for future redesigns of the system are two others.

Technical manuals of the old school were often written at a reading level that represented several years of education beyond that of the actual user (Chapanis, 1975). One common fault is that the texts often devoted considerable space to explaining how the machine worked from the machine's point of view, with user instructions interspersed at critical points. This format makes the user's experience a frustrating hunt for critical information.

A better format would separate the text into two documents. One document would contain the technical information that people might use if their goal was maintenance and repair. The other document would describe how the machine is organized, identify the controls and displays, and then proceed with operations instructions. Operations instructions should begin with simple and popular tasks, then move on to the more exotic capabilities of the system. Note the emphasis here on two user populations for text documents.

Online help facilities for computer programs have been convenient and popular. They work best when the users know what they want to do and can at least guess at the searchable keywords for finding an instruction. A natural language processor, which the user does not actually see, can translate a user-defined question into one or more possible questions that the system knows how to answer; see Chapter 12 for an elaboration of this point. The format of most searchable help facilities usually does not give a user a good overview, however, of what the system's capabilities are or how they are organized. Separate files that are often called "tours" or "getting started" can fill in this void.

Note, however, that users exercise two broad strategies in their searches for information. One is the directed search for targeted information. The other is exploration to determine what information and new ideas are available.

Despite the convenience of online help for many purposes, not to mention the cost savings to the manufacturer, there is still a role for a printed technical or user's manual. "Getting started" means just that—advanced users have probably developed needs for the deeper or more challenging features of the system, which would overwhelm a new user. Furthermore, a user can be in the middle of a job with several screens open and might wish to look something up without having to move windows around or disturb the work layout.

Training *programs* extend well beyond a printed or online manual into a structured learning experience where the user has an opportunity to gain hands-on experience with the system features and capabilities. The program structure is typically organized from simple to complex, such that each new module builds on what was learned in previous modules.

Cost–Benefit Analysis

As its name implies, the cost–benefit analysis is a mental exercise for potential system users to determine whether there is a benefit to adopting a new system. Costs in the concrete financial sense figure heavily into the process. Comparisons may be made between new potential systems or against a present system. Typical questions that might be asked at this stage would be the following.

Does the new system get the job done faster than the present system? Faster in this context often means delivering a manufacturer's product to the end purchasers faster so that the chain of payments is faster also. Laboratory analyses of control-display sequences often have the potential to signal reductions in operation time by a matter of seconds. Time savings on this scale might translate into saved labor costs, but in many cases the savings are in user satisfaction, which is a subjective experience and not readily quantifiable in the same manner as gross labor costs.

Does the new system replace a manually driven system or a computer-based system that reuses information that is time consuming to produce? Word processors replaced the conventional typewriter for this basic reason. Computer-driven metal lathes can reduce a task time from 2 weeks to 12 hr in some cases, and reduce the time to make the object a second time to significantly fewer hours because the basic instruction program can be saved. Machine time is expensive too.

What is the training and start-up time? A new system may be ideal for a user or user organization that has never used a version of the system before. The decision to adopt a particular system might be predicated on training time. A user with an overhead in one particular system is now considering whether the downtime for retraining is worth the effort. The potential for substituting higher skill, higher cost labor with lower skill, lower cost labor would factor into the decision.

How many different system goals are served by the new system? In the early days of desktop computers, tasks that involved intensive data analysis favored the mainframe computer over the little desktops by a time savings of 100 to 1. (A colleague of mine timed a representative job of his using a 286-series desktop and a university mainframe at midnight that he could access through a phone line.) The pretty printing that everyone has come to expect from a word processor, however, favored the desktop; mainframes were usually equipped with only rudimentary typographic capabilities. Small businesses could obtain a significant advantage by converting their operation to a desktop that was much less expensive than a mainframe or mainframe service. Businesses with large computational needs had the opposite experience.

Compatibility has become another issue. Lange, Oliva, and McDade (2000) compared the market uptake for two products that had similar goals. Market growth for each remained relatively equal for a number of years until one of the manufacturers expanded its offerings in system-compatible products. Other software manufacturers then jumped on the bandwagon by making their products compatible with that manufacturer's items.

System Reengineering

Hammer and Champy (1996) extolled the virtues of adopting new technologies that could improve an end user's financial experience by eliminating redundancy and substituting low-skill labor for high-skill labor. The benefit of redundancy as a means of improving system reliability was mentioned earlier. The downside of skill and knowledge substitution is discussed in Chapter 12 in the context of expert systems.

On the other hand, there are positive stories to tell. In the home construction industry, there is an expediter whose job is to coordinate numerous contractors on numerous sites all in the correct sequence. The expediter also needs to inspect finished phases of the work and to go to the sites at critical moments when something unplanned happens. Before the days of cell phones and laptop computers, the expediter had to make a series of phone calls from a hardwired phone in an office and physically drive out to sites where phones did not reach. With the new equipment, however, the expediter's office is now in his truck. Sites can be seen, calls can be made, information files are always handy, delays are reduced, final products are delivered sooner and with less overhead overall.

Sometimes an organization's rate of innovation outpaces the decline of older technologies. As new features of the technology (e.g., a telephone system) are added, new links have to be made among parts of the system. As new features are requested by users, the technology becomes brittle, meaning that a small change requires extensive relinking to many parts of the system so that the system can stay internally coordinated. At some point, the technology reaches a crisis point where it becomes more sensible to rewrite the big program from scratch. The workload demands for doing so can be extensive. At the same time, the pressure to change the system from scratch is closely tied to the organization's dwindling knowledge about how the product was initially assembled (Staudenmeyer & Lawless, 1999). Fortunately for one organization with this type of problem, there were a few people remaining in the organization who still remembered how to manipulate the original program code, all 15 million lines of it, after 20 years. The remaining problem was how to shut down the existing system while the new system was being installed. Staudenmeyer and Lawless (1999) named this type of barrier to technological change *socio-technical embeddedness*.

Communication, Information, and Entropy

The Shannon and Weaver (1949) communication model was first developed at Bell Labs as a concept for improving the quality of telephone communications. It can reasonably apply to any communications medium, however. The communications model was the outgrowth of another important conceptual development from the same source that pertained to the quantification of information.

Communication Model

The communication process (Figure 2.2) is conceptualized as a five-stage process. Errors in communication could occur at any stage. The first stage is the sender, who should be speaking the correct words clearly into the phone. In the second stage, the speaker's voice is encoded into electronic impulses that are used by the phone system. In the third stage, the message travels across telephone lines or telephone cells (or other airwaves). Noise, which can produce signal distortion, can possibly enter at this stage and is the target of many technological innovations. The signal is decoded into sound again at the receiver's telephone. The final stage is the receiver who must interpret what is heard the right way. As one might imagine, a good deal of human miscommunication can occur without the assistance of electronic devices.

There are three aspects of noise that are of importance to HFE. One aspect is the stray or unwanted signal component that obviates the signals that are intended as the real message. This aspect of noise is considered in Chapter 3 on psychophysics.

The second aspect of noise is the high-volume environmental noise. Not only can it obscure signals, it can also lead to hearing loss when operators are exposed to large quantities of it for prolonged periods of time. This aspect of noise is considered in Chapter 4 on auditory displays.

The third aspect of noise pertains to the psychological stress that it can produce. The effects of stress may be health related or performance related, even where the tasks are not primarily auditory. This aspect of noise is considered in Chapter 9 on stress and human performance.

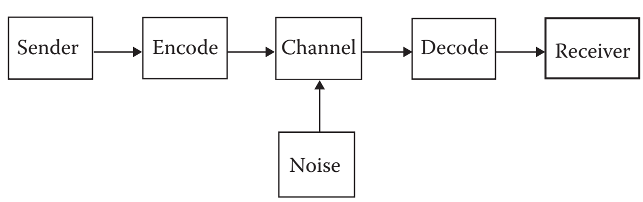

FIGURE 2.2
The Shannon–Weaver communications model.

Quantifying Information

A bit of information is the amount of information required to make a binary decision. Given that a system can take on two states, A and B, only one bit of information is required to know whether the state is A or B.

If the system can take on N possible states, we would need $N - 1$ bits of information to know which of the states is taking place. Each bit of information is quantified as 0 or 1. If the bits associated with states 1 through $N - 1$ are quantified as 0, then it follows that state N is taking place.

If we do not have the information that is required to predict the state of the system, then we have uncertainty or *entropy* in its place. Shannon (1948) and Ott, Sauer, and Yorke (1994) defined entropy (H_s) in Equation 2.3, where p_i is the probability associated with one categorical outcome in a set of r categories:

$$H_s = \sum_{i=1}^{r} p_i \ln(1/p_i). \tag{2.3}$$

Entropy is greatest when the odds of system states are equal, and it is less than maximum when the a priori odds of a state are unequal.

Entropy

The construct of entropy has undergone some important changes since it was introduced in the late 19th century. Shannon entropy (Equation 2.3) is actually the third paradigm of the construct to be introduced out of four. In its first historical epoch, the concept of entropy was introduced by Clausius in 1863 in conjunction with relationships among pressure, volume of a container, and temperature along with broader problems in heat transfer. Entropy thus became a concept of *heat loss* or *unavailable energy* (Ben-Naim, 2008), which gave rise to the second law of thermodynamics, which become controvertible a century later. According to the second law, heat loss is inevitable in any system, so that the eventual outcome is “heat death.” In this scenario, the system (if not the entire universe) will gravitate toward an equilibrium point where, in essence, nothing more will happen once the heat energy has been expended, and the system is completely disordered.

The second paradigm of entropy was associated with Boltzmann’s contributions to statistical mechanics, which concerned the movement of gas molecules throughout a container space. Although it is not possible to follow the movement of every molecule, it is possible to give an average value for the molecules in a particular condition. In one extreme we have the condition of absolute zero temperature where molecules of a gas stand still; in the other extreme we have Brownian motion, where molecules are bouncing randomly and completely filling the container’s space. This paradigm led to the concept of entropy as being one of *order versus disorder*.

The third paradigm of entropy was Shannon’s, which introduced the quantifiable construct of information. Information, which is not always completely available, is used to predict the motions of the system. Note that in the first three paradigms, entropy is motion that is induced by other system states such as heat, and detected by an outside observer who wants to predict system states. The perspective changes in the fourth paradigm that originated with nonlinear dynamic systems theory (NDS) and principles of self-organization

(Nicolis & Prigogine, 1989) contradicts the heat death scenario: Rather than dissipating energy, the system reorganizes itself—without any assistance from outside agents—to produce a more efficient use of its energy. Self-organizing systems are expanded in the next section of this chapter.

The metrics for quantifying entropy in the NDS paradigm often build on Equation 2.3 in some way (Guastello & Gregson, 2011). Although the full range of the metrics falls outside the scope of what human factors and ergonomics have been able to uptake to date, one important feature is worth remembering. Shannon entropy only quantifies the variation in elementary states; it does not account for *patterns of states* over time. For these purposes, constructs such as *approximate entropy*, *sample entropy*, and *topological entropy*, which are closely related to other phenomena in nonlinear time series analysis, have been developed. Symbolic dynamics analysis have been useful for identifying patterns in task switching, and it was in turn possible to apply entropy statistics to the patterns that were extracted.

Some basic constructs of NDS are considered next, and applications will pop up in various places throughout the text. Various concepts of complexity are introduced here and again in Chapter 13. Entropy constructs resurface again in Chapters 6 and 9.

System Events Change over Time

Perception of Change

One of the critical links between human and animal perception and their ability to survive is their ability to detect change either over spatial location, over time, or both (Gibson, 1979). When a machine system renders information about change in the environment, we can speak of an operator's *situation awareness* (Endsley, 1995), which involves several levels of depth in perception and cognition: (a) Did the operator notice that an event occurred such as a warning light going on or off, or objects entering or moving around the display screen? (b) Did the operator interpret any change in meaning associated with the perceived differences? (c) Did either the change in stimuli or change in meaning indicate a change in response? (d) Did any of the foregoing aspects of change also produce a change in the perceived state of the system?

Note that there are two things going on here: One is the more obvious difference in levels of processing. The other is the translation of change at one level of processing into change at another level. The former can be relatively fast, while the other is relatively slow. A critical level change is required at the lower level to promote a change at the higher level.

Information, as HFE tends to experience it, has different relationships to time. In the simplest relationship, the information is static; the information item remains the same, that is, stable until perhaps a control action has been taken or a noticeable change has occurred to denote a new stable state. Such a stable state might occur if a status light indicates "not ready for action," until such time as the light changes to mean "ready for action." Even here, however, something changes when the light goes on and off.

In other situations, the information is changing over time much more vibrantly. Sometimes the change over time is a deliberate contrivance of a mode of communication; Morse code, which went out of fashion at least a half century ago, is an example. The auditory signals convey all their meaning in terms of the duration of pulses that are heard over time, and the pattern of long and short pulses. Although still interesting in some ways, deliberate time-phased signals are not the central concern here. Rather the concern is for the changing events for which one must decide whether meaning exists; such decisions require some effort if not also some well-developed intuition.

In conventional thinking, a discernible pattern of system behavior is regarded as a *deterministic* process. If no discernible pattern exists then it is regarded as a *stochastic* process, which is to say a random event. Furthermore, the behavioral sciences have typically treated change as only one type of event. This is where NDS makes its contribution. What might have been interpreted as a stochastic process is often a deterministic process that only appears random over time. The changes over time can be captured by simple equations. The trick is to determine what the relevant equations might be. In NDS many different patterns of change over time are possible, some of which are more complex than others (Guastello & Liebovitch, 2009).

What Is Random?

There is a tendency among behavioral scientists to jump to the conclusion that random processes are normally distributed or vice versa. Actually, one is neither necessary nor sufficient to conclude the other. Random processes have a rectangular distribution such that all the system's states, or values of a variable, are equally likely, and one does not know which state or value will pop up at a given moment. When there are multiple random processes taking place, however, the concatenation of rectangular distributions produces a normal (Gaussian) distribution. Thus, a normal distribution can result from many random processes. However, it is also possible to predict normally distributed variables from knowledge of other variables that are also normally distributed. Social scientists and others do it all the time, and when they do, it is possible to conclude that the dependent measure is not so random after all because it can be predicted to *some* degree of accuracy.

The myth of the normal curve once led to the interpretation of poisson distributions in accident data as prima facie evidence of accident proneness. Eventually, it became clear that it was just a distribution and did not result necessarily from accident-prone individuals or deviant situations. Yet accidents can also be predicted with some degree of accuracy (Chapter 10).

NDS functions tend to involve other types of distributions, which are most often the power law, exponential, and complex exponential distribution. In fact every differentiable process or function has a unique distribution within the exponential family. Thus, statistical distributions do tell us something of the processes involved that produce the data. Because error usually exists to some extent, prediction is not perfect. Yet knowing that the relationship between variables is a function other than a straight line, together with other variables that go into the function can greatly enhance the prediction of a process (Guastello & Gregson, 2011).

So what is random? The medieval meaning of the word is perhaps the most applicable today: the motion of a horse that the rider cannot predict (Mandelbrot, 1983, p. 201). However, just because the rider cannot predict the motion, it does not mean the horse does not have perfect control (Guastello & Liebovitch, p. 2). Some basic NDS processes are considered next.

Simple Attractors

Attractors are spatial structures that characterize the motion of points when they enter the space. Although they exist in different structural varieties, the fixed-point type is widely applicable. Collectively, attractors represent patterns of events that were known in the past as "equilibria." In the fixed-point attractor (Figure 2.3), a point may enter the space, and once it does, it remains at a fixed point. Other points that are indefinitely close to the epicenter of the attractor are pulled into the attractor space. Points traveling outside the boundaries (or *basin*) of the attractor are not pulled into the attractor region. There are other types of attractors beyond the fixed point, but Figure 2.3 makes the case for the prototypic attractor. Fixed-point attractors can be taken for granted if they are not considered in a context with other more volatile change processes.

Fixed points might not appear especially interesting, but try to move a system out of one. For that challenge we need bifurcations and at least one other stable state to which the system may gravitate. A definition of *stability* would be helpful as well: A system is stable if all the points in the dynamic field are behaving according to the same rule. Thus, a fixed-point attractor would be stable, but not everything that is stable is a fixed point.

Oscillators are also structurally stable attractors, and they are often found in biological systems and economics. Human work systems that produce output on a regular basis can be characterized as oscillators also, although there is a tendency for operations managers to view the cycle as a whole and ascribe one output value to the whole work cycle and ignore the internal temporal dynamics. The behavior of oscillators (mostly in engineering contexts) is usually characterized by a regular sine wave function. Auditory signal waves, for example, are often superimposed on each other to make what we hear as complex sounds or multiple tones. Oscillators are commonly decomposed in signal processing using fast Fourier analysis, which produces an additive combination of signal functions that are occurring at different frequencies.

The analytic task becomes more challenging, however, when the oscillators interact. According to a theorem by Newhouse, Ruelle, and Takens (1978), three coupled oscillators are minimally sufficient to produce chaos (see below). Although not all combinations of three oscillators will do so, complex systems containing at least three coupled oscillators have the potential for chaotic behavior. Karwowski (2012) noted that because of the growing complexity of human–machine systems in virtually every economic sector, more examples of chaotic behavior should be anticipated.

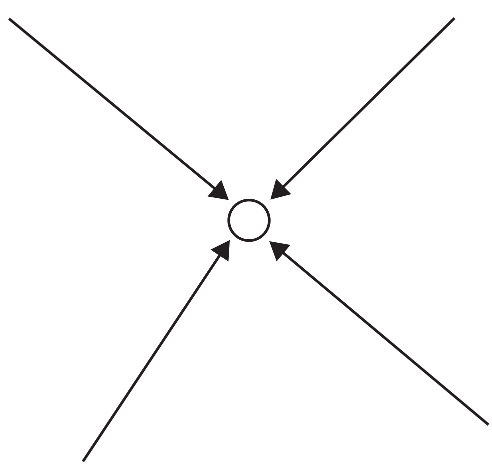

FIGURE 2.3
Fixed-point attractor.

Bifurcation

A *bifurcation* is a pattern of instability that is observed within a dynamic field. Bifurcations split a dynamic field into local subregions occupied by separate dynamics. One simple well-known bifurcation, the logistic map, changes a behavior pattern from a fixed-point attractor to a chaotic process, as the value of a *control parameter* changes. A control parameter is, in essence, an independent variable that is found in a dynamic system. Unlike the independent variables that are found in ordinary experiments and statistical analysis, control parameters within a system can be different from each other because of their different underlying functions, rather than simply their content. The logistic map is shown in Figure 2.4.

The equation for the logistic map is

$$X_2 = CX_1(1 - X_1), \tag{2.4}$$

where X starts with a value between 0 and 1. It is an iterative process, meaning that we start with a value of X_1, use it to calculate X_2, run X_2 through the same equation to produce X_3, and continue for many values of X_i. C is a control parameter. At low values of C, Equation 2.4 produces a series of X_i that denote a fixed-point attractor. When C becomes larger, X goes into oscillations. As C increases further, we start to see period doubling, or oscillations within oscillations. When C exceeds 3.6, the series of X becomes chaotic.

The logistic map has seen some useful applications in ecology and organizational behavior outside of any HFE applications. The concept of bifurcation is particularly relevant in catastrophe models for discontinuous change; catastrophe models have seen several useful applications in HFE.

A bifurcation may be as simple as a critical point, like the ones in the logistic map model, or it may be a continuous pattern of instability as found in the catastrophe models. Critical-point

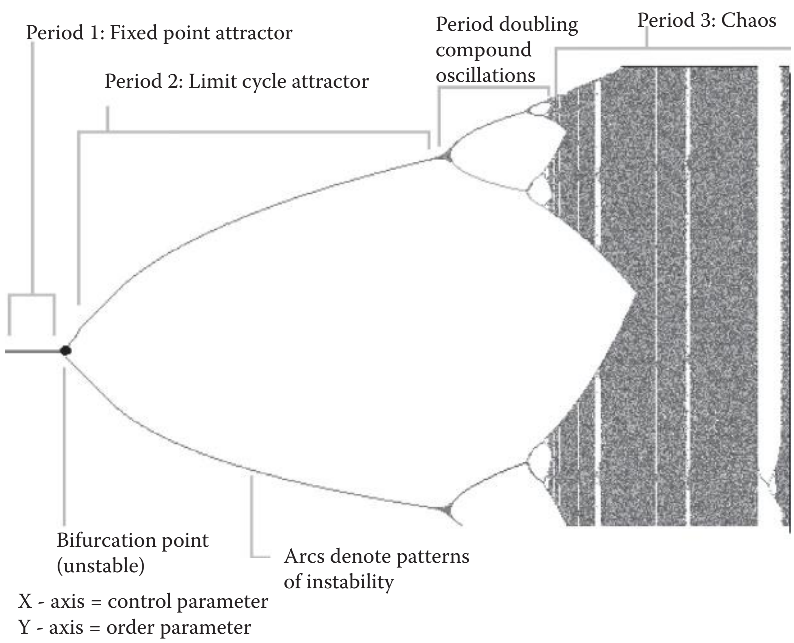

FIGURE 2.4
Logistic map.

bifurcations appear to be inherent in the automaticity function that occurs in cognitive processes (Chapter 6), or in the speed–accuracy trade-off (Chapter 8). There are more exotic and dramatic types of bifurcations, such as the *annihilation*, where two dynamic fields grow, then collide, then produce a very different dynamic field as a result. Urban sprawl and urban renewal comprise one of the few known real-world examples (Guastello, 1995). Fortunately, annihilations have not surfaced yet in HFE insofar as anyone has been able to determine.

Chaos

Chaos is a series of events that appear to be random in their appearance of values and incidences of occurrence, but are actually accountable by simple deterministic equations. In the ideal case, chaotic time series are nonrepeating, bounded, and sensitive to initial conditions (Kaplan & Glass, 1995; Sprott, 2003). Although chaotic behavior may appear random, chaos is ultimately a deterministic process, hence the initial fascination with the concept in the scientific community (Dore, 2009; Lorenz, 1963) and popular media (Dooley, 2009).

Boundedness means that the values of the observed variable stay within limits. Sensitivity to initial conditions means that if two points start arbitrarily close together, and each point is iterated to the same equation, then after a period of time (number of iterations) the two points become markedly further apart. The reader is encouraged to iterate some numbers through Equation 2.4 where $C = 4.0$ and see what happens.

A chaotic time series can result from an underlying attractor, but an attractor structure is not a requirement. In the case of a chaotic or "strange" attractor, the range of values for the observable variable would stay within boundary values, but between the boundary values it would exhibit the type of unpredictability associated with chaos. Ironically, a chaotic attractor is structurally stable, meaning that all points within its boundary are behaving according to the same (unpredictable) rule. Again, it is important to emphasize the "unpredictable" in this context is relative to conventional means of statistical prediction or conventional mental models of prediction that one uses. Within NDS "unpredictable" is better characterized as "nonrepeatable" (Guastello & Liebovitch, 2009).

One of the more visually appealing chaotic attractors appears in Figure 2.5. Chaotic behavior in real data is usually not as good looking. Figure 2.6 depicts chaotic data taken from an experiment in which two people were carrying on a conversation with electrodes

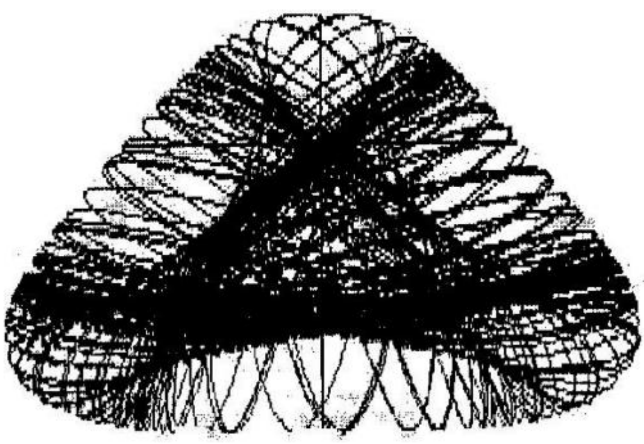

FIGURE 2.5
The Henon–Heiles chaotic attractor.

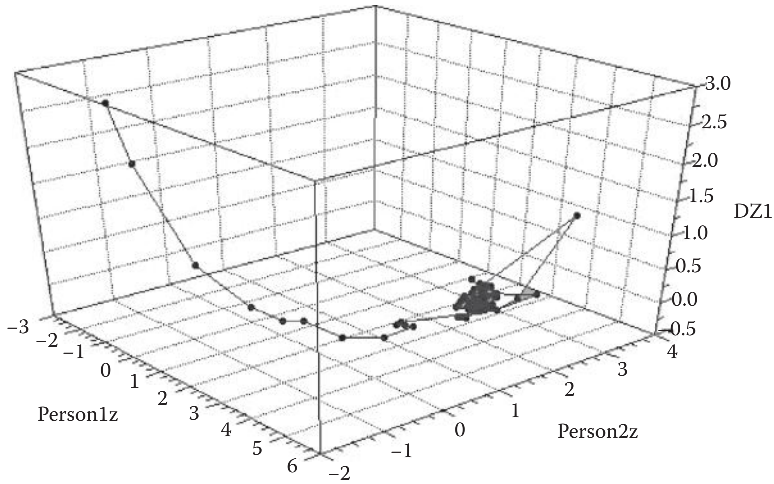

FIGURE 2.6
Chaotic time series for electrodermal responses from two people engaged in a conversation.

attached to their fingers to record electrodermal response (Guastello, Pincus, & Gunderson, 2006); the change in electrical conductivity of the skin is a signal of physiological excitement. The three axes of the plot are electrodermal response from person A at one point in time, electrodermal response from person B at the same point in time, and electrodermal response for person A taken 20 s later (vertical axis).

At present, there are at least five dozen formally defined chaotic mathematical systems (Sprott, 2003). Testing each and every one would be virtually impossible given that the behaviors in question could result from several possible chaotic systems and the control parameters that are known to drive the formally defined system might not be readily discernible in the real-world database. Thus, a simpler yet sufficient analytic strategy is to first determine whether the time series data reflect chaos at all, and at the same time ask a related question: How chaotic is the system? For these purposes, two types of metrics have been useful. One consists of entropy statistics which were described briefly earlier. The other is the Lyapunov exponent, which can be calculated directly on relatively noise-free data or statistically using nonlinear regression modeling (Guastello, 2009; Guastello & Gregson, 2011).

The concept nonetheless goes as follows: Chaotic time series exhibit both expansions and contractions over time. Sensitivity to initial conditions is, for the most part, an expansion. The Lyapunov exponent calculates sequential differences in the dependent measure over a short number of observations, then extrapolates those to produce a spectrum of values. In a chaotic system, the largest Lyapunov is a positive value, but there are negatives in the spectrum as well. A value of 0 is a perfect oscillator; slightly positive values indicate aperiodic oscillators, and slightly negative values indicate dampened oscillators. If only negative values are taken from a time series, they signify that the system is a fixed point. There is a relationship between the largest Lyapunov exponent and topological entropy under limiting conditions.

As mentioned earlier, humans must often deal with flows of information that appear irregularly over time often with the help of their display systems. People vary widely in their ability to track and respond to chaotic information, yet some *can* predict values from a chaotic sequence with a respectable degree of accuracy. The nature of this ability has not yet been determined. The widespread limitation has given rise to a class of intelligent displays and controls known as chaotic controllers, which are addressed in Chapter 7. Chaotic functions also show up in Chapter 13 where complex systems are considered in further detail.

Fractals

Another important property of chaotic attractors is that their outer rims, or basins, are fractal structures. A *fractal* is a geometric structure that is characterized by repetition of overall patterns at different levels of scale. Thus, if we were to zoom in on a small detail of the fractal, we would see a repetition of the overall shape in the enlarged area. This scaling property of fractals has led to some interesting questions such as whether the fractal nature of a time series of individual biometric data also appears at the level of group collective biometrics (Cooke et al., 2012); sometimes it does and sometimes it does not.

The dimension of a fractal is fractional, meaning that it does not occupy one of its spatial dimensions completely. Fractals have probably become famous for their pretty geometric designs as much as for their other characteristics. The geometric designs have not shown much direct applicability to HFE or human social behavior. One concept based on fractal theory that has had enormous impact on applications to organizational behavior and elsewhere is that of the *fractal dimension*. The fractal dimension is related to the Lyapunov exponent under limiting conditions; both have been used as measures of complexity within a time series. Despite their reticent role in HFE, fractals do connect the concepts of chaos and self-organization, which are more proximally related to HFE phenomena.

Self-Organization

Self-organization is a dynamic event whereby a system that is in a state of chaos creates its own order by forming feedback loops among its component parts. Although there are several mechanisms of self-organization, the feedback loops that transfer information are their common feature. The feedback itself may be positive or negative. It can be a constant flow, a periodic flow, or even chaotic over time. A general representation of a self-organized system appears in Figure 2.7. Note that both feedback and feedforward loops are possible.

The coupling of subsystems through information loops is the basis of the complexity in complexity theory when the latter is used in its restrictive form. (In its broader form, it is synonymous with chaos theory, which denotes the study of nonlinear dynamics, including chaos.) Additionally, the information flows are coupled with local rules by which agents or subsystems interact and information inflow is interpreted and transformed into a behavioral outflow.

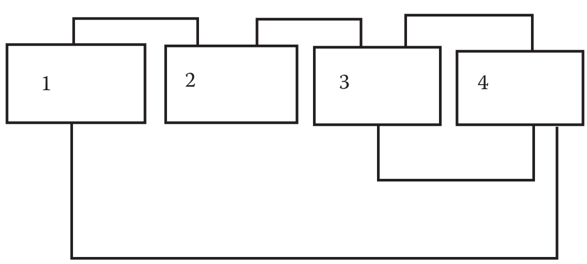

FIGURE 2.7
A set of feedback loops in a simple self-organized work process.

The spin-glass phenomenon serves as one example. If molten glass is subjected to high-intensity mixing, the molecules do not homogenize. Rather, molecules group together depending on the similarities of their electron spins; this principle has been studied in the context of living dynamic systems in a few different contexts (Kauffman, 1993; Sulis, 2008). In more broadly defined conditions, the mixing of agents produces hierarchical structures that serve to reduce entropy or disorder and keep the system apparently stable. The hierarchical structures often take the form of driver–slave relationships among the subsystems, such that the behavioral dynamics of one subsystem drive the dynamics of another (Haken, 1984, 1988). The key point is that the system increases its internal order spontaneously, and reduces its rate of lost potential energy. Within the context of human factors, an agent can be a human, a semiautonomous machine, or a compound agent consisting of both.

There is a line of thinking that complex adaptive systems toggle between states of chaos and self-organization (Waldrop, 1992). A chaotic state is one with a high potential for adaptation in many possible directions. A self-organized state is one where the organism has formed a pattern of behavior with lower entropy. Shocks from the environment or from internal sources, however, may propel the organization back into a chaotic state to prepare for an adaptation of a different type (Guastello, 2002). Some self-organizing processes are encountered in Chapter 13.

There are two sets of statistical footprints of a self-organized system. One set is the inverse power law relationship between the values of a variable X and the frequency distribution of X.

$$\text{Freq}(X) = aX^b \tag{2.5}$$

Equation 2.5 is saying that X has a power law distribution. If $b < 0$ and a noninteger (which it will be 99% of the time if we confine ourselves to two decimal places), then b is the fractal dimension of X, which in turn denotes that a self-organizing process is taking place (Bak, 1996). An example of an inverse (i.e., $b < 0$) power law distribution appears in Figure 2.8. Beware, however, that the presence of an inverse power law does not prove the

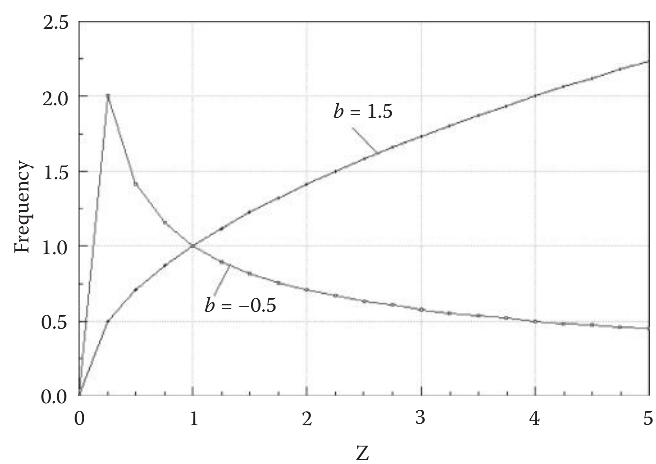

FIGURE 2.8
Power law distributions with $b = -0.5$ and 1.5.

existence of a self-organizing process. Functions such as Equation 2.5 often arise from self-organizing processes and are thus a signal to look further for the actual process that could have produced the function.

The second set of footprints for self-organization is the presence of a catastrophe model. Catastrophe models involve fixed-point attractors, multiple-system states, and bifurcations of varying levels of complexity.

Catastrophes

Catastrophes are discontinuous changes of events. According to the classification theorem (Thom, 1975), given a maximum of four control parameters, all discontinuous changes of events can be modeled by one of seven elementary topological models (with qualifications). The models describe change between (or among) qualitatively distinct forms for behavior. The elementary catastrophe models are hierarchical and vary in the complexity of the behavior spectra they encompass. Change in behavior is described by differential equations that represent the structure of the behavior spectrum, or *response surface*. The *cusp* response surface is three-dimensional and describes changes between two stable states of behavior, which are usually fixed-point attractors. The two attractors are separated by a bifurcation structure (manifold).

Movement of the system around its possible states is governed by two control parameters. The *asymmetry* parameter governs how close the system is to discontinuous change in behavior. The *bifurcation* parameter governs how large the change will be. In contrast, the unstable transients (represented in Figure 2.9 as the cusp point) represent behavioral regions of great instability and indeterminism. Some specific catastrophe models that concern stress and work performance appear in Chapter 9. Others that concern occupational accidents appear in Chapter 10.

Not all catastrophe functions are self-organized processes and not all self-organized processes can be framed as catastrophe models. There are enough occasions for overlap, however, to warrant at least a question of how catastrophe dynamics might produce a self-organized system or be the result of one. The general thinking here is that a self-organized process is often evidenced by a dramatic shift in the system's behavior, visual appearance, or internal organization. The shift is analogous to phase shifts among gasses, liquids, and

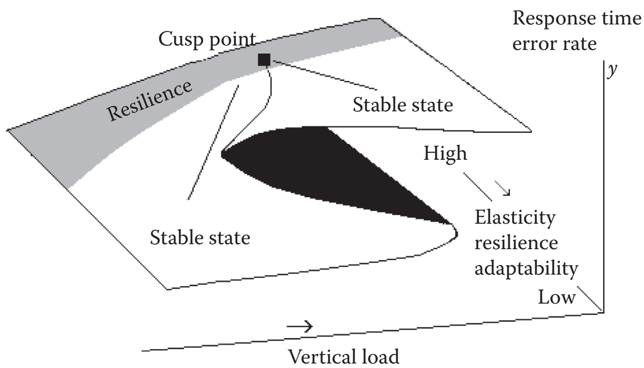

FIGURE 2.9
Cusp catastrophe model labeled for the cognitive workload and resilience problem.

solids, which are known to follow the cusp catastrophe equation literally (Gilmore, 1981). The equation of the cusp response surface for a process that changes over time is

$$dy/dt = y^3 - by - a \tag{2.6}$$

where y is the dependent measure, b is the bifurcation variable, and a is the asymmetry variable. The explanation of the statistical analysis for catastrophe models can be found in Guastello (1995, 2002) and Guastello and Gregson (2011).

The cusp model shown in Figure 2.9 is labeled for a problem that we revisit in later chapters. Although there are several constructs of resilience, this particular variety is a buckling problem for workload. The model depicts two stable states of performance; "stable" does not mean "desirable" any more than "catastrophe" means anything but a "sudden jump" from one state to another. There are two control variables involved. As increase load is put on a system, a rigid system will resist the change in performance very effectively up to a point until it snaps. A more flexible situation will waffle a bit under load, but will not approach the bifurcation where the performance errors suddenly appear. The more flexible state is considered "resilient." The downside of this insight, however, is that "resilient" is not necessarily stable, although it can be made so by first taking the system toward the cusp point, then redirecting it toward locking in the desirable characteristics that are needed to cope with the increase in workload.

Emergent Phenomena

Emergent phenomena are those that seem to appear out of nowhere, and when they do, they cannot be explained as the simple additive result of the actions of the agents within the system. Something new has taken place. Although there is some debate as to whether all emergent phenomena result from the self-organizing processes outlined above, a good many of them do so (Goldstein, 2011).

One contrarian result is the case of a variable that seems to play no role in the behavior of a system during one epoch of the system's existence but might indeed play an active role later on when some other aspect of the system underwent a change (Guastello, 2002). In a time series analysis, the problem is called *ergodicity*: Are the dynamics that are observed during one portion of the time series the same throughout the entire time series? Sometimes they are, but sometimes there are transient events that disturb the normal course of events. Because they are transient, the system reverts back to its earlier dynamics, but not exactly picking up where it left off (Guastello & Gregson, 2011; Gregson, 2013). Sensitivity to initial conditions may go a long way toward explaining why. Thus, there may be some value to the ancient wisdom, "You cannot step into the same river twice."

DISCUSSION QUESTIONS

1. Consider two different examples of the same type of product, such as a handheld computer device or a videocassette recorder. Imagine that you are looking at two prototypes. Sketch a plan in as much detail as possible for assessing the usability of the devices.
2. Consider a device or piece of equipment that greatly reduced human labor when it was first introduced. How did the allocation of function between the person and machine shift?

3. What would be some events in the life of a person–machine system that would fit the description of chaos?
4. How many events can you think of that would be distributed (based on your first impressions), as an inverse power law distribution? What would that distribution tell us about those events?
5. A system might make a dramatic change when it becomes so brittle that it reaches the point of no return. What dynamic model would describe this type of change? What variables might be part of the process that could make it easier or more difficult?

3

Psychophysics

How bright is a bright light? How loud is the sound? Did anyone notice that the lights became dimmer suddenly? Did anyone notice that the sound has gradually become louder? Why was the bottlecap in the beach sand mistaken for a coin? Why was the blip on the radar screen mistaken for an enemy aircraft? It was discovered early on that the answers to the foregoing questions was not as simple as noting the number of watts or candlepower in a light or decibels associated with a sound level. "Oops" may have been a good enough answer for the beach sand question, but not nearly good enough for the aircraft question.

Psychophysics is the study of how humans respond to the strengths of physical stimuli of all sorts. Although the strength of a psychological response to a physical stimulus is related to the physical strength of a stimulus, the relationship is not proportional. There are four principal theories of psychophysics. Classical psychophysics was the original approach, and signal detection theory is considered standard. There are also two frontiers, fuzzy signal detection theory and nonlinear psychophysics.

Classical Psychophysics

The foundational contributions are credited to Weber (1846/1978) and Fechner (1964), although much credit should be given to those who participated in the debates that ensued (Murray, 1990). We consider next the basic concepts that they introduced and the general laws that they derived.

Threshold Concepts

An *absolute threshold* is the minimum value of a stimulus that can be detected. A light does not get any dimmer before you cannot see it at all. A sound is never quieter, or weights on your hand ever lighter. In early times it was believed that such absolute values existed for all people, and the goal of the research was to determine what those critical values happened to be.

A *difference threshold* is the minimum change in the value of a stimulus that a person can detect. This amount of psychological difference was further known as a just noticeable difference, or JND, and one may speak of stimuli differing by two, three, or more JNDs.

In the classical paradigm, the absolute threshold for detecting a sound (loudness) would be determined in a simple experiment, the general format of which is depicted in Figure 3.1. Human subjects would be presented with sounds of increasing loudness. When the subject first indicates that a sound is present, an estimate of the absolute threshold is obtained. Next, the subject is presented with sounds of decreasing loudness. When the subject first indicates that the sound has disappeared, then the second estimate of the absolute threshold is obtained. Typically, the two estimates differ such that the first estimate is physically louder than the second. Thus, the average of the two estimates was taken as the real value.

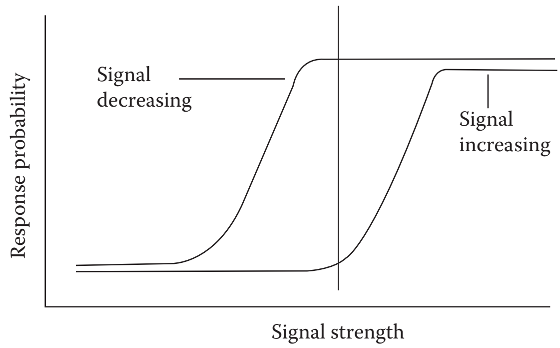

FIGURE 3.1
Charting absolute thresholds in a classical psychophysics experiment.

The y axis of Figure 3.1 is interchangeably known as the response criterion. Each person in the experiment has a criterion of signal strength that triggers the "yes" response. It is the experimenter's task to find out what that criterion happens to be. The notion of response probability was introduced a bit later in time when it became apparent that absolute thresholds were not so absolute; this issue is addressed by signal detection theory.

Difference thresholds in the classical paradigm were obtained in a similar fashion, except that the human participants would be presented with a standard stimulus and a second stimulus. The minimum value of the second stimulus that would produce a response from the subject saying that the two stimuli are different is taken as the JND.

Fundamental Laws

Classical psychophysics produced two important laws. According to Weber's law, the ratio of the change in intensity to initial intensity is a constant:

$$\Delta I/I = C \text{ or } \Delta I = CI \tag{3.1}$$

or alternatively, that change in perceived intensity is proportional to the initial intensity, where ΔI = JND, and I = initial intensity. C is constant for all values of a particular type of stimulus, such as loudness of a tone, brightness of a conventional lightbulb, and so forth. Not surprisingly, Weber's work began with the perception of weights. It was not until electric power became available, along with equipment to plug into it, that it was possible to deliver controlled stimuli of many other varieties.

Fechner's law was derived from Weber's law. According to Fechner's law, psychological signal strength is a logarithmic function of physical signal strength:

$$\Psi_s = k\log(I), \tag{3.2}$$

where Ψ_s is the psychological magnitude, k is a constant, and I is physical magnitude. Once again, k is specific to the type of stimulus (d'Amato, 1973).

A close variant of Equation 3.2 specifies that the sensation of a stimulus is relative to the absolute threshold, I' for the stimulus:

$$\Psi_s = k\log(I/I'). \tag{3.3}$$

According to Murray (1990), Fechner devoted a good portion of his next 30 years conducting experiments that would define Weber's law, on which his own function was based. There were also some important criticisms to counter. First, if two different thresholds are obtained for ascending and descending series, the absolute threshold is not so absolute as the initial theory would indicate. Psychologists discovered later that the two-threshold effect was an artifact now known as the *response set*: If the human subject has habituated to responding "yes" for a while, the person will continue to respond "yes" even when the stimulus strength has fallen below the response criterion. Similarly, the "no" response persists after the stimulus exceeds the threshold. The response set problem goes away when the experiment is designed so that the stimuli are presented in random order with respect to their intensity.

The second problem was a critique offered by physicists, that it was not possible to measure psychological sensations with the same kind of meaning as one would associate with physical measurements. People do use measures such as "very bright," "not so bright," "very dim," and even "twice as bright as" in common parlance, but this form of subjective rating is not objective measurement. Psychologists applied the requisite statistical elements to handled this problem, and the basic solution is presented in the next section of this chapter.

The third problem was that if a weak stimulus is presented with a strength below the absolute threshold (I' in Equation 3.3), the result is a negative sensation value. What is the meaning of a negative sensation? Although the Weber–Fechner laws lost accuracy for extreme values of I, they continued to be used with success for mid-range stimulus values for one-dimensional stimuli. The classical model incurred greater difficulty (long after Fechner's productive era) when multidimensional stimuli were involved, such as in the perception of small color differences (Wandell, 1982).

The fourth problem was that a competing theory was already launched that posited that the relationship between physical stimulus strength and sensation was based on an exponential function. It took more than half a century to settle that particular issue in the form of signal detection theory.

Scaling Procedures

This section elaborates a basic method for scaling the psychological values of physical stimuli that was refined by Thurstone (1927) and Torgerson (1958). The method is so basic that it can be applied to any stimulus, and the stimulus does not need to be limited to lights, sounds, weights, or other specific physical sensations. Indeed the method for scaling physical stimuli became the foundation of other psychological measurements such as attitudes. Although methods for attitude scaling are beyond the scope of this book, a brief foray is included nonetheless because there is a critical result that simplifies scaling procedures that can be used in a wide range of human factors applications.

Psychophysical Stimuli

The *method of paired comparisons* begins with many stimulus values that are presented in pairs to a human observer. Each pair of stimuli must be presented many times. Alternatively, many observers may be involved to make the complete set of comparisons. In theory, these stimuli will have average (mean, or M) values given to them by the observers. These values will have standard deviations associated with them also, so that the

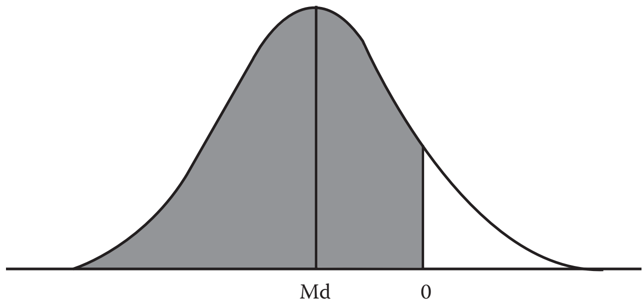

FIGURE 3.2
An underlying distribution of scaling comparisons.

distribution of psychological values will be normally distributed. Furthermore, for any pair of stimuli, we will have two means (M_1, M_2), and the difference between the means (*Md*) will be normally distributed as well.

Figure 3.2 shows a possible distribution of *Md*. The shaded region indicates the proportion of times the observers responded that the strength of Stimulus 2 was less than the strength of Stimulus 1. The unshaded region represents the proportion of times the observers gave the opposite response. If the shaded area represents 80% of the area under the curve, then the observers agreed 80% of the time that Stimulus 2 was less than Stimulus 1.

Consider an example involving four stimuli, A, B, C, and D. The stimuli are presented in pairs, and each pair has been judged 100 times before the experiment is over. Count the number of times that each stimulus has been rated as stronger than each of the other stimuli and put the frequencies into a frequency matrix as shown in Table 3.1. Each table entry represents the number of times the stimulus in the column was greater than the stimulus in the row. Note that it is not necessary to test each stimulus with itself.

The second step is to divide the frequencies by the number of observations per cell (N). Because N is 100, this is an easy calculation to imagine. Insert .50, however, on the diagonals, before proceeding to the next step.

The third step is to convert the table of probabilities to Z score units, which can be done using a normal curve table that is available in any basic statistics book. The result for this example is shown in Table 3.2. For the entry of column A, row B, the probability of 0.70 corresponds to a Z score of 0.52. Imagine that the shaded area in Figure 3.2 is 70% of the area under the curve. The corresponding unshaded area would be 30%; figure 30% of the area from the left of the normal curve, and the Z score is −0.52. The value of −0.52 is entered in the cell for column B, row A. Complete the table in the same fashion. Insert $Z = 0.00$ for the diagonals.

The fourth step is to add up the Z-score entries for the columns, which are shown in bold in Table 3.2. For the final step, it is convenient to add a constant to the ΣZ values to

TABLE 3.1
Frequency Matrix for a Scaling Experiment

	A	B	C	D
A	—	30	60	90
B	70	—	70	80
C	40	30	—	60
D	10	20	40	—

TABLE 3.2

Z Matrix for a Scaling Experiment Corresponding to Table 3.1

	A	B	C	D
A	0.00	−0.52	0.25	1.28
B	0.52	0.00	0.52	0.84
C	−0.25	−0.52	0.00	0.24
D	−1.28	−0.84	−0.24	0.00
ΣZ	**−0.98**	**−1.88**	**0.53**	**2.36**
ΣZ + 2	1.02	0.12	2.53	4.36

eliminate negative numbers; a constant of 2 works well here. The results show that B was the weakest of the four stimuli, and D was the strongest.

Note that the measurement scale is calibrated in arbitrary units. The only concern is that they be calibrated in statistical units. The origin of the scale is arbitrary also, and it does not need to be an integer.

Nonpsychophysical Stimuli

The earliest attempts to develop attitude scales, for example, a person's subjective impression of a social or political event, or liking for a job or a consumer product, were modeled on the method of paired comparisons. Research participants would be presented with pairs of statements such as, "The people who work the hardest are paid the least," and "The pay scales here are fair for what we do." Participants would indicate which of the two statements represented a more favorable attitude toward one's job. A few dozen statements would be compared in this fashion, and the calculation procedure for the method of paired comparisons would be deployed. The result would be a rating of favorableness for each survey item in the set, and a scoring system would be devised based on the favorability ratings. Then, an independent group of people who were really evaluating their job would indicate which statements they thought were true, and their attitude levels would reflect the most favorable statements that they would endorse.

Likert (1932) made an important breakthrough that simplified the development of rating scales by proving that all the information that was necessary to measure an attitude could be obtained by use of the 5-point Likert scale as it is now known. In this procedure, the participants would be presented with a statement for which they would select one of five responses valued as 1 = strongly disagree, 2 = disagree, 3 = ?, 4 = agree, and 5 = strongly agree. The ratings would be summed over all the survey items to produce an attitude score. Although there are many important factors that contribute to the reliability (1 minus the error variance of all ratings on the scale) of such measurements, an important one for present purposes is that the target attitude statements cover a full range of favorability levels.

It is possible to adapt the 5-point rating system to other types of rating objectives by changing the anchors, so long as the relative intensities of the anchors remain approximately the same. For instance consider an ergonomic problem where the goal is to ask participants to rate the comfort or functionality of a leg prosthesis. One of the items might be, "How stiff or flexible is your prosthesis?" The rating options could be 1 = Device is definitely too flexible, 2 = Device may be a little too flexible, 3 = It is comfortable the way it is now, 4 = Device may be a little too stiff, 5 = Device is definitely too stiff.

What is obviously missing in this example is any notion of what the physical features of the prosthesis could be that would generate any of the five levels of response. A study would need to be designed that would associate the design features with the ratings. Another difference between the Likert-type ratings and a conventional psychophysics study is that there are more than two response options, such as signal present versus signal not present. Instead, we have gradations of experience. The issue of graduated rather than dichotomous ratings is resumed later with fuzzy signal detection theory, which follows standard signal detection theory.

Signal Detection Theory

Signal detection theory was developed to address artifacts in the classical psychophysics experiments. We saw already that the random presentation of stimuli got rid of the response set problem, and that psychological ratings could be scaled into cogent measurements that could be used in other calculations and experiments. Other problems remained, however. Participants in psychophysics experiments make judgments that are based not only on the nature of the stimuli themselves, but also on the basis of their sensory acuity. In other words, people with hearing problems might miss sounds (which we now call *signals* more generally) that other people might detect. There are also more subtle individual differences in sensory acuity that would not qualify as problems but that would have the same effect.

The human participants also work under a motivation to respond in the experiment in much the same way as they do in real life. There are gains and penalties associated with certain responses and errors. All this affects discriminative behavior, and thus, subjects' threshold to respond to a stimulus. There remained the problem of predicting responses to stimuli with strengths close to the threshold.

Signal detection theory separates the various influences on human response. Although it was an outgrowth of at least a half-century of work, its consolidation and its centerpiece, the power law, is usually credited to Stevens (1951, 1957).

Threshold Concepts

Signal detection theory introduced several fundamental new concepts that distinguished it from the classical paradigm. The first major change from classical psychophysics was to redefine the threshold. Instead of absolute definitions of the absolute and difference thresholds, there are relativistic definitions. The absolute threshold in signal detection theory is defined as the point where 50% of the subjects perceive the stimulus. The difference threshold was similarly refined as the point where 50% of the subjects detect a difference. The frequency distribution of the number of people detecting a stimulus value as the threshold is thought to be normal (or Gaussian).

Discrimination Index

The second major change was to redefine the task of perceiving a signal: We do not detect simply a signal, we detect a signal relative to noise. Figure 3.3 shows two normal-density functions on an axis defined by signal strength. The left distribution represents the absolute threshold for perceiving a standard noise, and the distribution on the right is the total

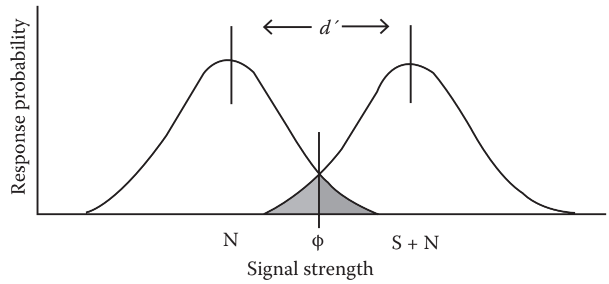

FIGURE 3.3
Distributions for thresholds to detect a noise or a signal plus noise.

strength of a target signal embedded in the noise. The difference in means is known as the discrimination index, or *d*′.

In a standardized signal detection experiment, the human participants are typically selected for having a qualifying level of sensory acuity, and they would thus be regarded as equal in ability (Robertson & Morgan, 1990). One may then proceed to assess the humans' response to the stimuli. Larger values of *d*′ are obtained for signals of greater strength. For signals of greatest strength, all people will detect the signal; those are not the most interesting cases. Midrange stimuli, which may be detected by only some of the people some of the time, are of much greater concern. Ultimately, a signal detection experiment would be performed to determine the error rates and to find a way to minimize them, especially if it is not possible to change the nature of the signal. On the other hand, it is a productive use of human factors knowledge to identify signals that need to be enhanced through some technological means and to determine how much enhancement they require.

Minimization of Errors

Having defined signal detection as a statistical phenomenon, it was possible to express the results of a signal detection experiment in the form of a 2 × 2 contingency table. Figure 3.4 expresses the probabilities of a subject, or a group of subjects giving the right and wrong answers. The 2 × 2 contingency table forms the basis of several types of statistical tests that draw inferences about the overall percentage of right and wrong judgments. The table also

Reality \ Judgment	S + N	N
S + N	"Hit"	"False rejection"
N	"False alarm"	"Correct rejection"

FIGURE 3.4
The 2 × 2 contingency table for signal detection responses.

forms the basis of decision theory where the decisions in question are far more complex than determining presence or absence of a simple physical signal.

The *likelihood ratio* is defined as the ratio of the probability of a hit to the probability of a false alarm. The likelihood ratio is optimal at a point where the two normal curves cross (ϕ) in Figure 3.3.

Probabilities, as defined in the chart, take into account subjects' response behavior under conditions where the noise and signal plus noise are equally likely. If the experiment were devised so that there was a 90% chance of a signal being present, a subject might be inclined to respond as if there were a signal on every trial, thereby obtaining a total accuracy of 90%; errors would be biased toward false alarms.

Subjects respond differently, however, when there are different payoffs attached to each type of response. For example, if a miner were panning for gold, a correct hit could be valuable, but a false alarm error would have a trivial impact; the coin and bottlecap example at the beginning of this chapter is another example of the same set of motivational conditions. In medical decision making, however, a premature rejection of a disease diagnosis could be very costly, thus physicians might prefer to err on the side of overly sensitive testing, and then use follow-up tests to confirm the diagnosis (Swets, Dawes, & Monahan, 2000).

Jury decisions in criminal trials, on the other hand, are often subject to a reverse set of motivational conditions. In the U.S. legal system, the defendant is considered innocent until proven guilty. The evidence, or signals, favoring the guilty verdict should be sufficiently strong that jurors have little reasonable doubt about their decision. Jury decisions are typically made without the interaction with machine systems such as those considered in this book, although individual pieces of evidence may in fact have been the result of a forensic analysis using such equipment. A human factors expert might do well to question the particular forensic decisions made in those contexts. It is a completely different, though relevant, question as to whether jury decisions would turn out the same if critical pieces of testimony were experienced directly in the courtroom or through videotape or telecommunications equipment.

In any case, the miner, the beachcomber, the physician, the forensic specialist, and the juror are making decisions with the same common elements. Something important is present in the real world or it is not present. A judgment is made as to whether the stimulus has occurred. The decision could be in error. There are two kinds of errors, and the goal of a human factors analysis is to minimize errors. Error is minimized when the motivational utilities associated with one or the other type of error are balanced to negate this bias (Robertson & Morgan, 1990).

ROC Curves

A *receiver operating characteristic curve*, or ROC curve, is a plot of the proportion of correct hits, or $P(S|s)$ to false alarms, or $P(S|n)$, for all subjects in an experiment (Figure 3.5). The data points would configure as a set of arcs, where each arc represents a value of d'. Thus, subjects of a given discrimination ability would respond anywhere along their arc. Position along the arc is the result of a response set. The points falling in the lower left of the arc would denote subjects who tend not to respond, and who accumulate many false rejections and correct rejections. The points falling in the upper right would denote subjects who have a low threshold to response to any stimulus, and would accumulate a large number of correct hits and false alarms.

The ROC curves in Figure 3.5 were drawn with their peaks lined up more or less on a common axis. This arrangement is neither necessary nor sufficient. ROC curves can

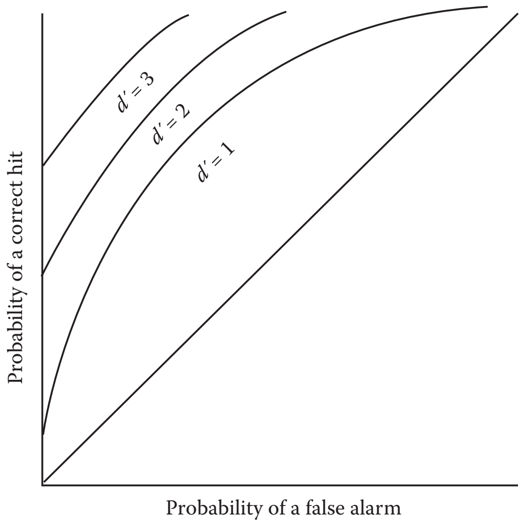

FIGURE 3.5
ROC curve.

be shaped with their peaks skewed to the left or right of their center points. Analysts have found that the particular shapes of the curves result from the standard deviations of noise and signal-plus-noise curves, such as those appearing in Figure 3.3 (Grieg, 1990). As a result, d' should be calculated in such a way as to accommodate the standard deviations of the two distributions in addition to the two means. Equation 3.4 is an appropriate solution:

$$d_A = (\mu_s - \mu_n)/\sqrt{(\sigma_s^2 + \sigma_n^2)/2}. \tag{3.4}$$

In Equation 3.4, d_A is the adjusted discrimination index, μ_s and μ_n are the means of the signal and noise distributions, respectively, and σ_s^2 and σ_n^2 are the variances of the signal and noise distributions, respectively. It is obvious the d' is a statistical value. Accurate values of d' will be obtained to the extent that the values of d' that are obtained from the research sample approximate the theoretical distributions (populations values) for the stimuli (Grieg). It now appears that population values are tenable when there are at least 50 observations for each particular signal strength used in the experiment. That means that each human participant will have to respond to every stimulus 50 times, or there must be at least 50 human participants who can respond to the set of stimuli at least once (Robertson & Morgan, 1990).

Next, consider an application of the ROC curve to resolve, or at least shed some light on, another forensic application, the best means of structuring a police lineup for identification of a perpetrator by an eyewitness. Faces and body shapes generate a lot of information that challenge an eyewitness's ability to identify the perpetrator from the lineup. This challenge often sits on top of the difficult circumstances in which the viewer saw the suspect. The standard procedure is for the witness to view a batch of perhaps five candidates simultaneously and pick out the perpetrator from the options presented. A new approach that has been gaining attention, however, is to view the candidates sequentially. Although the arguments for and against this method are complex, the relevant concern for present purposes is the contrast between probative value and the ROC method of evaluating the outcomes of the two procedures (Wixted & Mickes, 2012). Probative value is the odds of a correct conviction given the identification of a suspect.

To begin, it is possible that the instructions to the eyewitness could affect the correct hit and false alarm rates. Instructions could emphasize the goal of finding the perpetrator, exonerating the innocent, or both. The instructions, when combined with the witness's own proclivities, could result in liberal, neutral, conservative, very conservative, or extremely conservative responses by the witness, according to a hypothetical analysis by Wixted and Mickes. They compared ROC curves for the five response styles for the sequential and simultaneous lineup formats. "Extremely" and "very" conservative response styles produce hit-to-false-alarm ratios that were equivalent, but the ratios spread out for the other three styles. The best discrimination would be the ROC curve that was farthest away from the diagonal, which was obtained for the simultaneous approach.

Individual Differences

The basic signal detection experiment, the 2 × 2 decision chart, and the ROC curve can be used to assess individual differences in acuity. For such an experiment the stimuli would have to be all alike with respect to signal strength. The chosen level of signal strength would need to be close to the average person's absolute threshold or less. Given this setup, if the stimuli are all alike, then differences in d' would correspond to ability levels, with higher values of d' denoting people of greater sensory acuity.

Throughout this discussion, it has been assumed that the humans in our experiments were paying attention to the experimental stimuli. That might not be a challenging assumption in a controlled experiment, but real life introduces distractions and forms of noise that are not part of the planned experiment. For the most part, we can assess the impact of extraexperimental noise by conducting a signal detection experiment in the environment where the human subjects are likely to work. There all the forms of machine noise, verbal chatter among coworkers, and random yelling and screaming can do what they do; their contributions would be part of the noise curve in Figure 3.3.

Attention can be experimentally manipulated in a visual luminescence experiment, however, by varying the position of the stimulus on the display area. The experiment would begin with subjects being conditioned to respond to a stimulus or noise in the center of the display, thus focusing their attention on that location. Next, the stimuli would be presented in either the target location or in the periphery of the display. A suite of experiments by Hawkins et al. (1990) indicated that stimuli presented at the targeted location were more readily detected compared with stimuli that were presented in peripheral locations. Furthermore, the processing of luminance and location information was simultaneous rather than serial, wherein the brightness of the stimulus would be processed before the location or vice versa.

Power Law

Finally, the power law describes the relationship between the strength of a physical signal and the psychological perception thereof:

$$\Psi_s = k(S_s - S_0)^c, \tag{3.5}$$

where Ψ_s is the psychological rating of signal strength; k is a constant for the idiosyncrasies of the experiment, such as procedural errors that often appear minor or trivial or a particular payoff function; S_s is physical signal strength as before; S_0 is the absolute threshold value of S for each human subject; and C is the characteristic exponent for each type of stimulus. Examples of C are loudness of a sound, 0.6; brightness of a light, 0.33; smell of coffee, 0.55; electric shock, 60 HZ through the fingers 3.5, (Stevens, 1951).

The power law seems to hold for every type of signal detection except the detection of vibration, which seems to require a more complex function. Situations requiring multidimensional inputs and outputs also require more complex forms of analysis.

Although the power law is usually associated with Stevens' work, the first experiments that hypothesized such a law date back to Plateau in the 19th century, which were posed as serious contenders for the Weber and Fechner laws (Murray, 1990). According to Gregson (1998), Herbart published the first known theory of psychophysics in 1812, which contained the earliest version of the power law.

Decisions Revisited

So far we considered the simple psychophysical judgment as a decision between two distinct outcomes—presence or absence of something important. We can now put these simple decisions into a broader context that includes not only simple psychophysical judgment, but also other forms of dichotomous decision such as the detection of a medical abnormality based on a diagnostic test, or the forecasting of the weather conditions from meteorological data. Even though some of the information sources produced by machine systems produce results that are continuously valued, rather than categorical, the human operator is often charged with making a dichotomous action decision based on the information that is supplied. For instance, there may be a 40% chance of rain, but the behavioral decision is "Do I turn on the irrigation system today or not?"

Swets et al. (2000) presented an ROC curve for a real decision-making task showing the slope of the curve at two points to the left and right of the center. For combinations of low false-positives (false alarms) and low hit rates, the respondents had set their response criteria so high that signals were never reported. For combinations of high false-positive rates and high hit rates, the respondents had set their response criteria so low that they hit all the signals, but the false-positive rate was very high. If the costs associated with errors are equal for false-positives and false-negatives, the point of optimal decision making is located at the point where the slope of a d' curve is equal to 1.00; this point corresponds to ϕ in Figure 3.3.

If the errors are of unequal worth, then the decision makers would need to calculate the optimal slope on the ROC curve in order to maximize their utilities. There is a formula for making an exact calculation of the optimal slope for a given decision (Peterson, Birdsall, & Fox, 1954; Swets et al., 2000), but it would require specific values, perhaps actual monetary costs, associated with the decision errors. Often these numerical specifics are not available, and a decision maker may have to trade off a quantifiable monetary cost of one kind of error with the subjective cost of another.

Fuzzy Signal Detection Theory

Standard set theory and the use of Venn diagrams should be familiar to most readers. In standard set theory an object either belongs to a set or it does not belong. In fuzzy set theory (Zadeh, 1965), however, the boundaries of a set are indeed fuzzy; an object has a probability between 0 and 1 of belonging to a set and can even have imperfect odds of belonging to more than one set. Fuzzy set theory has a history of applications in engineering, and its utilization bears a close resemblance to statistical methods for classification of objects into categories. It eventually did cross over into human factors applications in the form of fuzzy signal detection theory (FSDT).

In FSDT, the dichotomy of a signal being present or absent is replaced with a subjective rating of the probability (s) of a stimulus matching a target signal, with values closer to 1.0, or matching noise, with values closer to 0.0 (Masalonis & Parasuraman, 2003; Parasuraman, Masalonis, & Hancock, 2000). The response (r) can remain dichotomous with values [0,1] or be continuously valued, as when some signals are more urgent than others. In an illustrative example, an air traffic controller must keep planes 5 nautical miles (nm) apart horizontally, and 1 nm apart vertically. If two planes become too close together, a direction is issued to one of the pilots. If two planes are 4.9 nm, one plane is just barely within the window for a response, but if two planes suddenly become 3 nm apart, a response to that condition takes high priority. The evaluation of the severity of the situation could depend on how close they are vertically as well as horizontally.

Hits, misses, false alarms, and correct rejections were valued [0,1] in conventional signal detection theory, but now have partial values in FSDT (Masalonis & Parasuraman, 2003):

$$\text{Hit (H)} = \min\,(s, r) \tag{3.6}$$

$$\text{Miss (M)} = \max\,(s - r, 0) \tag{3.7}$$

$$\text{False alarm (FA)} = \max\,(r - s, 0) \tag{3.8}$$

$$\text{Correct rejection (CR)} = \min\,(1 - s, 1 - r) \tag{3.9}$$

The rates for the four outcomes from a batch of stimuli are (Masalonis & Parasuraman, 2003)

$$\text{Hit rate} = \Sigma(H_i)/\Sigma s_i \text{ for } i = 1 \text{ to } N \text{ stimuli} \tag{3.10}$$

$$\text{False alarm rate} = \Sigma(FA_i)/\Sigma(1 - s_i) \tag{3.11}$$

$$\text{Miss rate} = \Sigma(M_i)/\Sigma(s_i) \tag{3.12}$$

$$\text{Correct rejection rate} = \Sigma(CR_i)/\Sigma(1 - s_i) \tag{3.13}$$

One motivation for reframing a signal detection task in FSDT form is to obtain a better understanding of the response to stimuli that are near the signal threshold. Near-misses would be classified as no-signal in the conventional approach, but would be recognized as such in FSDT. A trend that appears in the comparison of conventional and FSDT analysis is that FSDT produces lower hit rates and lower false alarm rates. Stimuli that are closer to the threshold can be ambiguous for multiple reasons. Luggage screening in airports is challenged by differences in luggage design, many possible shapes of several classes of threatening objects, and many forms of distracting objects. Signal detection ability is also influenced by time pressure and emotional context. Culley and Madhaven (2012) reported an experiment wherein 315 students were exposed to x-ray pictures of luggage in which they had to find one of five possible targets. Participants were exposed to the pictures for 2, 4, or 6 s. Prior to the main task, the participants were primed for an emotional condition of anger, fear, sadness, or a no-emotion control condition. The near-miss rate was significantly higher in the 2-s condition for all emotion groups compared with the longer exposures. Significantly different near-miss rates were obtained in descending order for fear, sadness, anger, and controls.

Multidimensional Stimuli

Conventional signal detection theory is rooted in the assumption that the signal has a one-dimensional characteristic for discrimination, such as its brightness or loudness. Real-world stimuli, however, are seldom that simple. Brightness is only one characteristic of a light, which is its black–gray–white continuum. Lights also vary by hue (e.g., the chromatic colors that we know as yellow, red, green, blue, orange, and purple) and saturation (the relative amounts of chromic versus black–white content). We have also seen that lights are detected in a location, which thus adds one or more spatial components to the stimuli.

Sounds are not only loud, but they have pitch, timbre (set of harmonic frequencies that is associated with the fundamental frequency), and location as well. Loudness interacts with location such that our ability to locate the direction of a sound declines as the sound becomes louder (Soderquist & Shilling, 1990).

Taste and smell are also multidimensional. Although we seldom utilize these senses in person–machine systems, small differences in taste and smell are particularly important in food processing and perfume manufacturing. There are two notable exceptions, however: The gas utility companies put a chemical called mercaptan into the natural gas precisely so that we can smell it if gas is leaking in a system somewhere; natural gas is naturally odorless. The other exception is that if a machine system starts to produce an odor, it is usually a signal that something undesirable is going to happen very soon.

In any case, it is possible to scale multidimensional stimuli. Multidimensional scaling techniques can determine how many dimensions are required to explain the apparent similarities and differences among various combinations of stimuli. They compute scale values on each of the resulting dimensions that look very similar to the unidimensional scales we produced earlier in this chapter.

Multidimensional Scaling

One of the first multidimensional scaling experiments and statistical procedures used the spatial distance between U.S. cities as stimuli. Human participants were presented with pairs of major cities and asked to report how close they were in miles. All cities in the set were paired with all other cities in the set. The matrix of distances was then analyzed by principal component analysis. Principal component analysis is a close relative of factor analysis; it is an algorithm to determine the smallest number of underlying dimensions that account for the distances among objects (cities in this case). The final numerical result is an $N \times 2$ matrix of scaling coefficients, where N is the number of original stimuli and 2 is the resulting number of underlying dimensions. Matrix entries are the scaling coordinates. In this example, the two dimensions corresponded to the geographic axes of north–south and east–west (Schiffman, Reynolds, & Young, 1981).

If one were to do a similar experiment on chromatic colors, the stimuli would be the colors, which would be presented to the human participants in pairs. Participants would rate the pairs for how similar they appear to be. The results of the scaling analysis would produce a matrix of scaling coefficients in two dimensions. If we were to plot the pairs of coefficients for each stimulus we would obtain a circular arrangement, which is the familiar color wheel (Shepard, 1962), as shown in Figure 3.6. This is a favorite exercise for students to replicate. Multidimensional scaling programs can be obtained on popular statistical software packages.

After producing the set of scale values there may be further questions regarding how different the stimuli would appear to be. To determine the difference thresholds for, say, a pair of colored lights, a signal detection experiment would have to be set up in which the lights are presented in pairs of pairs. Subjects would then respond as to whether pair AB is as different as, or less different than, pair CD. The statistical analysis would be an extrapolation of the techniques already discussed; the interested reader should consult Wandell (1982).

Multidimensional Nonlinear Psychophysics

So far we have progressed from simple stimuli to multidimensional stimuli. Next, we consider the dynamics of perception once the stimuli are perceived over time. The observer could be standing still while the stimuli move, or the observer could be moving through an environment where stimuli are occurring. It would be possible to invoke FSDT here where observers would rate the odds of an event based on their subjective notion of how often

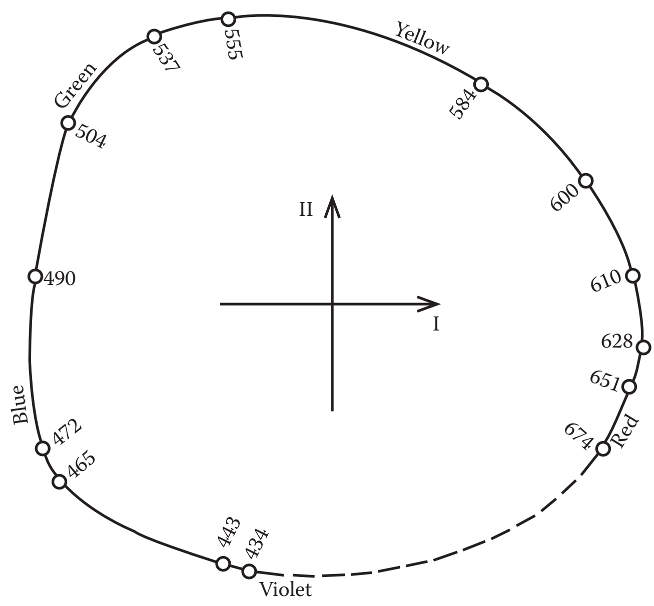

FIGURE 3.6
Multidimensional scaling of perceptual similarity among colors. (From "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Functions II," by R. N. Shepard, 1962, *Psychometrika, 27*, p. 236. Copyright 1962 by the Psychonomic Society. Reprinted with permission.)

a stimulus appears in a series. The FSDT approach, however, renders the time series into static form, and thus does not capture the multistability of events over time or any important temporal patterning. Furthermore, the ratings could be unduly influenced by illusory perceptions of connections between events over time (Botella, Barriopedro, & Suero, 2001). The illusory relationships are essentially autocorrelated errors. Both problems are handled particularly well with NDS models generally (Guastello & Gregson, 2011), and in nonlinear psychophysics specifically (Gregson, 2009).

Although the ogive curves drawn in Figure 3.1 have been assumed to result from a Gaussian probability function ever since they were introduced in the early 1890s, they could just as readily result from a variety of NDS processes (Gregson, 2009):

> It can be shown that the existence of a trajectory that begins close to an unstable fixed point … is repelled from this point as [stimulus intensity] increases … but then suddenly snaps back precisely to this unstable fixed point. This property of the dynamics is sufficient to imply chaos …. [I]t has [also] been shown that some neural networks can embody snap-back trajectories …. Bringing together these modeling features in one context is worth serious consideration in seeking dynamic representations of psychophysical stimuli (p. 114).

It is noteworthy, furthermore, that the two ogives shown in Figure 3.1 when taken together represent hysteresis between the unstable (no signal) and stable (signal present) states, which further supports an NDS process taking place.

The theoretical development (Gregson, 1988) began with the limitation of preexisting theory to handle multidimensional inputs and outputs. He noted that olfactory stimuli, which are generally regarded as multidimensional, have a two-stage decay in effectiveness that is well known but not incorporated in standard psychophysical theory. The gamma recursive function was developed therefore to provide such an expandable model of psychophysical processes, which is based on the logistic map (Gregson, 1992):

$$\Gamma : Y_{j+i} = -a(Y_j - 1)(Y_j - ie)(Y_j + ie). \tag{3.14}$$

Equation 3.14 states that the strength of a response to a stimulus, Y, at time $j + 1$ is a function of the response at a previous point in time, j, a control parameter representing physical signal strength a, a situational control parameter e, and the imaginary number −1. Response strength might be measured as a subjective numerical rating, although the preferred method would be to measure response time. The shorter response times will be required for stronger stimuli. It is noteworthy that as the signal strength becomes sufficiently weak, the pattern of responses over time becomes chaotic.

Equation 3.14 can be expanded to two response parameters that may be useful for modeling cross-modality responses, such as interpreting a hue of a colored light from the perception of its brightness only, or studying the size–weight illusion. Two-dimensional outcome expansion is accomplished by substituting a real number, x, for i (Gregson, 1992, p. 27):

$$X_{j+1} = a\left(e^2 - e^2 X_j + X_j^2 - Y_j^2 - 3X_j Y_j^2 - X_j^3\right) \tag{3.15}$$

$$Y_{j+1} = aY_j\left(-e^2 + 2X_j - 3X_j^2 + Y_j^2\right). \tag{3.16}$$

Γ is further expandable into multidimensional inputs and outputs that were applied to real situations such as perceiving the taste of wine. For superimposed stimuli, Γ has the convenient property

$$Y_{TOT} = \Gamma(a_1 + a_2) = \Gamma(a_1) + \Gamma(a_2) \quad (3.17)$$

in the parts of Γ involving real numbers (Gregson, p. 165). The presence of imaginary numbers is particularly provocative, inasmuch as psychology has had a difficult enough time connecting psychological experience to real stimuli with real numbers. We saw already, however, that cross-modal perception can be tracked by replacing the imaginary number component with a second real component. In other words, the imaginary numbers in the Γ function indicates that a second function should be present, even though we might be able to harness only one at a time. If the two stimulus parameters have different intensities, and the experimental effects e_i are also separate, one can expect further interactions between a_1 with e_2 and a_2 with e_1 (Gregson, 2009).

Every psychophysical process, according to Gregson (1992), is the result of at least two functions, one involving the proximal connection between stimulus and response, and one pertaining to the delivery of the stimuli over time. Virtually all the stimuli in conventional experiments utilize a fixed-interval stimulus-delivery time. No such regularity can be assumed in the real world. Rather, coupled oscillators should be prevalent. For instance, if a system consists of a chaotic slave, which is proximal to the perceiver's response, coupled to a periodic driver that delivers the stimuli, the overall pattern of response should appear as a periodic function that is peppered with bursts of chaotic behavior.

One plausible explanation for why stimulus perception could be chaotic near the absolute threshold or why there could be ghost or imaginary functions of sensory responses might be found in an analysis of the neurological events that take place in the early stages of any perceptual process. According to Hess (1990), the neurons in the visual system are dedicated to specific tasks such as the encoding of brightness and color, but also of contours such as straight lines, curves, repetitive sequences, and so forth. If an incoming stimulus is just strong enough to activate neurons, it would follow that all pertinent systems would be activated. It does not matter whether the experimenter is interested in just one aspect of the incoming stimuli or more aspects; the systems will work as intended anyway. Furthermore, the sensitivity to certain ranges of signal types becomes more acute if the human operator is charged with making distinctions among similar stimuli within that range. The many facets of a stimulus can be expected to interact in complex ways to produce a complex pattern of responses over time (Gregson, 2009).

In other words, the human cognitive system is more complex than any model that scientists have developed to date. We should bear this point in mind as we proceed through our further adventures in human factors engineering.

DISCUSSION QUESTIONS

The following situations represent applications of signal detection theory to person–machine systems. Although the theory covered thus far was developed with the intention of explicating what happens in a wide range of real-world situations, real-world situations are not always limited to single phenomena.

1. In 1986, a radar operator aboard the *U.S.S. Vincennes* noted a signal that he interpreted as an enemy missile. He responded immediately with the result that a civilian aircraft was blown out of the sky, killing hundreds of people on board.

How could disasters like this one be prevented in the future? Hint 1: The representations of enemy missiles and civilian aircraft on a radar screen are very similar. The representations appear for only a brief period as the radar beam sweeps the sky. Hint 2: At least one of the possible remedies for this situation appears in a previous chapter.

2. The Air Force wants to know whether a new pilot–cockpit interface will help fighter pilots destroy enemy aircraft in air-to-air combat situations. How can signal detection concepts be used to assess pilot performance (in a simulator)? If training is needed, what indices should be compared before and after training? This was an actual problem that was reported in Eubanks and Killeen (1983).
3. A sophisticated printing company is in the business of printing wood grain designs on paper or vinyl that, after the print works are sold, are annealed to plywood substrates that are sold for decorating homes, offices, and furniture. The printing process involves four to eight separate colors. The original printing cylinder for each color is engraved separately and loaded onto the printing machine. After long-term use, the printing cylinders wear down and do not pick up enough ink; differences in print quality will eventually be detected. A problem situation arose in the manufacturing operation when it was noticed that, for long print runs, the print quality had deteriorated so that the entire roll of printed work (anywhere from 5000 to 20,000 ft) had to be destroyed. What is the signal detection problem here? How complex is it? What could be done to prevent this type of industrial waste? Note: The foregoing was an actual problem reported in Guastello (1982). Note that we are not considering any contribution of occupational stress at this time.
4. A medical school wants to train technicians to recognize the locations of cancers from magnetic resonance images. How can principles of signal detection be used to track the progress of the trainees?
5. Technicians use ultrasound equipment to detect possible cracks in aircraft wings. This method allows testing without dismantling the aircraft. Most of the results from ultrasound analysis indicate subcritical events or conditions in the metal. The technicians must make a judgment as to whether the conditions are severe enough to warrant taking a plane out of service to make a repair. What are the utilities associated with the judgment errors in the situation? How would the technicians set optimal thresholds or response criteria?

4

Visual Displays

This chapter describes the central principles for designing effective visual displays. It begins with a synopsis of the human visual system and principles of perception. Expansions of these principles can be found in a standard text on general psychology and in advanced-level texts on sensation and perception. The scope of visual sensation and perception for this chapter is necessarily confined to the principles that are most closely related to the design of person–machine systems.

Visual displays for conventional person–machine systems are considered next, followed by some key points concerning signs. Although the same principles carry over to computer-based systems as well, computer-based systems introduced some new issues and potentialities for control–display *interaction*, which are considered in Chapters 11 and 12. The last section of this chapter pertains to lighting and glare. It would be equally feasible to consider lighting and glare as part of a chapter on workspace or environmental design. For present purposes, however, the full value of a visual display is appreciated and evaluated within the environmental context in which it is viewed.

Sense of Vision

Visible light is a relatively narrow band of electromagnetic energy. Its wavelength ranges from 300 nm for violet-colored light to 500 nm for red-orange light. The colors that people recognize as pure shades of blue, green, and yellow correspond to approximately 375, 525, and 625 nm, respectively. The color regarded as pure red is actually not on the visible light spectrum; it is produced as an additive mixture of red-orange and blue light.

Figure 4.1 shows a diagram of the eye. Light, as reflected off objects in view, enters through the pupil, then is focused by the lens to project on the retina. The retina contains receptor cells. When light reaches the receptor cells, a chemical reaction occurs that is very similar to the chemical reaction on photographic film. As the receptor cells are activated by light, they stimulate adjacent neurons in turn. The stimulated neurons aggregate their signals in the optic nerve, and transfer the signal to the thalamus region of the brain. The thalamus is a relay center in the brain, and it is located in the central cerebrum below the cortex (part of the limbic system, located below the cortex). The visual signals are then transferred from the thalamus to the visual projection area in the occipital lobe (back of the head).

Visual Acuity

The clarity of a visual image is the degree to which specific points on the retina are stimulated by the source image. The shape of the lens relative to the length of the eyeball affects visual clarity. Eyeglasses and contact lenses routinely correct deficits in clarity.

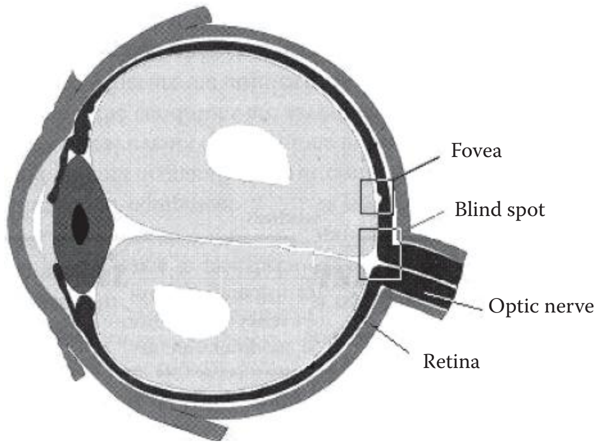

FIGURE 4.1
Structure of the human eye.

There are two different types of receptor cells, known as *rods* and *cones*. The rods have a lower threshold to respond to light, compared with the cones, and comprise the primary visual receptors in night vision or low light conditions. Rods are indifferent to color. Cones, on the other hand, do their best work in daylight conditions and are responsive to color.

The *fovea* is a region of the retina that contains a high density of cones, relative to other sections of the retina. The *optic disk* on the retina is also known as the *blind spot.* It is the location where the optic nerve joins the retina, and there are no receptor cells in that particular region. Anyone who has learned to drive an automobile has learned the importance of the blind spot when attempting to change lanes. Once upon a time, automobiles did not have passenger-side mirrors. Drivers were required to turn their heads about 120° to see if they were likely to hit another vehicle entering their blind spot. Passenger-side mirrors do assist drivers, although many still rely on turning their heads if no danger is visible in the mirror. The mirrors often introduce a distortion in depth perception such that objects are somewhat closer than they might appear in the mirror. Turning one's head involves taking one's eyes off the roadway in front, which can be dangerous under some traffic circumstances, yet it is still the recommended response before changing lanes if the mirror looks as though the space is clearly open.

The quality of a visual image source can be defined in terms of the specific points within the image source that the viewer can discern. For instance, a common television screen is a matrix of 207 dots per cm (525 dots per inch, dpi). "High-resolution" television is twice as dense. For the purposes of the following formula (Kantowitz & Sorkin, 1983), we only rely on the number of horizontal lines, 525 lines per inch. The quality of a visual image can be defined in terms of the number of discriminable lines, l:

$$l = 57.29\ h\ N/R\ (fov), \tag{4.1}$$

where h is the projected image height, N is the number of lines inherent in the picture that is being displayed, R is the distance between the viewer and the target image, and fov is the system field of view, measured in degrees. The field of view is an angle created by the left edge of the image, the eye, and the right edge of the image. Its close relative, the field of vision, is considered later on in this chapter. It follows from Equation 4.1 that high resolution television affords a clearer image at a longer viewing distance. For a period, very large televisions (>91 cm or 36 in diagonal) were popular enough to support their manufacture, but not universally so. Users placed them in rooms in homes that had shorter than optimal viewing distances. As a result, images projected from a 525 dpi source were too grainy for most tastes because *l* exceeded the line capability of the source and not necessarily worth the space in the room they would occupy. The combination of wall-mounted flat screens and high-resolution source images seem to resolve the primary issues in image quality.

In an experiment with Air Force personnel viewing aerial photographs, Turpin (1981) systematically varied the relative amounts of blur and noise in the photographs. Image detection accuracy was greatly affected by the presence of noise, relative to the presence of blur. In other words, noise adversely affected visual signal detection to a far greater extent than the blur.

Color Vision

The cones are not all alike; they respond to different colors. Two lines of experimentation explain color vision: the trichromatic theory and the opponent process theory (Sternberg, 2004; Wandell, 1982). In the trichromatic theory of color vision, which explains why primary colors are primary, three wavelength thresholds have been isolated that correspond to red, blue, and green light. The cones respond to wavelength at different thresholds.

The opponent process theory of color vision specifies that some cones are responsive to red and green, whereas others are responsive to blue or yellow. More precisely, one subgroup of cones is responsive to red or green interchangeably. If a particle of red light shines on that cone, it will respond as red, but if it responds as red, the same cone cannot respond as green. The adjacent red–green cone, however, can respond as green if green light hits it. The other subgroup of cones responds to blue or yellow. Unlike the red–green cone, however, it can respond as blue, but yellow is a default response from the result of "no red, green, or blue."

As a practical application, all chromatic colors can be decomposed as combinations of red, blue, and green light sources. The printing industry has thus developed the four-color printing process, wherein any multicolor image, such as a photograph, can be reproduced as relative combinations of red, blue, green, and black inks of different densities. Similarly, the electron guns behind every television screen and computer terminal produce a full-color image by projecting red, blue, green, and black onto the image shown to the viewer.

Another form of support for the opponent process theory is the phenomenon of *negative afterimages*. The classic demonstration is to focus one's eyes for 2 min on an image of the U.S. flag that is presented in opposite colors, green, yellow, and black. Keep staring as the image is removed. One then sees an afterimage of the flag in the correct colors, red, white, and blue. Prolonged activation of a color sensor in a cone triggers the activation of the opposite color once the stimulus is removed. The same process work spatially: An intense activation of red on one cone triggers a green response from the adjacent cone on top of whatever is supposed to be adjacent.

Graphic artists knew the secret of negative afterimages for years and could sharpen a multicolor painted image by applying a thin border of black between two adjacent color regions. As another example, the first color televisions were not universally accepted. Although price was a factor, many viewers preferred the sharpness of black-and-white to the fuzziness of color. The breakthrough came when Quasar (a television manufacturer) introduced an imaging technique that surrounded the projected dots of color with black, which negated the lateral afterimage. The result was a sharper color image.

There are two types of color-mixing processes, additive and subtractive. The additive case is exemplified by theater lighting, wherein white light is blasted through a colored gel and projected onto a target. If lights projected through red (actually reddish orange), blue, and green gels are projected onto the same spot on the stage, the viewer sees white. If, on the other hand, white light were projected through red, blue, and green gels (filters) simultaneously, the result on the stage would be black. That is to say, triple-screened light would be no light at all. An additive mixture of two opponent colors, such as blue and yellow, produces gray light.

The mixing of paints and dyes exemplifies the subtractive mixture. Each color in the mixture is actually a range of wavelengths. One range of wavelengths filters the other range of wavelengths. The only wavelength range that is allowed to pass is the color the viewer sees. Thus, blue and yellow crayons produce green.

Colors, are distinguishable in terms of three properties: hue, brightness, and saturation. Hue is the distinction among red, blue, green, yellow, and so on. Brightness is the relative black–white dimension, with all shades of gray located between the two extremes. Saturation is the relative of the hue plus the gray-scale dimension. Thus, red plus white equals pink, and red plus black equals maroon.

The complete range of perceived colors can be represented as a 3-D display, shaped like two cones with the bases connected together, and the points at opposite poles. The cone bases in the middle correspond to the color wheel where the spectrum of hues is depicted along the outer edge. The brightness axis, which ends at the two points of the cones, is transverse to the color wheel. Highly saturated colors are positioned on the outer edge of the cone bases, and unsaturated colors are positioned closer to the central brightness axis. The color patch strips that are available in the paint departments of hardware stores are simply slices of this 3-D array.

Color Vision Abnormalities

Color vision abnormalities, commonly known as colorblindness, affect 8% of males and 0.3% of females. The most poignant deficit is where the individual cannot distinguish between red and green because of a missing pigment in the receptors. Blue-blindness is known, but it is relatively rare. More common, however, is the form of distorted color perception that is caused by deficits in the opponent process where color mixing takes place.

The Ishihara tests for colorblindness are often used to screen colorblind individuals for jobs where color vision is important. The stimuli are round stimuli that are composed of small round color patches of different sizes and tones. The colored patterns, which look a bit like mosaics overall, show a two-digit number as part of the pattern. The color-normal person sees one two-digit number, but a color-abnormal person sees a different two-digit number. There is also a version of the test that triggers either a recognition of the correct number or no number at all.

Ishihara screening has been used advantageously by the Air Force. People who fail the test fly as reconnaissance agents. Their task is to watch the ground for movements of enemy troops and equipment. Whereas the color-normal person would be fooled by camouflage, the Ishihara-blind person would see the disguised equipment or people clearly. Colorblind and color-normal people are about as different in their abilities to discern small woodland animals in a forest, just as another example.

Many industries that require tight controls over the color production of painted or printed products utilize Ishihara tests to screen out employees with the most blatant cases of color misperception. Industrial color matching, however, is subject to additional influences from the task itself and the environment that affect the accuracy of color matching by humans. These include habituation to a small color band, matching two or more color elements at one time or in one display, time pressures imposed on the human to make the match, glare, and ambient lighting (Guastello, 1982). Similarly, computer-supported color-matching equipment can be foiled by the effect of the substrate into which the colored ink or paint pigment is suspended, and thus humans are required to make the final color adjustments to the paint or ink.

Photopic and Scotopic Functions

There are people who have abnormal color vision by the usual standards, but who can distinguish colors nonetheless. They do so by the relative brightness of the colors. In the case of traffic lights, the standard placement of red, amber, and green lights in the signal box also facilitates proper recognition of traffic lights.

People with normal color vision, however, also respond to the relative brightness of colors. In the case of lighted displays, it is also true that lightbulbs of the same wattage will look brighter or darker depending on the color of the light. The *photopic function* is thus the effect of a color on subjective brightness of a light under normal daylight conditions. Its values, V_λ, can be found in tables, and range in value from 1×10^{-4} to 1.0, such that

$$F = \Sigma\, P_\lambda V_\lambda. \tag{4.2}$$

In Equation 4.2, F is the total photometric intensity of a lighted display, P_λ is the physical intensity of the light (e.g., wattage), and V_λ is associated with a color (Kantowitz & Sorkin, 1983; Kaufman, 1974).

The *scotopic function* is similar to the photopic function, with the major exception that the viewer has been in the dark for 40 min before the color light stimuli were presented and judged for brightness. Its values, V'_λ, range from 1.8×10^{-5} to 1.0, and they can be used in the same Equation 4.2 to calculate the photometric intensity of a lighted display when it is viewed at night.

Perception

Perception refers to how we make sense of a visual stimulus once the receptor cells and neurons have done their jobs. Psychologists have had some classic arguments as to how perceptual processes occur, but contemporary reasoning grants a lot of credibility to each of the major viewpoints. Helmholtz (1909/1962), the physicist who produced the first psychological theories of color vision, nd audition, argued that perception was a completely learned process.

Gibson (1950, 1979), on the other hand, argued that it would be a survival advantage for any organism to evolve with the ability to perceive the physical environment in as much detail as it actually contains. Furthermore, we have the ability to read the handwriting of someone whose handwriting we have never seen before. It must, therefore, be true that we have some innate capabilities and predisposition in our perceptual mechanism. This reasoning led to the discovery of *feature detectors*, wherein specific neurons or patterns of neurons in the optic nerve fire when exposed to distinct but elementary perceptual units such as vertical lines, horizontal lines, curves, diagonals, round things, lines that intersect, and so on. A sampling of elementary features appears in Figure 4.2. The correspondence between the elements and the capital letters of the Roman alphabet should be apparent.

Gibson (1950, 1979) also identified *invariant processes* in perception, which are basic processes of perception that are usually associated with depth and motion. The principle of size–distance constancy is an example. If an object occupies a small portion of our visual field, we regard it as small. If we know that it is truly a large object, such as an elephant, we interpret the elephant as located relatively far away. As the elephant gets closer to us, it occupies a greater portion of our visual field. Obviously, the viewer must have a clue about the real size of the object in order to interpret its distance; fortunately, relative size is only one cue for distance perception. The presence of invariant processes, nonetheless, explains why optical illusions produce their illusory effects. In an illusion, some combination of invariant processes is invoked to induce a false conclusion about reality.

Kohler, Koffka, and Wertheimer, who were collectively known as the gestalt psychologists, also concluded that there are fundamental innate perceptual processes (Sternberg, 2004). The gestalt psychologists were concerned about the perception of form, and their work predated the discovery of feature detectors. In their reasoning, the brain (somehow) organizes the elements of a visual stimulus, and creates an interpretation of the whole image, hence the word *gestalt.* They argued that if we could understand how the brain organizes an image, we would be in a better position to understand how the brain works overall. In any case, their thinking led to the famous dictum, "The whole is greater than the sum of its parts."

Hochberg, much later, introduced the constructivist perspective on perception that was, in a way, a throwback to the thinking of some 19th-century philosophers. In essence, our perceptions are not simply a view of what is presented, but a mixture of what is presented as it interacts with our past. We store up elements that are experienced and learned in the past, which in turn assist in the perception of wholes even though we might be given only parts to look at. From a human factors standpoint, Hochberg's thinking, at the very least,

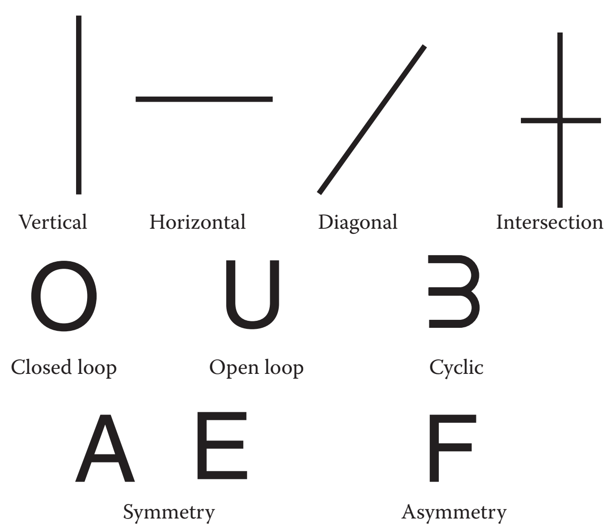

FIGURE 4.2
Feature detection.

gives one more reason to beware of the differences not only between novices and experts but also among experts in different areas of expertise. A broad variety of expertise within a group of system users could readily produce different perceptions of the same stimuli that are novel to all of them.

Form

Figure versus Ground

There are six basic gestalt principles to consider here: figure–ground distinction, similarity, proximity, good continuation, symmetry, and closure. *Figure–ground* distinction is the extent to which we can discern a visual target from a background containing it. Camouflaged objects are intentionally prepared to minimize figure–ground distinction. The vase–face perceptual illusion (Figure 4.3) that has become iconic in psychology works the way it does because the image equalizes the black and white portions of the image so that one can see a vase as the figure in one moment with the faces in the background portion, but the faces pointed nose-to-nose as the figure at another moment with the vase as the background.

The figure–ground distinction led to a concept of field independence versus field dependence that characterizes individual differences in the ability to separate figures from ground. The prototype measurement is the Group Embedded Figures Test (GEFT; Witkin, Oltman, Raskin, & Karp, 1971), which has a long history of use in research on "cognitive style." Field independence has gained some relevance for problems that are not literally perceptual in nature, but require the individual to identify relevant pieces of information for solving a problem from a mixture of relevant and irrelevant source material. Examples include solving chemistry problems (Stamovlasis, 2006, 2011; Stamovlasis & Tsarparlis, 2012), which require the chemist to isolate critical information from extraneous information and hold a number of pieces of information in mind when solving the problem. Field

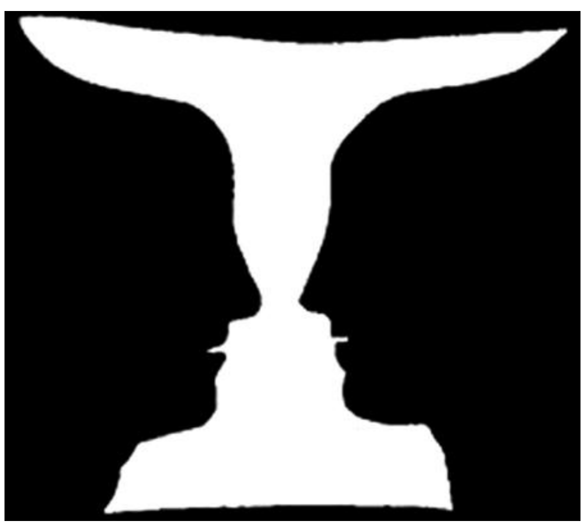

FIGURE 4.3
The vase–face illusion.

independent people are better at solving the problems, or financial decision making, which require the investment analyst to isolate information that speaks to expected future value and the odds of future gains (Mykytyn, 1989; Guastello, Shircel et al., 2013).

Witkin et al. (1971) remark further that the GEFT does not predict performance on all types of spatial perception tasks, only those that require the individual to isolate figure from a ground. According to Pascule-Leone (1970) field-independent people apply more of their innate mental channel capacity (Chapter 6) to a task than field-dependence individuals; the jury is still out on this point, however.

Other Principles

The remaining principles of form appear in Figure 4.4. The upper left display depicts the principle of *similarity.* The two groupings of dots are exactly the same in terms of their size, shape, and relative position. Their patterns of coloration differ, however, such that the grouping on the left induces the perception of rows, whereas the grouping on the right induces the perception of columns.

The upper right portion of Figure 4.4 depicts the principle of *proximity.* There are two rows, each of which contains six vertical lines. The relative proximity of the lines leads the viewer to perceive two groups of three lines in one case, and three groups of two lines in the other.

The principle of *good continuation* is analogous to a road map. If we were driving from Point A to Point O and wanted to continue on the same path, would we end up heading to Point B, C, or D? Most people would head toward B, as it makes the smoothest continuation from A.

The principle of *closure* is depicted in the lower right portion of Figure 4.4. Given the four unshaded dots, most people would place a fifth dot in the lower left of the image where the shaded dot now appears. There is a psychologically unfinished sense of the diagram without the shaded dot. Different graphic displays might produce different reasons why an additional graphic element would produce an experience of closure or not. In this

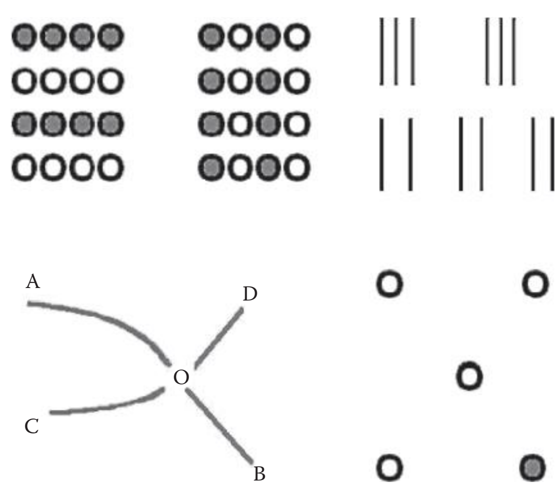

FIGURE 4.4
The Gestalt laws of form perception.

case, the experience of closure relies in part on *symmetry.* People generally prefer designs that are symmetrically organized around a central axis rather than those that are not so organized. On the other hand, aesthetic judgments that are regarded as simple tend to be simplistic, whereas those that are complex embrace asymmetry. From a human factors standpoint, however, it remains to be shown empirically whether a particular symmetric or asymmetric display layout is simultaneously more functional with regard to human performance criteria.

Depth

Binocular Cues

There are several means of perceiving depth. The first is *binocular disparity*, which involves the use of the two eyes. When a scene is viewed with both eyes, slightly different images appear on each retina. The two retinal images are similar to the extent that the portion of the scene is further away. Near-field portions of the image produce relatively different retinal images.

The first 3-D display technology actually appeared in 1847. The Brewster stereoscope facilitated the viewing of 3-D images by allowing the user to mount a pair of stereoptic photographs on one end of the device and look through the eyepiece at the opposite end. The photographs were taken with a double-lens camera that produced two images that corresponded to the binocular disparity that occurs when an image reaches the retinas of two eyes (Faris, 1997). The device was in use until after World War II. It became tedious to mount pairs of pictures that could be viewed by only one person at a time, when a simple photo album or projected slideshow were faster and more accommodating. A more contemporary version of the stereopticon (made of hard plastic) appears in Figure 4.5.

3-D photography resurged as a movie format in the 1950s. The viewer needed to wear disposable 3-D glasses in order to get the 3-D effects (e.g., vampire bats flying off the screen into the audience). The technique may have been successful enough to induce people to see some movies, but not enough to win movie awards; other factors would be involved, of course.

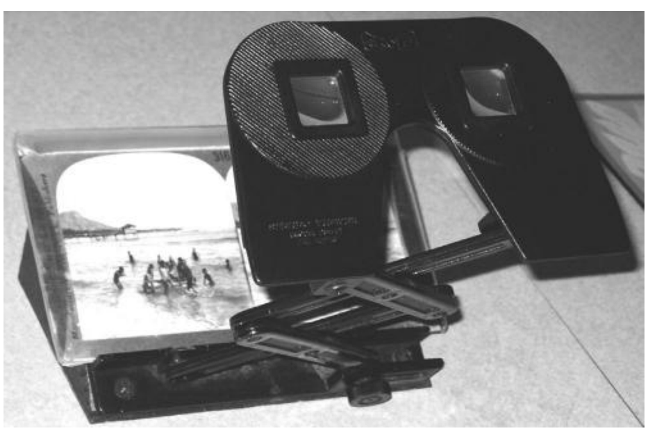

FIGURE 4.5
Stereoscope.

3-D photography resurged (slightly) again in the early 2000s as digital cameras supported taking two pictures and holding the double image. Unfortunately, the camera only had one lens, so the photographer needed to rely on personal judgment to position the separation of the two images. The viewer still needed 3-D glasses to get the effect. It does not appear that digital 3-D skills have become a very widespread.

Monocular Cues

The remaining depth perception cues are monocular cues, meaning that it is possible to determine depth using only one eye. Thus, it is possible for a person with only one eye to perceive depth. Depth can be perceived in terms of relative size. An object takes up a larger portion of the visual field if it is relatively close, and a smaller portion if it is further away. It is important to know the real size of the object in order to judge its distance correctly.

Linear perspective in drawings is the technique for slanting parallel lines so that the lines converge in the distance. Parallel lines, such as railroad tracks and straight roadways viewed in the near field look parallel, but they appear to converge in the distance at a *vanishing point* toward the horizon. Linear perspective combined with size–distance constancy explains the *moon illusion*: The moon looks its usual size when viewed overhead from earth, but looks comparably huge when it is seen rising over the horizon; the moon itself did not become any larger.

Occlusion is the result of one object blocking out part of another object in our view. The object that is blocked is regarded as further away. The U.S. Department of Defense (2012) specifications warn against false occlusions in visual displays. Although they are not an issue with meters and numerical gauges, scenic graphics, which might be cartoonized in some virtual reality or synthetic vision systems, could provide opportunities for this type of design error.

Two objects may be rendered to indicate different distances by placing one higher than the other relative to the bottom of the 2-D canvas. This technique of relative height is more convincing when the two objects are also drawn to different sizes in proportion to their alleged distances.

Texture gradients can be observed in visual images where there is a lot of repeating elements. For instance, if one were to view a field of corn, the corn closest to the viewer would be perceived as separate and relatively detailed stalks. The corn positioned toward the far field, however, would become progressively less distinct.

Lighting and shadows also provide depth information. Portions of the image that are meant to be closer to the light source appear brighter, whereas the portions that are receding from the light source appear darker.

Finally, *motion parallax* is a cue for determining depth while the viewer is in motion. Note that this is a depth cue, not a means of perceiving motion. If we were driving down the highway, the telephone poles at the edge of the road would whip by quickly, whereas the water tower in the distance would float by slowly. Thus, the different speeds of movement of objects in our visual field convey signals of depth. The U.S. Department of Defense (2012) specifications warn against mixed messages from conflicting depth cues: “Conflicting depth cues shall not be permitted. An example of a conflicting depth cue is if a dog was occluded by a house, but the dog was larger than the house. As the dog is behind the house, it implies farther distance; but as it is bigger, this implies that the dog is closer. Depth cues shall assimilate natural depth; exaggerated depth cues shall be avoided” (p. 111).

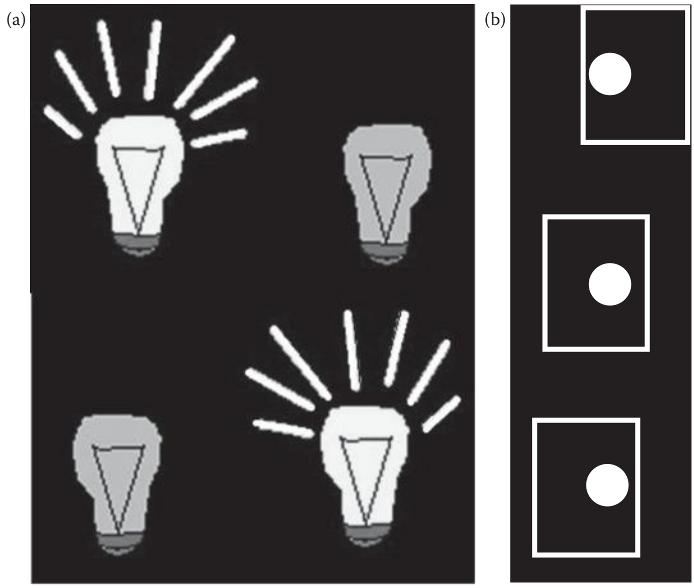

FIGURE 4.6
Motion perception: (a) stroboscopic and (b) apparent motion.

Motion

The perception of motion requires a change in the perceived visual image. The first method is direct perception. When we are standing on the ski slope, the tree in our view becomes larger and larger as it moves toward us.

Stroboscopic motion involves the change in the position of a light source (Figure 4.6). The classic demonstration is to place a viewer in a darkened room with two lightbulbs positioned at a fixed distance apart. The experimenter then flips one light on leaving the other one off. Then the alternate light is turned on while the first light is turn off, and the process is repeated in alternation. The viewer perceives that the light is moving from one location to another.

The strobe lights in the dance halls have the opposite effect because the light position does not move, but the people are moving relative to the light. As the high-intensity light flashes on and off at sufficient intervals, the viewer sees different views of the people in the room. The different views are sufficiently disjointed so that the viewer loses the continuity that is associated with motion. Movie film works on this principle: The basic movie is a sequence of still frames that are rendered fast enough to provide the viewer with the same continuity of visual change as would be perceived from real-world events.

Holding the light source steady and moving the contextual image around the light can induce *apparent motion*. The viewer will perceive that the light is moving. A common example occurs when viewing the moon on a partially cloudy night. For a period of a few minutes, the moon is not moving its location in the sky, but the clouds are rolling by. It appears to the viewer, however, that the moon is rolling through the clouds—until the viewer notices that the moon is not really getting anywhere as fast as it appears to be moving.

Principles of Display Design

Although it is convenient to describe classic display concepts one at a time, the real world at this time of writing is full of displays that are built on more than one principle. The human factors engineer's skill is required to pull the principles apart and detect which features of the display are desirable and which features can be improved. Similarly, it is tempting to discuss displays without any reference to controls, and that is the approach taken here so as not to confuse important issues. It should be remembered, however, that the control and displays often work together to produce a particular result.

Types of Basic Displays

Four classic display models are shown in Figure 4.7. The design in the upper left shows three system values in a piece of equipment located in an electric power plant. The needle on the left, which indicates a value of 65, displays a value that is currently in operation. The needle on the right, which indicates a value of 0, stays there until the operator turns the knob underneath. Turning the knob moves the needle upward to a desired value; the needle reverts to the 0 position when the knob is released. The third system parameter is the AUTO/MAN lights. The lights are mounted on the button itself. One would need to push the MAN button to allow the system to accept a manual control operation. This display design was mounted in the early 1980s.

The display in the lower left is showing four values from related system areas in one display. Again the needle rides up and down over a convex drum. The relevant controls are not closely related to the display. This display design was mounted in the mid-1960s. Other similar-looking designs have been built whereby the needle stays stationary and the numerical drum rolls under the needle. The convex drum in the two left displays predispose the operator to possible depth-perception errors while reading the displays if the indicated values are too extreme to the top or bottom of the display.

The display in the lower right is a circular meter wherein the needle crosses a flat calibrated background in a circular fashion. This design does not invoke the depth-perception problem, but it does require more physical space on the display panel than do some of the drum-type displays. It is susceptible to parallax effect (see later discussion) if the operator tries to read the display from an angle instead of head-on.

The display in the upper right uses a lighted bar to indicate the system value that is currently in operation. The electronic display method obviates the need to roll the needle over a convex drum, thus reducing the chances for errors of depth perception and parallax. The numerical value is also displayed digitally on the top of the display. This display design was placed into the system in the early 1990s.

Display Design Criteria

Inevitably, a designer would ask, "Which type of display is best?" The short answer to this question is, "Pick the right tool for the right job." Testing is required to determine which method of display would work best in a given situation. Theoretical considerations include not only the properties of human vision, but also the context in which the device is used. Are there other controls and displays that go with it or surround it? How many other tasks are going on at the same time?

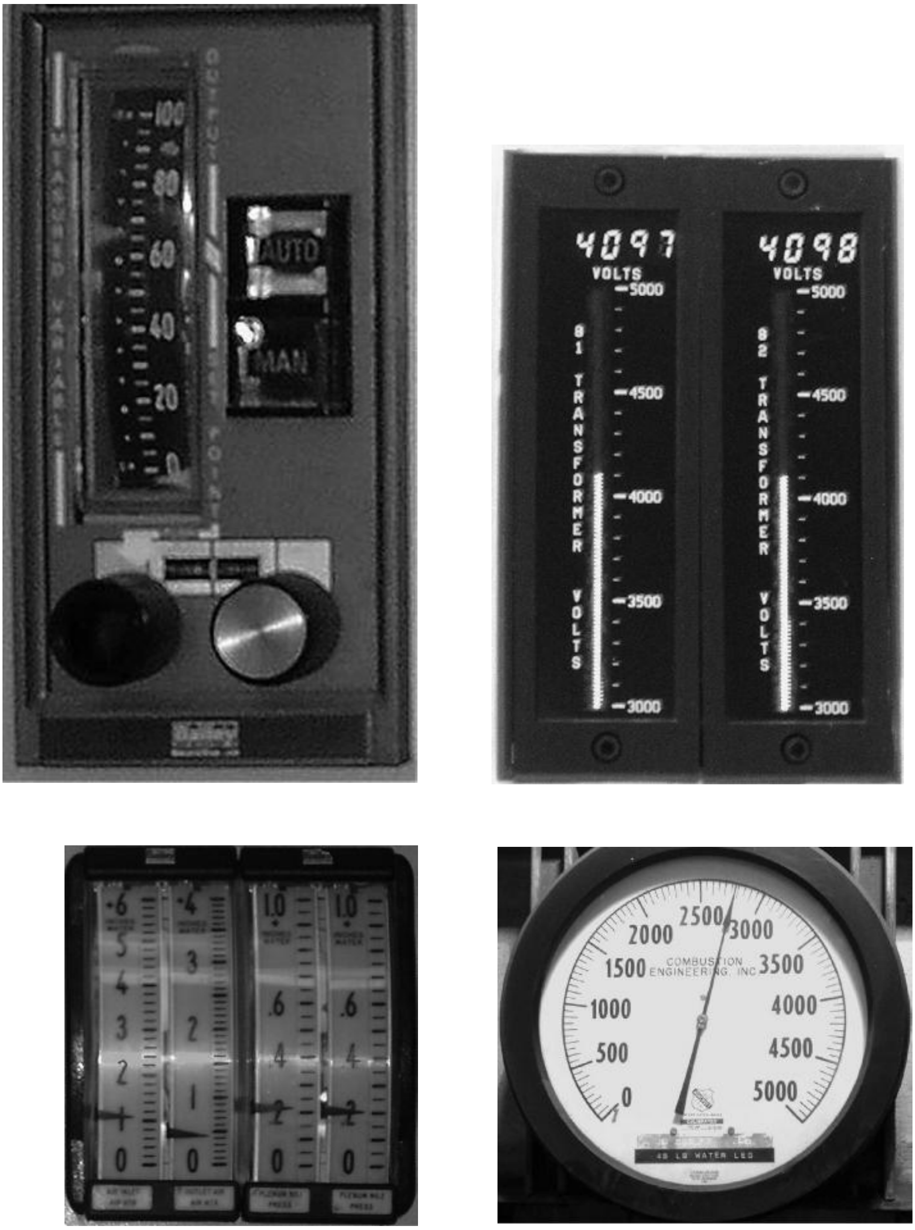

FIGURE 4.7
Common types of displays.

The top criteria for evaluating a display design are: (a) speed of interpretation or reading, (b) accuracy of interpretation or reading, (c) speed of learning, (d) comfort, (e) absence of fatigue with long-term use, (f) small differences in human performance that can be traced to differences in the populations using the device, and (g) the stability of performance under poor environmental conditions or stress (Kantowitz & Sorkin, 1983). Fortunately, the displays that can be read more quickly are also those that can be read more accurately. Thus, it is not always necessary to sacrifice one criterion in favor of another. Some of the more important aspects of display design that translate into performance outcomes are considered next.

Visibility

There are three aspects to visibility. The first is image clarity, which was considered earlier. Image clarity is perhaps most germane to photographic and computer-projected images.

The second aspect of visibility is the location of the display within the operator's field of vision. Figure 4.4 illustrates that the normal line of sight is not directly horizontal from the viewer, but is oriented at a 15° downward slope (Figure 4.8). The optimal orientation of a control panel would be a 45° angle to the angle of vision. The latter is the same as a 30° angle relative to the directly horizontal line of sight (from U.S. Department of Defense, 2012).

Next, consider a panel of displays facing the operator. The most important displays should be centrally located within the field of vision, and the less important displays located on the periphery. (The determination of "important" is another matter, which is considered in a later section of this chapter on display panel organization.) The amount of space within the central location depends on whether the operators are allowed to move their eyes only, turn their heads, or whether they are expected to turn their heads and roll their eyes. The horizontal field is ±15° to the left and right of direct horizontal viewing for eyes only, ±65° for turning heads, and ±95° for eye rolling and head turning (Figure 4.9).

The vertical latitude for the field of vision is also shown in Figure 4.9. If only eye movements are allowed, the accepted lower limit is 15° below the normal line of sight (30° below direct horizontal), and 40° above the normal line of sight. If head movements are allowed, the range is between 65° above the normal line of sight and 35° below. Note, however, that

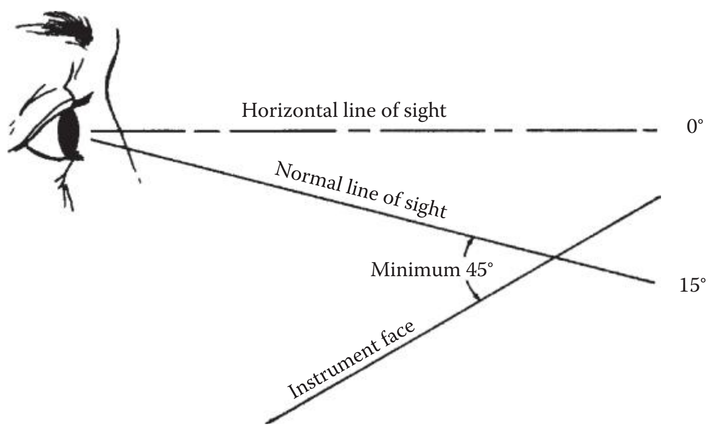

FIGURE 4.8
Line of sight. (From U.S. Department of Defense, *Department of Defense design criteria standard: Human engineering*, Author, Washington, DC, retrieved April 4, 2012 from http://www.assistdocs.com, 2012, in the public domain.)

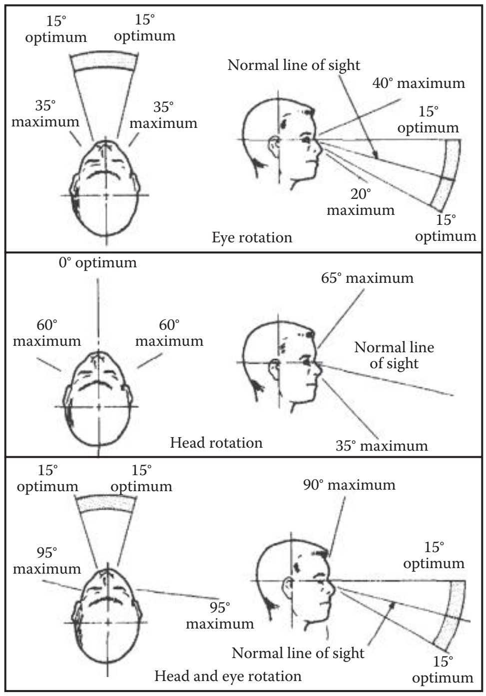

FIGURE 4.9
Field of vision. (From U.S. Department of Defense, *Department of Defense design criteria standard: Human engineering*, Author, Washington, DC, retrieved April 4, 2012 from http://www.assistdocs.com, 2012, in the public domain.)

if head and eye movements are allowed, the acceptable range is between 90° above the normal line of sight (75° above direct horizontal) and 15° below the normal line of sight. In other words, heads are allowed to turn upward and back to center, and eyes are allowed to roll downward and back to center.

Even in the case of a single display, the question of whether an operator could see the display can be answered in terms of the location of the display in the field of vision. Figure 4.10 shows two aberrant examples. On the left, a lonely display is mounted in a busy visual field in which the operator must walk up a ladder platform to see the meter reading. On the right, the meter is located within the vertical head-turning field, but it is upside down. This example of 1935 technology could not be installed right-side-up because the flow of steam within the pipe flowed from ceiling to floor.

Distinguishability

Distinguishability is whether the important elements of information that the operator is supposed to capture from the display are sufficiently clear. Figure 4.11 illustrates the importance of making the tickmarks on an analog display distinguishable from each other. The U.S. Department of Defense (2012) standards specify heights, thicknesses, and spaces between tickmarks. The actual sizes in millimeters depend on the viewing distance.

FIGURE 4.10
Two displays with visibility problems.

Demand is the extent to which the display is indicating a serious condition in the system. Aircraft and environmental control systems often specify "alarm" conditions. Alarm conditions need to be easily distinguished from "normal conditions." Flashing red lights are often a good way to convey alarm.

Figure 4.12 is a reconstruction of a "starfield" display from a computer-based system for which this author was a part of the design team. The dots in the display indicate different sensor activity in five areas of the system. The central axis represents set points, and deviation from the center indicates the change in value for the sensor point over a period. Different colors were used to denote the separate parts of the system. An operator who desired more information about a sensor's activity could then click the dot and additional information would pop up. Alarms were initially denoted as white. The design recommendation was to make the alarms larger and to make them flash. Flashing white against a dark background was not a bad choice for that application. Military equipment

FIGURE 4.10 (Continued)
Two displays with visibility problems.

standards, however, specify flashing red for situations where emergency action is required (U.S. Department of Defense, 2012).

Designers need to give some consideration to what value of a system function constitutes an alarm condition. Ideally, an alarm should mean that the operator must take action immediately or close to it. When alarm criteria are set too low, the operator's likely response is to reset the alarm to a higher value. If the response to reset the system comes too often or too easily, the operator runs the risk of missing an important condition in the system (Chapanis, 1975).

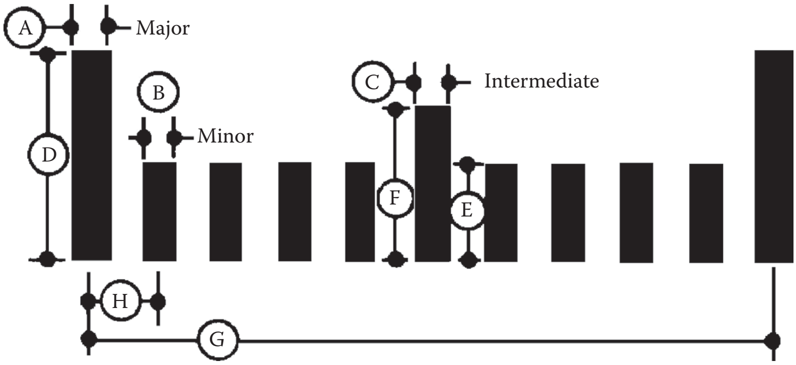

FIGURE 4.11
Analysis of tickmarks for display readability. (From U.S. Department of Defense, *Department of Defense design criteria standard: Human engineering*, Author, Washington, DC, retrieved April 4, 2012 from http://www.assistdocs.com, 2012, in the public domain.)

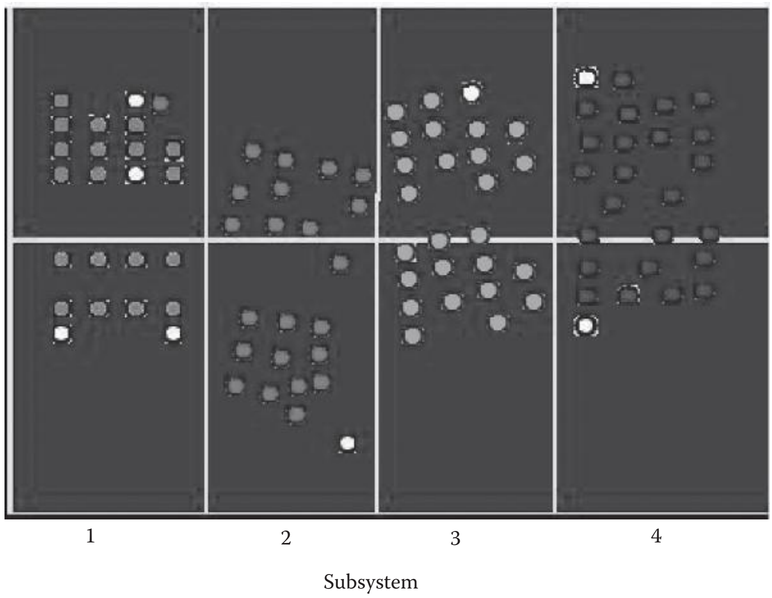

FIGURE 4.12
Prototype "starfield" display.

Military equipment design guidelines recommend that visual alarms be designed so that an alarm means only one condition. The nature of the condition should be apparent to the viewer so that the viewer is not required to read other displays in order to interpret the nature of the emergency (U.S. Department of Defense, 2012).

Interpretability

Interpretability is the extent to which the operator can view a display and get a sense of what the information means and what to do about it, if anything. This aspect of display design gave rise to smart displays in computer-based systems. Traditional displays indicate an aspect of the system from the system's point of view—its machinery, its levels of one thing or another. Those displays typically did not convey any direct information to the user regarding what to do with that information.

A *smart display*, however, will internalize some of the operator's decision-making strategy and produce information in a form that coincides with the decisions that the operator is trying to make. For instance, a military pilot needs to recognize another aircraft as friendly or unfriendly, its type of attack power, and whether it is in shooting range (or within the firing envelope). A relatively dumb display might show that another aircraft exists within radar-detectable limits. A smart display will provide signals regarding the type and friendliness of the aircraft and code its proximity within the effective firing envelope (Post, Geiselman, & Goodyear, 1999). This is a complex signal detection and perception problem that can now be assisted by the machine, rather than done by the human alone.

A smart display relies on a system of meanings that are shared by the population of operators (Bennett & Flach, 2013). If the situation is [X], what would the operators do in response, and why? Are multiple scenarios and outcomes possible? This principle is expanded in Chapters 11 and 12.

Of course, knowing what to display in any form is a major step forward. The Model T Ford had no machine displays, thus no dashboard, and no fuel gauge! One can only imagine the joy of discovery when the designers figured that one out. Speedometers were yet to appear as well; perhaps they were not necessary in the days when traffic kept pace with the horses, and the horses were not equipped with speedometers either.

Similarly, there were airplanes before there were display panels. The phrase "flying by the seat of your pants" meant that pilots relied on the physical contact between themselves and the plane to determine changes in altitude. The altimeter was invented eventually.

Completeness

Completeness of a display denotes whether all the information that an operator requires and that is meant to be provided by the display is actually provided. A simple case would be to have a full range of numerical values shown if indeed all those values were both possible and meaningful (Kantowitz & Sorkin, 1983). The range of values shown on a display is the *space* of the display.

For many years, automobile speedometers were calibrated from 0 to 120 mph in the United States. A federal law was passed in 1975 making 55 mph the speed limit on interstate highways. The motivation of the law was to reduce gasoline consumption and highway deaths. Automobile manufacturers responded by changing the speedometer so that "55" was placed about two thirds of the distance from one end of the meter to the other, with the upper limit shown as 85, instead of 120. When the car exceeded 85 mph, the needle continued to spin past the visible upper limit. (The national speed limit was later rescinded because of the increased cost of traveling slower across large expanses of the country.)

Parallax Effect

Parallax effect is the extent to which a reading of a display becomes distorted if the operator is viewing the display from an angle instead of head-on. Parallax effects are likely when the display involves the use of a needle that rides above the calibrated face panel, as in the example in Figure 4.7.

A common example of the parallax effect occurs in automobile speedometers. The driver sees one speed, but the front seat passenger sees another and asks, "How fast are you going? It sure looks faster than that from here." The design solution, which may have been inadvertent, was to shape the dashboard so that the passenger cannot see the speedometer.

Color and Contrast

Color is sometimes helpful, but there is a temptation for designers to apply it frivolously. For the most part, higher contrasts between the image components will produce better human performance, particularly among older users (Kantowitz & Sorkin, 1983). Black and white elements have the largest contrast. Color should be used only if it conveys actual meaning. The starfield display attempted to use color in that fashion to represent components of a system.

Yet another speedometer story is applicable here. The speedometer in the 1960 Oldsmobile was laid out in a horizontal line from 0 to 120 mph. The indicator was a light that expanded from 0 to the indicated speed; this arrangement is not too different from the lighted display in Figure 4.7, except that it was horizontally oriented and changed colors. The indicator light, however, changed color from green to orange at 35 mph, and from orange to red at 65 mph. The contrast between the green and orange against the dark gray background was perhaps pretty equivalent, but the red might have blended too easily into the background. This was not a particularly good design choice for high-speed travel in a car that was equipped with an overdrive transmission (some models only). On the other hand, the design idea did not seem to last long enough to support published usability studies.

Historical and Predictive Displays

Historical displays tell the operator something about the activity of a system over a period of time. Figure 4.13 shows two classic designs. In the upper left, a calibrated paper is placed over a drum. A pen marks the numerical value of the system as it changes over time. The paper in this example is changed every 24 hr, stored, and possibly examined later. A possible limitation to easy reading is that the single line of information is wrapped around a 3-D drum, thus requiring a bit of depth correction by the operator who is trying to read it.

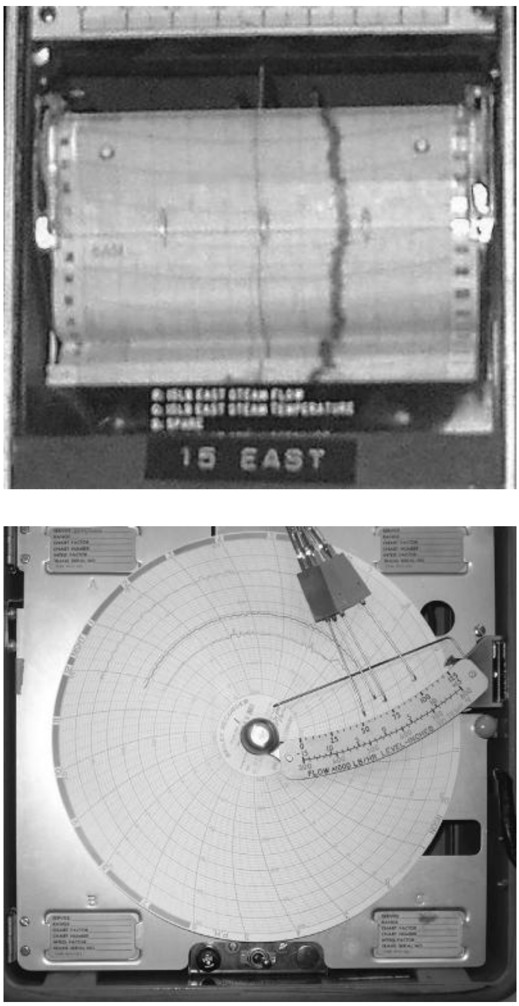

FIGURE 4.13
Historical displays.

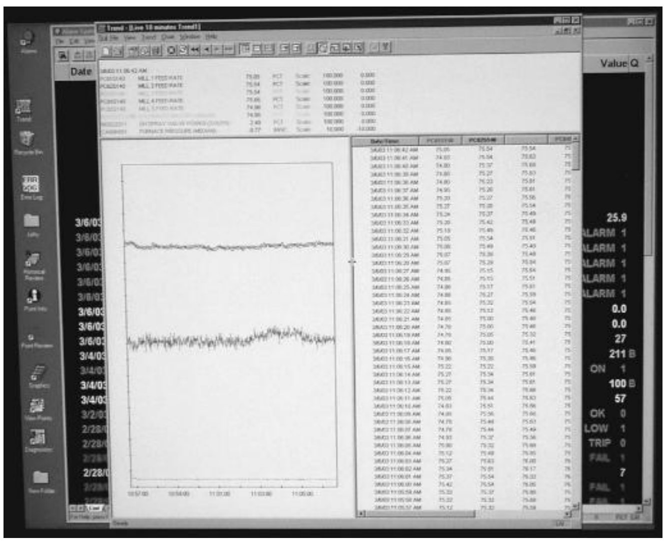

FIGURE 4.13 (Continued)
Historical displays.

In the lower left of Figure 4.13, a three-color pen assembly is recording values of a system on a rotating disk. The radius of the circle distorts the scale of the values; it is not possible to put all four indicators on the outer edge. On the other hand, the depth-perception conversion is not necessary, and the new values are centrally located on the display.

The disc design for a historical display is not appreciably different from the tachometer recordings on delivery trucks that were used in the mid-1970s. A tachometer would show the speed of the truck over an 8-hr period. Someone reading it could tell when the truck stopped and for how long while making deliveries. If a fleet of trucks was involved, which was usually the case, one would have to have the drivers' itinerary in hand to interpret the starts, stops, and movements.

In the upper right of Figure 4.13, a power plant operator is recording system values on a clipboard for other operators to examine later if necessary. This procedure, dated to the mid-1960s, was obsolesced by computer-based data-acquisition methods. An example of a computer-based display is shown in the lower right of Figure 4.13. The screen facing the viewer gives both a graphic and a numerical history of a system sensor. The historical

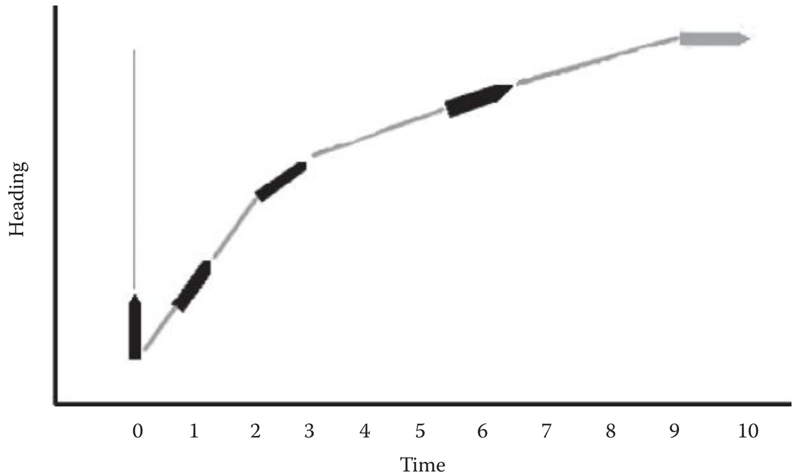

FIGURE 4.14
Basic format for a predictive display.

information screen is a pop-up window that is accessed by pointing a mouse cursor to a line of information about a sensor on an alarm screen. The alarm screen, which is occluded by the historical display in this photograph, uses color coding to denote the levels of criticality associated with the sensors.

A predictive display conveys the state of a system at a future point in time. Predictive displays may be necessary in slow-acting systems, such as the steering mechanisms of ships, in which there is a substantial delay between the time of the control action of the operator and the time the result is achieved. Slow-acting systems may induce the operator to make the control operation a second time and thus do too much of an otherwise good thing. The layout shown in Figure 4.14 is a generic display design and not a system in actual use.

3-D Displays

Artists have been able to deliver high-fidelity renderings of 3-D information on a two-dimensional canvas for a very long time; the skill set expanded during the Renaissance. The advent of photography in the 19th century made another leap forward in this regard, as did television nearly a century later. One must not forget the use of sculpture, which dates back millennia. An actual 3-D display, however, is not a rendering of an image to a two-dimensional medium; it is a rendering of a 3-D image. For instance, the Brewster stereoscope and movies that allowed the viewers to experience bats flying off the screen into their faces serve as the prototype.

The technology for generating 3-D displays in line-graphic or photographic form became available in the late 1970s. A review of the first stage of research indicated that 3-D displays were better than 2-D for displays involving terrain maps and perhaps other real-world situations where the original source is 3-D (Getty, 1983). Three-dimensional imaging reappeared in the 1990s when computer display graphics technology improved and new needs were identified. For instance, projections of the human brain can be compiled from a functional magnetic resonance image (fMRI). They are 2-D displays in the sense that they convey depth through monocular depth cues rather than binocular disparity, although they are generated from true 3-D source information. Nonetheless, 3-D monocular images have considerable value in biomedical applications and elsewhere.

Three-dimensional displays also reemerged in the 1990s in conjunction with virtual reality applications where the user is actually immersed in the visual image, and not just looking at it (see Chapter 11). Other technologies have become available since that time that are intended in part as a supplement to virtual reality construction; although some require special glasses one technological goal is to make the glasses unnecessary. Potential applications to toys and games proliferate. Work-oriented applications are often centered on terrain-based applications, such as aviation guidance, oil and gas exploration, telerobotic control, and visualization of room interiors.

There are some known assets and limitations to 3-D displays, according to Park and Woldstad (2000). On the one hand, it is possible to perceive features of the image that would not be apparent in a 2-D rendering. On the other hand, perceptual ambiguity along the line of sight has been documented. The interpretation of one visual feature may be distorted by variations along a different visual parameter. Some of the perceptual ambiguity can be corrected, however, by adding graphic enhancements to the 3-D display such as calibrations and other symbolic markers.

On the frontier of 3-D visualization technologies, *volumetric displays* (will eventually) allow users to walk around the display and view all sides of the object (Figure 4.15; NEOS Technologies, 2004). The object is projected into the center with a system of laser beams (Nayar & Anand, 2007, 2013). A possible application could include air traffic control, according to NEOS Technologies, where controllers could visualize airplanes swarming about in a contained simulated space. Some prototypes made by different manufacturers actually exist, with miniature buildings, toy characters, human heads, abstract geometric objects, or luminescent bunny-rabbits projected into the display bulb as demonstrations. Another format that is undergoing active experimentation would project the object into a free-standing space. Thus, a person could grasp an object that looks as though it is present

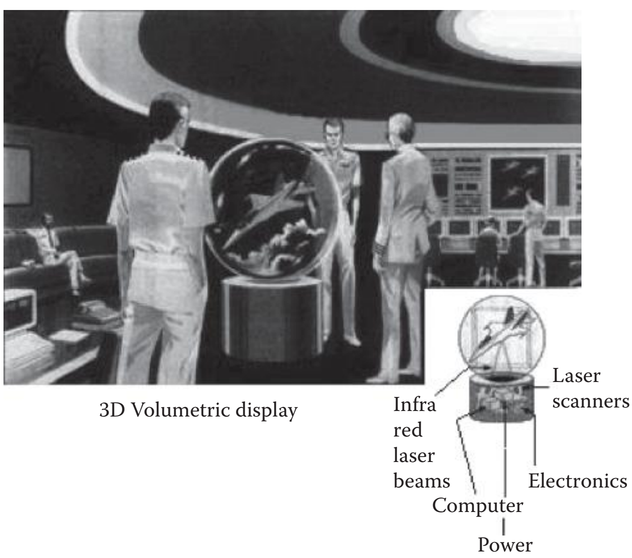

FIGURE 4.15
Volumetric display prototype. (Reprinted with permission of NEOS Technologies.)

but really is not, enjoy a hologram stage show, or encounter a salesperson that pops up out of nowhere in front of them and gives a sales pitch.

The technology is still too raw to support meaningful usability studies. If the goal were to portray real-world events as they occur, such as air traffic, the sensor system would need to be very sophisticated, and the time that would be required to compute the holographic image might be too long to be useful. For the pop-up people paradigm, is the user the sender of the message, or the unsuspecting shoppers who could run into a pop-up everywhere they turn?

Digital Versus Analog Displays

Perhaps the best-known example is the clock. There are the round ones with the hands, and the varieties that specify the time in digits. Which is better? If it is necessary to see how close one is getting to a critical time, such as the hour or half-hour, then the round clock is usually preferred because it automatically makes the subtraction. On the other hand, if the specific time is needed, perhaps in conjunction with programming an appliance, then the digital design is preferred. The example display in Figure 4.16 utilized both

FIGURE 4.16
Analog and digital cockpit display panels. (From National Aeronautics and Space Administration, *Lesson 2: Technology interface design*, retrieved March 24, 2004, from http://humanfactors.arc.nasa.gov/web/hfl01/technology.html, 2004, in the public domain.)

analog and digital representations of information; perhaps the operator relied on both forms of representation for different tasks.

For many years the contrasts between analog and digital displays began and ended with the different uses of clocks, which applied fairly well to design choices in other situations. Developments in computer graphics in the late 1990s, however, have made a significant impact on the design features of airplane cockpits. Here, digital is not confined to showing the digits; rather, it refers to the mechanism behind the production of the display.

Figure 4.16 contrasts a strictly analog cockpit design with another that is equipped with digital displays (from National Aeronautics and Space Administration [NASA], 2004). According to NASA, the new digital display designs allow the pilot to zoom in on legends and details for better legibility. It is possible to produce graphic information as well as numeric information. Color coding can be utilized to distinguish different types of related information such as flight paths and traffic flows.

Multifunction displays facilitate a less cluttered control panel, as NASA (2004) also noted. It remains an open question, however, whether stacking information from two or more individual displays into one display is always beneficial; see the section of Chapter 6 on "mode error."

Heads-Up Displays

Machine displays often compete with the real world for the operator's attention. There is a limit to how much attention a driver can divert from the road, or the pilot from the skyscape. Hence, the heads-up display (HUD) puts critical information in front of viewers' eyes at all times through use of a specially devised helmet and visual screen. Alternatively, the heads-up image can be projected onto the windshield of the automobile or airplane. The electronic trick, however, is to project the image in such a way that the operator does not loose the sense of depth and spatial information that usually comes in from outside the windshield. In other words, HUDs can produce attentional tunneling, which must be assessed in any HUD design (Tufano, 1997). When properly projected, however, they can reduce the ocular readjustments that the pilot makes when viewing a close-up control panel and a full-depth view through a windshield (Fadden, Ververs, & Wickens, 2001). A windshield-mounted HUD for pilots is depicted in Figure 4.17 (NASA, 2004).

There is usually a trade-off between the clutter that occurs when too much information is placed in the heads-up display and the scanning-time cost associated with the heads-down (standard control panel) display. Although many HUDs that have been tested survive the trade-off, far-field vision becomes compromised when clutter becomes too great (Yeh, Merlo, Wickens, & Brandenburg, 2003). The difference between valuable information and clutter depends in part on what is task relevant (Hammer, 1999).

Display Panel Organization

At this point, it is possible to state some broad rules for organizing a panel of displays (Kantowitz & Sorkin, 1983). The first is to organize the displays according to importance. How critical or central is the information from a display to a task that the operator is trying to perform? "Importance" is a vague term, however, although it is usually equivalent to frequency of use.

Second, the most frequently used, or otherwise important, displays should be positioned centrally in the field of vision. The constraints for the field of vision were defined earlier in this chapter.

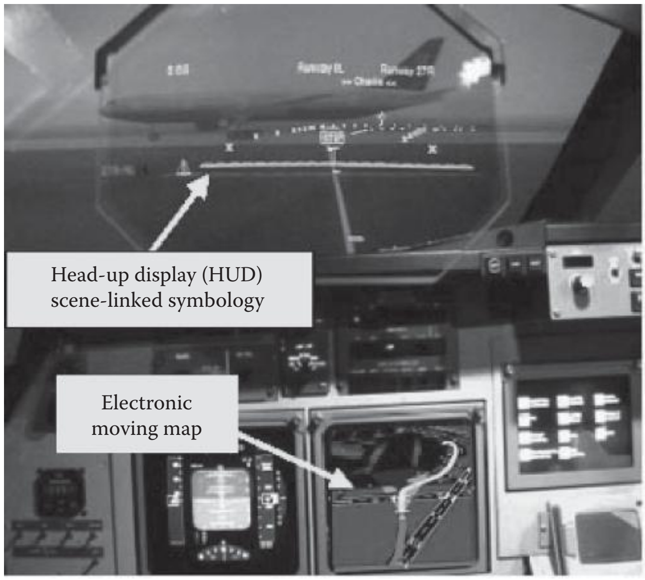

FIGURE 4.17
Heads-up display for aircraft windshield. (National Aeronautics and Space Administration, *Lesson 2: Technology interface design*, retrieved March 24, 2004, from http://humanfactors.arc.nasa.gov/web/hfl01/technology.html, 2004, in the public domain.)

Third, displays should be organized by order of use. If a task involves several sequential steps, the arrangement of displays should follow the same order.

Fourth, displays that are used together should appear together. For instance, in the case of the set of 2-D displays that convey a 3-D image as mentioned earlier, we would not want to separate the set of 2-D images.

Fifth, displays should be grouped by function. If a combination of displays pertain to one function of the system, they should not be mixed up visually with displays pertaining to another function of the system.

Frequency of use, order of use, and togetherness can be determined empirically through linking analysis. One would need to set up a study, perhaps with eye-tracking equipment, whereby the movements of the operator are examined as the operator views one display after another while performing a standard task. How often does the use of one display follow the use of any other display? The most frequently appearing combinations of displays should be given the most central location on the control panel. The displays that are used less often should be located more toward the periphery, and in reasonable proximity to the displays to which they are sometimes linked.

Eye-tracking equipment requires head-mounted sensors that follow saccadic eye movements. Sample output appears in Figure 4.18. The operator is actually driving an automobile. If all the displays of concern appear on a computer screen, eye-tracking equipment is now available that can capture eye movements without a head-mounted device. A sensor is placed at the bottom of the computer screen. The operator stares at it until the sensor calibrates the eye positions. It then follows the eyes as they move around the displays. The movement information is stored on a second computer and does not interfere with the test display or related operations.

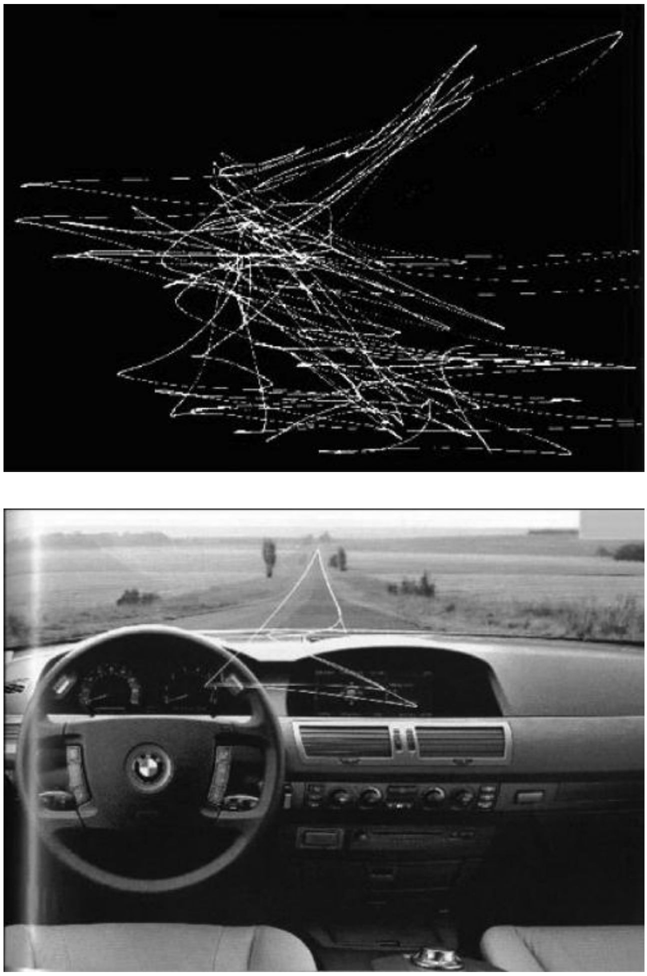

FIGURE 4.18
Saccadic eye movements while driving a motor vehicle. (Reprinted with permission of BMW, Inc.)

Signs of Importance

Signs are static visual displays. The concern here is for their information-conveying quality, and not for the impact of decorative signs or advertising logos. There is a line of reasoning, nonetheless, that aesthetics affect usability because an aesthetically attractive display induces the user to pay more attention to the sign or device.

Design

Pictorial signs are particularly useful in multilingual environments such as a European railroad system. A good design will convey a concrete noun or verb rather than an

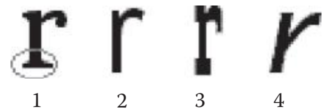

FIGURE 4.19
Serif on typographical characters.

abstraction. The design should capitalize on figure–ground distinction and the ease by which one sign image is distinguishable from another.

Verbal signs for instruction or emergency information should be prepared with clear and unembellished lettering. Sans serif type styles are preferable to styles with serifs. A comparison of four type styles is shown in Figure 4.19 that featuring the lowercase letter *r*. The example designated at 1 is Roman-style type that is commonly used in book print. The serif is the little flanges at the endpoints of the letters. The serif in 1 is circled. Example 2 is Helvetica type, which is a popular choice for figure captions and headings in book print as well as for informational signs; there is a family of related type styles that are equivalent with regard to their lack of serifs, hence sans serif.

Example 3 is a version of the famous Western (U.S.) style of type used for the word WANTED in a Wanted poster. It still attracts attention today, perhaps more because of its nostalgic reference than because of its clarity. Example 4 is a decorative sans serif type; because of its lack of simplicity, it is not recommended for informational or emergency signs. Further digressions into the dynamics of type styles are better reserved for the experts on typography (e.g., Spencer, 2004).

In any case, the width/height ratio of lettering on informational signs should stick close to a 3:5 ratio. The signs should be visible from a realistic distance, which means that they should be large enough overall. The ambient lighting should be sufficient to read the sign.

Standards

Some standards exist in the United States for certain types of signs placed in public situations. The American National Standards Institute (ANSI) has devised a standard sign for DANGER advisement. The danger sign appears in Figure 4.20. The upper portion is a black field with a white oval encompassing a red oval. The red oval contains the word *DANGER.* The lower portion is a white field with the name of the danger in black lettering. If additional explanations are needed, they should be placed in black on a white field to the right of the main part of the sign.

An ANSI standard caution sign appears in Figure 4.20 also. It is characterized by a bright yellow triangle containing a large exclamation point that appears to the left of the word *CAUTION.* It is followed by an explanation in black on a yellow or white field. The glyph with the triangle and exclamation point carries over to instruction manuals and related texts. An older style caution sign that might be found on some equipment still in existence is a yellow field with thick black spines around the perimeter. The text of the message is displayed in black beginning with the word *CAUTION.*

OSHA (1996) adopted standards for danger and caution signs in the workplace. The basic shape is square with a black border. The danger sign has a wide red stripe across the top with the word *DANGER* written in white. The description of the danger is written in black on a white field in the lower portion of the sign. The caution sign has a bright yellow stripe on top with the word *CAUTION* written in black. Again, the nature of the

FIGURE 4.20
Danger and warning signs: danger sign in a public situation, currently used caution sign in a public situation, danger warning on a product label earlier design for a caution sign in a public situation.

warning is written in black on a white field in the lower portion of the sign. OSHA also has companion designs for safety (with a green stripe) and notice (with a blue stripe) signs.

The motivation behind the designation of standard sign formats is to condition an instant recognition of a danger situation on the part of the viewing public. Other regulating entities have developed sets of standards as well. The U.S. Department of Transportation (2003) issued revised guidelines for the design and production of highway signs. Not only do they follow standard production formats for symbols, colors, numerals, and verbiage, but there are also specific sizes required for different roadways depending on the travel speed. The viewing distance must therefore be long enough for the driver to perceive the sign, register its meaning, and execute an appropriate action safely.

OSHA or ANSI designs are not always sufficient for warning labels that are placed on consumer products. Nonetheless, the standard designs are recommended for their recognition value. An example of an ANSI danger warning from a 6-ft stepladder also appears in Figure 4.20. Note the use of pictorial information that is placed in the box for additional explanations. Note also the angular drawing with heavy lines and the world-famous “do not” ring. Caution and danger warnings on consumer products play an important role in the threat of product liability litigation, as mentioned in Chapter 1.

Behavioral Impact

Although voluminous research was performed long ago on the legibility and intelligibility of signs and labels, there is comparatively little information about their behavioral impact. One classic study that was conducted in a British steel factory (Laner & Sell, 1960) indicated that safety signs (with pictures) that stated the proper safe behavior were more effective at reducing a particular type of accident than signs that contained only generic safety awareness information. The rule of stating the proper safe behavior carries over to roadway safety signs also (Lonero et al., 1994).

In broader situations, people who might be prone to risk taking are making a trade-off between the benefits of breaking the rule and the likelihood of a dangerous outcome. The benefits of breaking a rule, or *counterdemand,* may be matters of convenience such as getting a job done faster, minimizing physical discomfort, or avoiding the task of figuring out an alternative method for performing the task.

Illumination and Glare

Proper lighting is required to view any visual display or sign or to find the correct controls. The light may be part of the display itself, which is a good idea if the display is used in an environment that is sometimes dark. Environmental lighting in the room should be sufficient for normal working conditions.

Illumination

Lumens are the measurement of the total amount of light given off by a source in all directions; lumens are closely related to the wattage of a new lightbulb. *Candelus* is the measurement of light in one direction of the light source. *Illuminance*, or light reaching a surface of an object, is measured in *footcandles*, which is the number of lumens per square foot of surface. The *IES Lighting Handbook* (Illuminating Engineering Society, 1981) gives recommended lighting levels for a variety of work situations. In general, if there is relatively less contrast in the coloration of the work source or relatively small objects are involved, then relatively more ambient lighting is required (Kantowitz & Sorkin, 1983).

It is possible to conduct signal detection studies to determine the minimum lighting needed to read a display or do a piece of work. Sign legibility is optimal, however, at luminance contrasts of 10:1 between the target letters and background. Legibility actually tapers off by a small amount at greater contrast ratios (Sival, Olson, & Pastalan, 1981).

Lights that are not intentionally manufactured to produce colored light often produce almost-white light with hues of yellow, blue, or pink, depending on the materials involved. These trace amounts of color present little problem in most types of work, but an industrial color-matching task may be severely affected by the exact type of lighting used. If the manufacturer of a color-specific product and the customer do not agree to the reference lighting standard that they would be using, disagreements can ensue regarding the quality of the product. (This writer has seen it and heard it.)

Urban streetlights often give off a colored glow. Darnell (1997) reported street lighting in San Jose that was so yellow it could be confused with the amber light of a traffic signal. He wrote, "Imagine you are driving along at night and don't notice the yellow lights ahead are yellow traffic lights. You might be very surprised when they turn red."

Glare

Glare is the amount of light deflecting from a surface into the viewer's eyes. Discomfort glare is the amount of subjective annoyance experienced by the viewer. Disability glare is the reduction in visibility caused by the deflecting light source. Deflection patterns differ depending on the smoothness of the deflecting surface.

Glare can be reduced by placing the light source high over the viewer's head and behind the viewer. If the glare source is located too close to the viewer's line of sight and is located between the viewer and the surface, disability glare increases sharply (Sanders & McCormick, 1993). Typical solutions involve moving the work station relative to the ceiling lights, changing the location of the light sources, and working with glare-controlling surfaces.

Some standards from the Department of Defense that would carry over to many other environments include the following: (a) "Surfaces adjacent to the display screen shall have a matte finish." (b) "The luminance range of surfaces immediately adjacent to display screens shall be between 10 percent and 100 percent of screen background luminance." (c) "In vehicles, ships, observation structures, and other spaces where users may need to be dark adapted (e.g., cockpit, ship bridge, vehicle cabs) the interior walls or bulkheads shall be a matte black. This will minimize specular reflection from interior light sources."

DISCUSSION QUESTIONS

1. Mirrors are often used to provide a display in locations where the actual events are not visible. When are mirrors valuable? What types of human perception problems can result from the use of mirrors?
2. Besides clocks, what other types of information can be represented with both digital and analog displays? When would one type of information be preferred over another in those situations?
3. Design a prototype experiment that would determine if there is too much clutter in a visual display. Describe some display design techniques that would reduce

clutter if it were necessary to do so. Do the clutter-reducing designs suggest the potential for other types of display or control problems?

4. Suppose we had a group of displays like the one shown in Figure 4.21. The percentages given in the diagram represent the proportions of times the operator views one display after viewing another. How should the display panel be reorganized if the operators spent 48% of their viewing time with C, 10% with F, and 26% with E? What if the operators spent more time viewing E after viewing A more often than they spent viewing B after viewing A?
5. What would be some plausible new uses for volumetric displays once the technology has been perfected?
6. Figure 4.22 depicts a set of displays for chiller units in an HVAC system. What is wrong with this picture? (Hint: The photo was taken without a flash.)

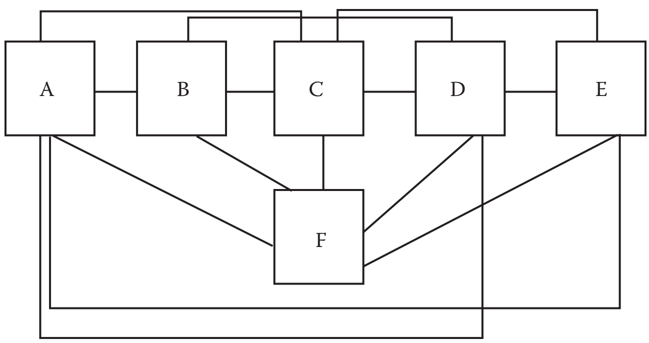

FIGURE 4.21
Hypothetical set of displays and links.

FIGURE 4.22
Display panel for a set of HVAC chiller units.

5

Auditory and Tactile Displays

This chapter describes the central principles for designing effective auditory displays. It begins with a synopsis of the human auditory system and principles of sound perception. Expansions on these principles can be found in a standard text on general psychology and in advanced texts on sensation and perception. Of necessity, the scope of this chapter is confined to the principles that are most closely related to the design of human–machine systems.

Auditory displays for conventional human–machine systems, which can be nonverbal or speech displays, are considered next. These principles carry over to computer-based systems as well, although once again the computer presents some new challenges.

The second to last section of this chapter pertains to noise and noise control. It would be equally feasible to consider the noise as a topic in workspace or environmental design. For present purposes, however, the full value of an auditory display cannot be fully appreciated or evaluated without some concern for the environmental context in which it is heard. The final section of this chapter describes some practical uses of the sense of touch.

Sense of Hearing

Sound is a vibration that creates disturbances in air pressure. Its three parameters are loudness, pitch, and timbre. A diagram of the interior of the human ear appears in Figure 5.1. The sound vibrations enter the side of the head where the cotton swabs usually go. It travels down the auditory canal until it reaches the eardrum. The eardrum vibrates, and in doing so transfers the vibrations to three little bones known as the hammer, anvil, and stirrup. The stirrup connects and transfers vibration to the cochlea. The cochlea is filled with fluid and contains the basilar membrane. The basilar membrane contains the receptor cells, which are also known as hair cells because of their antenna-like hair that protrudes and detects the exact nature of the incoming vibration.

The frequencies of the incoming sound are processed on the basilar membrane in terms of both the threshold of frequency that triggers a cell's activation and the location on the basilar membrane that is responding. High frequencies are registered in the area of the cochlea that is closest to the connection to the middle ear; low frequencies are registered further away. The basilar membrane does not contain enough cells to permit a one-for-one correspondence between all the shades of frequency that a person can hear and particular cells. Thus, volley theory is the prevailing explanation for how we can hear so many shades of frequency. The particular frequencies we experience are a result of a pattern of cells firing; the specifics of the pattern have not conclusively determined yet, however.

The responses of the hair cells are gathered together by ganglia in the cochlear nucleus and transferred to the auditory nerve, which connects to the auditory cortex in the temporal

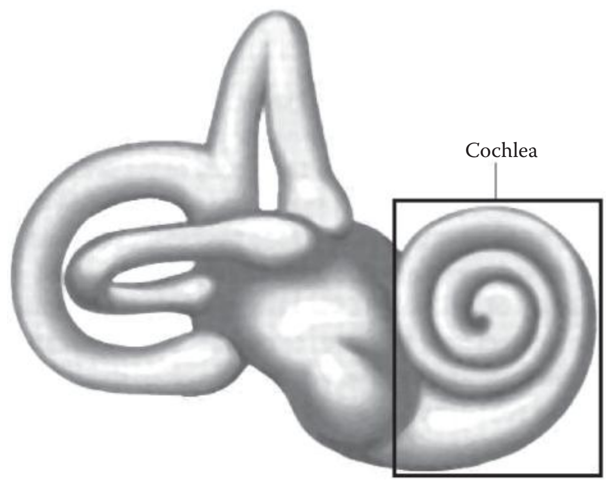

FIGURE 5.1
The inner ear.

lobe. The auditory cortex is conveniently located close to the thalamus and speech interpretation area of the brain. It consists of three components: the core, the belt, and the parabelt (Snyder & Alain, 2007). The belt surrounds the core, and the parabelt is off to one side of the belt. The three components of the auditory cortex are responsible for the early, middle, and late stages of auditory processing, respectively. Sounds are encoded by frequency when they leave the cochlear nucleus, and frequency information is processed first. Later stages process harmonics (see below), temporal patterns, and location.

There are connections running in both directions between each two of the three components. There are further connections in both directions to the cortex and thalamus. The cortex contributes interpretations of the sound and directs attention to the incoming sound source. The thalamus coordinates sounds with visual cues and psychomotor responses.

Loudness

A sound is a wave, not unlike the one shown in Figure 5.2. A wave has an amplitude, a frequency, f, and a wavelength, λ, such that $f = c/\lambda$, where c = speed of sound, or 720 mph. The amplitude of the sound wave produces loudness.

Physical loudness of a sound is measured in decibels (dB). A whisper produces a sound of roughly 20 dB. A typical residence with the television off generates sound of about

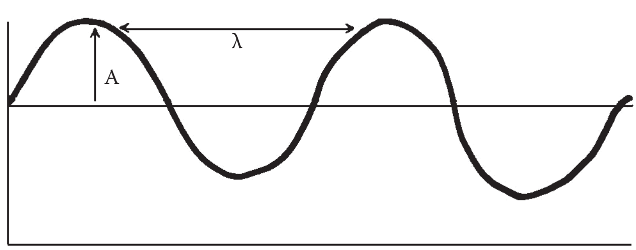

FIGURE 5.2
Basic soundwave showing amplitude (A) and wavelength (λ).

30 dB. An office is usually producing sound in the range of 40 to 55 dB. A noisy city intersection is closer to 70 dB. A train produces sound close to 100 dB in the near field of the track. Aircraft typically produce sound in the 125 to 130 dB range.

If we were to ask people to increase the loudness of a standard tone until it is "twice as loud as" a reference tone, they would give us an increment of about 10 dB. The logarithmic relationship between physical loudness and psychological loudness should be apparent. Psychological loudness is measured in sones, *S*, such that (Kantowitz & Sorkin, 1983)

$$S = 10^{(0.03\text{dB}-40)}. \tag{5.1}$$

Our perception of loudness, however, is dependent on the frequency of the tone. Tones with high frequency have shorter wavelengths, and thus do not travel as far and fade out a bit before they reach our ears. An alternative measure of psychological loudness is the phone. It is measured the same way as the sone, except that it is figured relative to 1000 Hz (cycles per second).

Pitch

Pitch is related to the frequency of the sound wave. The tone known as A above middle C on a piano (for those who are familiar with this device) is tuned to 440 Hz. An A that is one octave above the reference A is generating a frequency of 2 × 440, that is, the reference frequency is multiplied by a factor of 2. An A that is one octave below the primary A is generating a frequency of 1/2 × 440, that is, the reference frequency is divided by a factor of 2.

In Western music an octave is divided into 12 intervals of equal size, known as semitones. Thus, the frequency of any note is the 12th root of 2 (approximately 1.06) times the frequency of the note immediately below it. (Note that this explanation is an oversimplification for some musical instruments. The note known as G-sharp is the same note on a piano as the note known as A-flat, but on a violin, the G-sharp is a little sharper and the A-flat is a little flatter than the piano rendition of the notes.) Eight of the 12 semitones comprise the common diatonic musical scale. Although the remaining four semitones may be used generously in a musical composition, they are interpreted by the listener as members of a different scale, or accidentals.

Music of other cultures may divide the octave into more intervals or fewer intervals. A guitar allows us to bend the strings to reach the quarter-tones between the semitones. The point for our purposes, however, is that listeners have expectations of what tonal sequences are considered pleasant, unpleasant, familiar, or surprising. These expectations may carry over to the interpretation of auditory displays that are based on tonal sequences.

The perception of pitch has a well-known quirk. If we were to present listeners with two tones an octave apart, they would have little difficulty determining which tone was a higher or lower frequency. At the same time, they would recognize the two tones as the same note from the standard scale. If we were to present the listeners with two tones that were just one or two semitones apart, they would once again detect which had the higher frequency without much difficulty. On the other hand, if we were to present two tones that were five semitones apart (or a half-octave) confusion would set in as to which was the higher note (Deutsch, 1992; Shepard, 1982).

The *pitch helix* (Figure 5.3) depicts these pitch-perception relationships. Although there is a psychological linear relationship between actual and perceived pitch from one octave to another, there is a nonlinear involution in the middle of an octave. Thus, auditory display

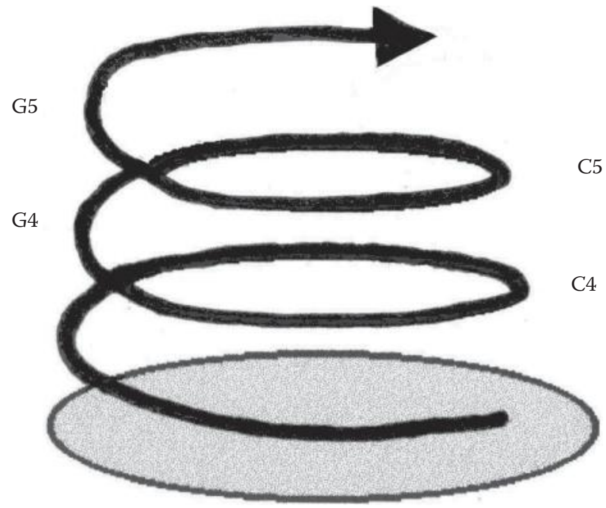

FIGURE 5.3
Pitch helix based on D. Deutsch (The tritone paradox: Implications for the representation and communication of pitch structures, in M. R. Jones & S. Holleran (Eds.), *Cognitive Bases of Musical Communication*, pp. 115–138, American Psychological Association, Washington, DC, 1992, with permission) and R. N. Shepard (*Psychometrika, 27*, 219–246, 1982, with permission).

systems should not require the operator to make any important decisions based on certain combinations of tone differences. Try another means of person–machine communication.

To make matters more interesting, if we generated two tones and captured the resulting sound wave on an oscilloscope, the resulting wave would take on a complex form that would be a merger of the two fundamental waves. The listener would hear not only the two fundamental tones, but also a third tone that is the sum of the two frequencies, and a fourth that is the difference of the two frequencies. The foregoing does not yet include the influence of harmonics.

Timbre

Why does the same musical note sound different on a violin, flute, or trumpet? One might immediately point to the materials used to make the instruments and the shapes of the instruments. Indeed, shape and material are relevant because they produce different series of harmonic tones. *Harmonics* are integers multiplied by the fundamental frequency to produce additional frequencies. Different musical instruments produce different harmonics and emphasize them to varying extents. *Timbre* is the set of harmonics associated with a sound source.

The various harmonics of a musical instrument are heard together in the form of different characteristic sound waves; some examples appear in Figure 5.4. Flutes produce sinusoidal sound waves. String instruments produce saw-toothed waves. Brass instruments produce square waves. Percussion instruments have noninteger harmonics.

The wave in the lower right of Figure 5.4 is a generic sound wave that forms the starting point for digital sound synthesis. The generic wave is divided into eight steps. The eight steps are allocated among three parts of the wave: the attack, the sustain, and the decay. Each step can be manipulated for relative amplitude and relative pitch. In this way any

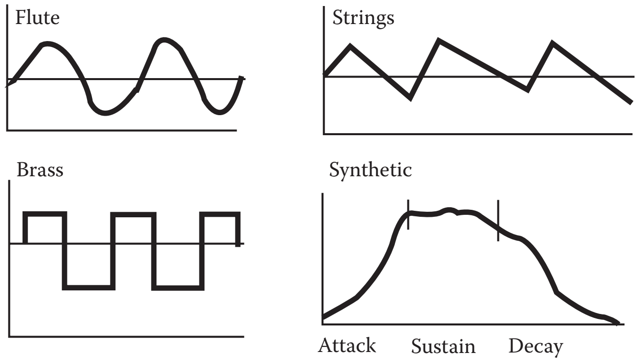

FIGURE 5.4
Typical waves for musical instruments.

musical instrument can be mimicked in principle, and unique sounds can be generated as well.

Even with a well-designed interface, the demands on the user for sound wave knowledge and memory can be challenging. Thus, the second generation of sound synthesis equipment made much greater use of sampling techniques. A recording of the sound would be entered into the machine. The machine would take a statistical sample of the wave and reproduce it. Higher fidelity replications require higher rates of sampling.

Binaural Hearing

The direction of a sound is detected by virtue of having two ears. The process is akin to binocular disparity in depth perception. The sound reaching each ear is not quite the same, and based on that difference, the listener can locate the sound. A directional sound reaches one ear about 5 ms before it reaches the other ear. The sound waves reaching each ear thus become slightly out of phase, but only enough so as to allow the detection of direction. If the delay should become as long as 20 ms, the listener will perceive two separate sounds (Kramer, 1994).

Regrettably, humans' detection of sound sources is not entirely accurate. Vertical direction is more difficult to determine than horizontal detection. To make matters more challenging, sounds in the real world bounce off sides of buildings or mountains or off interior walls. Thus, the listener may be getting sound from reflected sources rather than original sources, or from both sources at the same time. Anyone who has ever driven an automobile through a city and wondered where the police siren was coming from can vouch for the complications that could exist in locating a sound source. Contemporary concert halls, on the other hand, are engineered so that almost every listener in the audience will receive the complete sound generated from the stage with a minimum of distracting reflected sounds and with minimum loss of sound information in any one seat (Kantowitz & Sorkin, 1983).

Recorded sound during the first half of the 20th century was produced monophonically, meaning that even high-fidelity recordings gave the spatial impression of coming from a single hole in the wall. High-fidelity recording allowed the reproduction of a wider range of frequencies and harmonics, which were intrinsic to the development of stereophonic sound. Stereophonic sound, which was developed in the 1950s, required at least two sound sources, for example, two microphones strategically positioned in an array of musical instruments. Different sound patterns reached each microphone. Two speakers were required for proper playback. The result was that the listener could detect where the sound elements were located in 3-D space. For any room equipped with a stereo system, there is an apex point in the room where the detection of full stereophonic sound is optimal.

Nonverbal Auditory Displays

Auditory displays can be conveniently grouped into two types: those that contain speech and those that contain only nonverbal sounds. The nonverbal types will be considered first because they are more common in conventional person–machine systems and because the basic properties of the nonverbal displays carry over to the understanding of what makes a speech display work well.

According to the classic wisdom (Deatherage, 1972, p. 124), auditory displays are preferable to visual displays when the message is simple, short, or when the operator will not be required to refer to it later. They are particularly good for describing events in time or for calling an operator to immediate action. They are viable options when the visual system is degraded or overburdened. They can reach an operator who must move about continually.

In contrast, visual displays are preferable to auditory displays when the message is complex or long, or if the operator must refer to it later. Visual displays are preferable for conveying information about the location of objects in space and for when an immediate action is not required. They are preferable to auditory signals when the auditory system is degraded or noisy. Visual displays assume that the receiver remains in one position more or less.

There are two caveats to the foregoing recommendations. First, contemporary digital methods for generating auditory displays have triggered some experimentation for using complex sound patterns to depict complex visual information such as a chart of data from a research report (Kramer, 1994; Madhyastha & Reed, 1994). Such interfaces can be used in automatic reading devices for the blind, but they are also intended for sighted people who might be able to detect patterns in their data that might not be discernible visually.

Second, the mobility of the receiver can be managed with print documents in some cases. Laptop computers address the same issue posed by the desktop or mainframe computer. Anyone who has taken a romp through an airport will remember multiple panels of monitor screens displaying flight time arrivals and departures. People do need to refer to that information more than once if a flight has been delayed, and delayed, and delayed again. The auditory system in airports can easily become degraded with noise or overburdened with too many flight numbers, gates, and departure time updates spoken by multiple airlines workers.

Types of Nonverbal Auditory Displays

Conventional auditory displays are usually produced as single tones, patterns of two or more tones, or tones that have a more complex wave envelope. Examples of the latter variety might be characterized as a wail or yeow. A beep is a single tone displayed repeatedly with a clear definition between onsets; the beeping sound continues until an operator takes action, or the alert condition has otherwise ended.

Which type of display is better? Part of the answer lies in how many auditory displays are already in the system. Another part is connected to the intensity of the sound. Louder sounds prompt faster response times (Adams & Trucks, 1976), although particular sounds may be better than others at specific intensity levels. If in doubt, which may be often, a response–time experiment should be conducted where the operator is expected to produce a realistic response to the stimulus.

Masking experiments provide another means for determining the comparative effectiveness of auditory display designs (Teas, 1990). In the typical design a display is presented for a brief period, before a noise source is introduced. The noise source may be white noise, which contains a balanced representation of frequency bands within the audible range, or a conflicting sound that is more natural to the listeners' environment.

Three factors are involved in the detection of signals: the signal-to-noise ratio, the degrees of spatial separation, and the correlation between signal and noise sources (Doll, Hanna, & Russotti, 1992). As discussed in Chapter 3, louder signals relative to noise are more readily detected. Spatial separation is the extent to which sounds come from different sources. Greater separation usually allows the listener to attend to one source and ignore another. Alternatively, spatial separation allows the listener to monitor the secondary sound source in case anything interesting is happening.

Correlation in this context refers to the similarity of pitch and tone patterns in the signal and noise source. A pulsing beep signal against a background of speech babble would produce a low correlation, and greater detectability. A speech signal displayed against a background of speech babble would be highly correlated and thus less detectable. This concept of correlation is also known as *articulation* or *dissociability*; it is an auditory analogue to figure–ground distinction in visual perception (Sanders & McCormick, 1993; Williams, 1994).

According to the U.S. Department of Defense (2012) and U.S. Air Force (Doll, Folds, & Leiker, 1984), speech displays are superior to tones or complex sounds if auditory displays are to be used to convey quantitative, qualitative, or tracking information. Nonverbal displays are superior to speech when trying to convey status information. These trends were based on the assumption of relatively simple tonal displays rather than the complex sound patterns that can be produced through digital sound synthesis. Auditory displays that have been designed to convey trends in numerical data graphs can produce human interpretations that are equivalent in quality to interpretations of visual graphs (Flowers, Buhman, & Turnage, 1997). The applications to data displays often use pitch and loudness to represent independent facets of the data; sometimes interactions are detected whereby a particular pitch level will affect the perception and interpretation of loudness (Neuhoff, Wayand, & Kramer, 2002).

Gestalt Laws of Perception

Complex tone patterns can be characterized in terms of whether they conform to the Gestalt laws of perception of visual forms that were described in Chapter 4 (Williams, 1994). In addition to similarity, proximity, good continuation, closure, and articulation, auditory patterns can be characterized with regard to their familiarity, common fate, and stability. Sound elements are similar or proximal with regard to their pitch similarity and closeness of presentation in time. Good continuation occurs when tones of different frequencies are connected by tonal glides or separated by silences. Closure appears to be evident when a tone pattern is started with a particular tone, which creates a bit of auditory fixation. When other tones are introduced, closure occurs when the sequence returns to the fixation tone. Articulation was mentioned earlier.

Common fate occurs when two or more tones have wave envelopes that produce increasing frequency, or decreasing frequency. Stability occurs when a pattern of two or three tones repeats in rapid succession with a pause between sequences. Introduction of unique tones within or between primary sequences would disturb the listener's sense of stability.

Streaming

Although the Gestalt laws of visual perception have not been applied to auditory perception in a systemic fashion (yet), they appear to be operating during auditory streaming. Streaming is the process of detecting a meaningful auditory pattern from a complex array of sounds. One of the earliest recorded examples in psychology was the cocktail party effect (Cherry, 1953). One walks into a room that is buzzing with voices and joins a conversation. Although the majority of one's attention is on the present conversation, a portion of one's attention is devoted to scanning the auditory scene for key words and familiar voices that should be checked out next. The effect was initially reported as evidence that attention is *not* an all-or-none phenomenon; it can be divided among stimulus sources. The current debate is how much of the streaming process actually requires directed attention and how much carries on automatically (Snyder & Alain, 2007).

Streaming is a process of detecting similarities in frequency and loudness, proximity of frequencies (high vs. low pitch), temporal continuity, and spatial source. At some point, the totality of similarities build up to produce a *fission* in which an auditory stream is distinguished from all background sound and recognized as a pattern. Fission is essentially a bifurcation in NDS theory. The unresolved theoretical issues current surround which parts of the streaming process, such as frequency patterns or timbre, are processed in which subareas of the auditory cortex. From a human factors standpoint, the current knowledge of the streaming process has not been systematically translated into principles of good display design, but streaming phenomena would appear to form the crux of pattern confusability and the effectiveness of 3-D auditory displays, as describe forthwith.

Classic Problems and Solutions

Localization

Auditory displays have some classic problems. First, they are often difficult to localize. For reasons explained earlier in conjuction with binaural hearing.

Confusability

Auditory displays can be confused with other warnings. As a general rule, confusion sets in after five auditory displays in a system (Kantowitz & Sorkin, 1983). Hold that thought and consider that military fighter aircraft that were in use during the 1980s utilized 7 to 10 auditory displays (Doll et al., 1984), and up to 17 have been reported for other aircraft including some speech displays (Kantowitz & Sorkin). The problem compounds with multiple speech displays (Teas, 1990). Pattern confusion can be traced in part to the sheer number of tone patterns that a person must remember, similarities among the patterns, and similarities in the locations of the sounds. The Gestalt laws (Williams, 1994) may predict which tone patterns will be confused with others, although there is little empirical support yet for this suggestion.

Hospitals can be a cacophony of sound emanating from several pieces of equipment that monitor different patients. Nurses, who are likely engaged in a task requiring substantial mental concentration need to identify critical alarms, find the source patient, and respond. In an attempt to minimize the confusability of alarms, the International Electrotechnical Commission (2005) issued standard IEC 60602-1-8, which consisted of specified tone patterns for situations involving oxygen, ventilation, cardiovascular events, temperature, infusion, perfusion, and power failures.

The standards were issued without any prior human performance testing. Thus, in an experiment to determine the efficacy of the tonal patterns (Lacherez, Seah, & Sanderson, 2007) nurses were required to identify an IEC alarm that sounded while they were working on an arithmetic task. Nurses with prior musical training reached 80% accuracy on alarm identification, whereas the accuracy of untrained nurses was 50%. Nurses overall did not perform much better than nonnurses despite being aware of the medical implications of the alarms. Their accuracy dropped when the alarms were overlapping or partially overlapping. Although standard alarm formats could in principle have advantages over unstandardized formats, the IEC series did not induce the level of streaming necessary to make the alarms sufficiently effective.

Subsequent research that was also motivated by the confusability of IEC displays has ventured toward strategies that map physiological conditions, such as oxygen supply and heart rate to sound parameters such as frequency patterns (Janata & Edwards, 2013). Changes in the frequencies would then denote changes in the physiological condition. The usability study showed that there were substantial individual differences among medical personnel with regard to accuracy and interpretation of the meaning of the patterns. Some personnel were able to learn and respond to the displays with 100% accuracy, while others did not come close to the criterion level of 87.5%. Thus, the research on sonification strategies continues.

Desensitization

Desensitization occurs when the listener hears so many whizzers and bells going off on a regular basis that the signals lose their attention-getting properties. Operators often respond by resetting the alarm to a higher critical threshold. The danger here is that a serious condition will be overlooked (Chapanis, 1975). A related phenomenon is the cry-wolf effect: Alarms that produce false-positive indicators of a real situation will eventually be ignored. The eventual problem is that the real danger condition will be ignored as well. The basic solution here is to set the alarm controls at a sufficiently high threshold to maximize response to true conditions and minimize false-positive indicators.

Recommendations

Sanders and McCormick (1993) identified some principles for good auditory display design in addition to keeping them few in number and distinguishable in content. The effectiveness of a display can be enhanced by use of a two-stage signal. The first signal should be a simple alert to the operator that a signal with important content is coming up next. For instance, an alerting tone could precede a speech announcement.

Invariance and compatibility are two more related considerations. *Invariance* means that the auditory signals always have the same meaning through out the system or similar systems. *Compatibility* means that they have some relationship to the type of response that the operator is supposed to take. Again it should be recognized from one situation to another.

FIGURE 5.5
Pulse sequence for standard fire alarm.

An example is the national fire-warning standard depicted in Figure 5.5. The signal of about 400 Hz is presented in two bursts of 0.5 s, with 0.5 s in between, and followed by a 1-s burst and silence of about 1 more second (some variability from system to system is tolerated). The loudness should be about 10 to 15 dB greater than the ambient sound level. Some systems have the loudness cranked up to 100 to 115 dB just to make sure everyone's attention has been captured.

On the other hand, nonemergency auditory displays that are too loud can produce shock in the listener when an emergency response is unwarranted. Although it is true that louder displays are more likely to be heard and processed correctly, there is an upper limit when discomfort sets in (Baldwin & Struckman-Johnson, 2002; Edworthy, 1994). A good rule of thumb, based on available experiments, is to produce the display at 10 dB greater than the ambient sound level of the room or system and to conduct user tests if that amount exceeds 65 dB.

A final recommendation is to avoid extreme frequencies. In general, display tones of 200 to 3000 Hz are good. Tones under 1000 Hz are recommended if the listener is more than 1000 ft from the origin of the sound. Tones under 500 Hz are recommended if the listener is located around a corner or otherwise not in a direct path of the traveling sound (Sanders & McCormick, 1993).

Vigilance

Auditory displays are helpful supplements to a primarily visual interface when a vigilance task is involved. An operator may be watching a radar screen for activities of the little blips, but may not notice a new one coming on to the screen. A simple tone would alert the operator to something new taking place (Colquhoun, 1975) and thus enhance performance.

The use of an auditory signal to enhance vigilance is consistent with the principle that redundancy enhances reliability. There are situations coming to the foreground, however, showing as inhibitory effect between auditory and visual signals (Mkrtchyan, Macbeth, Solovey, Ryan, & Cummings, 2012). The effect has become known as cross-modal attentional blink, and occurs when the two signals are presented within 300 msec of each other (Dux & Marois, 2009; Haroush, Deouell, & Hochstein, 2011; Kelly & Dux, 2011; Marti, Sigman, & Dehaene, 2012; Raymond, Shapiro, & Arnell, 1992; Theeuwes, van der Burg, Olivers, & Bronkhorst, 2007). The source of the blink has been traced to a bottleneck through a common circuit in the frontal lobe that is associated with the executive functions of working memory. Cross-model blink does not happen to all operators in laboratory experiments each time the signals are so close together in time, however. The limiting conditions for when this effect occurs is still under investigation.

A different application is the introduction of phone calls into the routine of a night watchman. On a good night, nothing happens, and there is little to watch. By 3:00 a.m. the watchman is likely to fall asleep. A series of phone calls through the night requesting a nominal report of activities keeps the watchman alert.

3-D Auditory Displays

The development of 3-D auditory displays was based on the principle of binaural hearing, and they are intended for the auditory component of a head-mounted display. The sound sources are processed to be slightly out of phase with each other as they would be in free field directional hearing. The signal processing is such that the soundtracks can be artificially located in space to the left or right, front or back, and horizontally.

Airline pilots are typically presented with three streams of speech displays coming from air traffic control, other aircraft, and communications within the airplane itself. The objective of the 3-D display is to project the sound sources in space so that the pilot can attend to one source or another without undue masking from the other sound sources (Wenzel, 1994; Wenzel, Wightman, & Foster, 1988).

Masking experiments in which operators' performance using 3-D head-mounted displays were compared against free-field directional perception show that localization perception was about as accurate with the 3-D audio device as with free-field perception overall, although there were more front-back reversals with the 3-D synthesized medium (Wightman & Kistler, 1989). Doll et al. (1992) determined, however, that greater degrees of sound separation produced lower thresholds of loudness that were required to detect a signal when a mask was generated. The effect was more pronounced when the signal and noise were less correlated.

Three-dimensional synthesized audio displays can enhance pilot performance in some types of tasks. Bronkhorst, Veltman, and van Breda (1996) prepared a 3-D audio track to accompany a primarily visual task on a flight simulator. The participating pilots were chasing another aircraft that disappeared at critical points in the flight. The participants were required to locate the target aircraft. The researchers found that the combination of visual and 3-D audio signals produced shorter search times than either visual or 3-D audio displays alone. Ratings of workload were not affected by the introduction of 3-D audio.

Wenzel, Wightman, and Foster (1988) also found that 3-D audio displays can make serious alarms more salient without having to make them louder. The alarm preparation for wind sheer involved tonal signals with a synthetic voice overlay saying, "Wind sheer, wind sheer." The signal was phased so that it would oscillate rapidly from the left ear to the right.

3-D audio also showed an advantage for detecting infrequent speech signals from a background stream of irrelevant speech (McAnally & Martin, 2007). The advantage of 3-D seems to be attenuated by a greater total density of speech, making the basic signal detection task more challenging.

Speech Displays

The elementary connection between human speech and machine speech is the phoneme. The *phoneme* is the smallest unit of speech that contributes a change in the meaning of a word. Consider the set of words *pig*, *pit*, *pet*, *get*, *gem*. Each change in letter is a change in a phoneme. There are actually more phonemes in the English language than letters of the alphabet—long and short vowels, letters that sound alike, and diphthongs—but the elementary nature of the phoneme is about as elementary as a change in letter. Further up the linguistic scale we have *morphemes*, which are root words, prefixes, suffixes, compounded words, and similar attachments. Beyond the morpheme are the word, phrase, and sentence.

Speech displays occur in telephone systems, machine interfaces, and annotated computer programs. The criteria for system performance go beyond signal detectability to include intelligibility and acceptability (Voiers, Sharpley, & Panzer, 2002). Identification of the speaker is another important feature of telephone communication.

Intelligibility is the extent to which listeners can determine particular words from the source. Here it is possible to contrive experiments that require listeners to select the word they hear from a multiple-choice list. The real word and the distractors are typically only one phoneme apart. Factors that affect intelligibility include the signal-to-noise ratio, familiarity of the message, amount of vocal effort from the speaker, and electronic contributions such as distortion and peak clipping. Familiarity helps not only to recognize the particular words because they have been heard before, but also to provide some context for interpreting new information.

Acceptability is the extent to which noise or other electronic compromises are annoying. Acceptability is typically measured using a subjective rating. At the most favorable end of the scale, the speech is both intelligible and not annoying. In the middle ranges, it becomes more annoying. At the opposite extreme the speech is both unintelligible and annoying.

Speaker identification can be assessed through an experiment wherein listeners are presented with clear examples of several target individuals' speech. They are then presented with speech samples from the display system that might or might not be examples of the targets' voices. The listeners should be able to associate test samples with the proper voices.

Speech Spectrograms

A *speech spectrogram* (Figure 5.6) is a display of frequencies generated over time as a person pronounces a word or makes a longer utterance. The ordinate is time. The abscissa is the octave band of the frequency. The speaker in this example is saying, "Don't ask me to carry an oily rag like that."

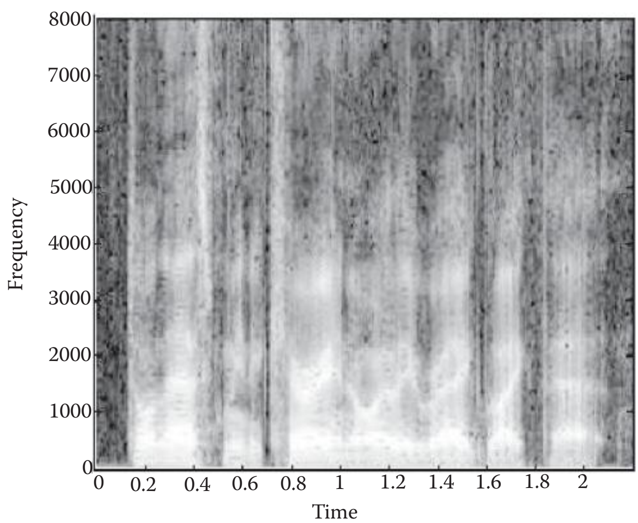

FIGURE 5.6
Speech spectrogram for "Don't ask me to carry an oily rag like that," a phonetically balanced sentence used as normalization, from the TIMIT database. (Michael Johnson, Speech Laboratory, Marquette University School of Engineering, copyright 2004. Reprinted with permission.)

In the early days of speech analysis and processing, scientists conjectured that it might be possible to link specific phonemes to specific frequency arrays. Unfortunately, the speech-to-signal correspondence is far more complex. People often emit different frequencies with the same word, depending on mood and context of the sentence. For instance, we can make a statement sound like a question by simply raising our voices at the end of the sentence without changing a word.

The creation of synthetic speech displays begins with the access of the spoken information. The technology is similar to the digital sampling method used to assimilate synthetic musical instruments. The complexity of frequencies emitted by voice, however, had made digital sampling a challenge. The common solution, however, is to record spoken keywords, phrases, and statements into little files. The computer program will then execute the right combination of files to produce the message. Telephone directory assistance in the United States is an example. The recorded files include, "The number is," words for each of the 10 digits, "hundred," "thousand," "repeat," and the closing message about automatic dialing at no extra charge. The recorded files preserve the original speaker's vocal inflections and regional accents. The listener can detect a little choppiness between the elementary word files, but not enough to present a problem. The difficulty, however, is to extend that technique to situations where a more extensive vocabulary is needed to compose many possible individualized messages.

The literature on speech communication and processing is extensive, although much of it pertains to the engineering aspects of the system rather than the human factors involved. Of course, human factors limitations have undoubtedly spurred the electrical engineers to greater heights. Speech-related controls as considered in Chapter 7.

Noise

Noise, for the purposes of this exposition, is any unwanted sound. Chapter 3 already explained the role of noise with respect to the detection of signals. This section elucidates a few more important aspects of auditory noise.

Colors of Noise

Noise specialists characterize noise as having four basic colors: white, pink, brown, and black (Ward, 2002). White noise is an equally weighted composite of sound from the full range of frequency octaves. If we were to plot a frequency distribution of the number of times a frequency was emitted from a white noise source by the frequency of the particular tone, the distribution would be square. If we were to take repeated random samples of white noise from a square distribution and add the samples together, a Gaussian or normal distribution would be obtained.

The colored noises occur when the distribution of frequencies from a noise source are not as random as "random" should be. Instead there is some internal deterministic structure to the noise source. Distributions of colored noise can be expressed as:

$$F(\mathrm{H}) = kH^{-a}, \tag{5.2}$$

where H is the frequency of the sound in cycles per second, $F(H)$ is the frequency, that is, number of times the tonal frequency occurs in the sample, a is a constant of the distribution that determines the color of the noise, and *k* is a constant of no particular consequence. (Don't you just hate it when there are two different uses of the word *frequency* in the same sentence?)

The noise is designated as pink if $|a| = 1$. The noise is brown if $|a| = 2$. The noise is black if $|a| > 2$. As the noise changes from pink to black, there is greater determinism in the source. At the black extreme, we have an unwavering, sustained, and probably annoying single tone. Colored noise with $|a| < 2$ thus indicates that a self-organized process is taking place; *a* is, incidentally, a fractal dimension for the distribution. In any case, white noise is the noise of choice for the preparation of experimental stimuli, but colored noise is closer to what occurs in reality.

More Signal Detection

Although the loudness of a noise can obviate the perception of a stimulus, the frequency of the noise tones can also interfere with the perception of the true stimulus. As one would expect, the interference is greater when the noise and signal are close in frequency. For reasons not yet known, noise tones that are slightly lower in frequency than the signal induce more interference than noise tones that are slightly higher in frequency than the signal (Kantowitz & Sorkin, 1983).

Just as it seems safe to conclude that noise is always bad for signal detection, we must introduce the concept of *stochastic resonance*. This is a signal detection phenomenon whereby a very weak signal can actually be detected more often if a modicum of noise is added to it. Ward (2002) characterized stochastic resonance as a double-well phenomenon, which is essentially a cusp catastrophe function as depicted in Figure 5.7. The curve in Figure 5.7 represents an underlying potential energy function. As we travel across the ordinate we move from one low-entropy well to another. The well on the left corresponds to the perceptual conclusion that no signal exists. The well on the right corresponds to the conclusion that the signal plus noise is present. The addition of the noise pulls the perceiver out of the fixed conclusion that the signal does not exist and introduces enough entropy to suggest that perhaps it could exist. The addition of noise to the weak signal, however,

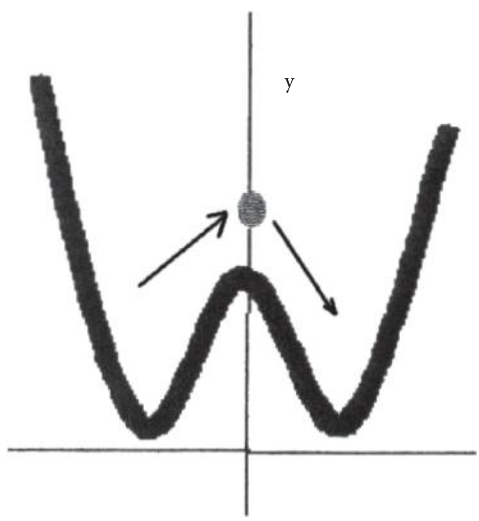

FIGURE 5.7
Two-well phase shift for stochastic resonance.

does not guarantee that the perceivers will always detect the signal. It only enhances the chances of them doing so.

Hearing Loss and Noise Exposure

Extremely loud blast noises can damage the eardrums, but the majority of occupationally related hearing loss is gradual and chronic. A good deal of what has been learned about hearing impairment comes from studies in which a temporary threshold shift has been induced. If a person is exposed to a loud noise source, their threshold to detect a sound becomes elevated after the noise has been removed. The threshold drops backs to normal soon. With prolonged and repeated occupational exposures of 10 years or more, however, the threshold eventually does not return to normal but stays elevated. The perception of frequencies greater than 4000 Hz are impaired first. With additional exposures the impairment spreads to the lower frequencies as well (Kantowitz & Sorkin, 1983).

Environmental noise levels can be measured with sound meters. There are two settings that pertain to OSHA standards (1996). The FAST or SLOW setting allows the device to change its reading as the operator moves through an environment; the OSHA standards are set to SLOW. The contour control places different weights on the various frequency bands when the bands are sampled and the overall sound level is calculated. The standards require the A contour, which places greater weight on frequencies greater than 1000 Hz compared with the other contours. OSHA's exposure limit is 90 dB for the composite total noise for an 8-hr daily exposure. The allowable total exposure decays rapidly thereafter, so that only 15 min per day are allowed at 115 dB of ambient noise.

Figure 5.8 depicts a noise criterion curve for the OSHA standards. The various curves indicate a distribution of frequency intensities associated with a particular decibel level

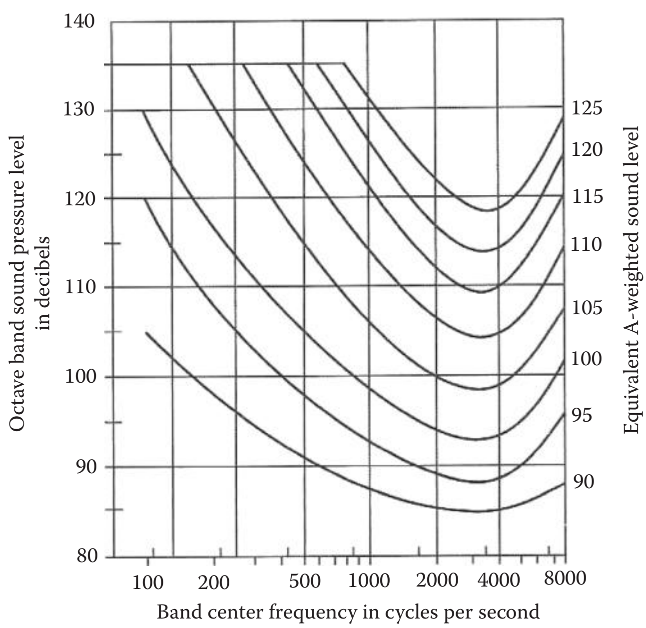

FIGURE 5.8
Noise criterion curves for OSHA standard 1910.75.

that is going to be used for allowable exposure. The goal of system engineering is to control noise within each frequency band so that all noises stay below the curve. The basic concept of a noise criterion curve can also be applied to any work environment that is subcritical for total noise as defined in the OSHA (1996) standard.

Noise can be controlled with structural engineering or personal protection equipment. In the case of the environment, the architects can build sound-resistant rooms for machine operators if their work patterns allow them some chances to escape from the noise of the machines (Kantowitz & Sorkin, 1983). Such rooms are built with sound-absorbing walls, double-pane glass, and seals around the doors. Employers who are building a factory from scratch might consider designing the factory floor as a series of platforms that are supported separately from below. Heavy machinery that creates vibration should be placed on separate pieces of flooring, and the pieces of flooring can then be connected with vibration-absorbing seals. In this fashion the vibration from one machine will not transfer to all parts of the factory environment.

Roof fans produce another source of noise. In short, fans with more blades are better than fans with fewer blades. Fans with a greater number of blades still produce noise, but they do so at a higher frequency (pitch) that does not travel as far as the low-frequency noises.

If structural redesign is not possible, operators need to use ear protection. Earplugs are cheap and disposable. Their advantage is that they are inexpensive and can be handed out to visitors readily. Their disadvantage is that the users often do not install them properly, thus they do no good and perhaps give people a false sense of protection. A more certain alternative are earmuffs that are designed like a set of headphones, except that they do not convey sound; they insulate against it. They are sturdy and effective. They cost more to purchase, but they are reasonably permanent devices.

Tactile Displays

Human skin sensors respond to stimulation from mechanical, thermal, chemical, or electrical sources. Free nerve endings are stimulated, usually mechanically, and connect to the Pancinian corpuscles located deep in the skin. Experimental evidence does not support a one-to-one correspondence between the nerve endings and particular sensations. Sensations on separate areas of the body, however, can be used as information conveyed by vibrotactile displays.

Gloves

The sense of touch is often compromised by the use of gloves. Gloves can be compared on the amount of sheer force that they exert on the operator. Sheer forces block incoming information and the more obvious aspects of the glove often inhibit manual and finger dexterity.

Haptic Perception

Touch is primarily an active process in human factors applications. Roughness is the basic perceptual element. This principle extends to *haptic perception*, which is the perception of objects' shapes. Although it is well known that sensations received through the hands and all other parts of the body are transmitted to the sensory cortex in the brain, the means by which a person can assemble the sensations into a perception of an object's form is mostly uncharted territory. The subject has been gaining some momentum, however, in light of design challenges associated with virtual reality programs; in those application`s the sense of touch must be introduced synthetically to give the individual a realistic experience with the computer program. Virtual reality is discussed further in Chapter 11.

The sense of touch receives more attention in the context of controls. It is usually not the primary source of information that a person receives, but a form of feedback from the control action. On the other hand, visually or aurally degraded or overloaded conditions could make haptic controls more attractive.

Knobs

Tactile displays are far less common than visual or auditory displays, although there are a few classic applications (from Kantowitz & Sorkin, 1983). Airplanes of the World War II era were equipped with several joysticks that were equipped with handles of similar if not identical shapes. Too often the pilot who was trying to keep an eye on the sky reached for a stick and operated the wrong one with disastrous results. The solution to the problem, credited to Alphonse Chapanis, was to mount handles of different shapes on each of the joysticks so that the operator could feel the correct stick (Roscoe, 2006). It worked.

Navy underwater divers often have to rely on the feel of knobs to know which control they are operating. Although the shape of knobs is not a display in the usual sense, the tactile sense is used to substitute for the labels that cannot be seen in that environment. A set of knob shapes or joystick handles that the U.S. Department of Defense (2012) considers maximally distinguishable is shown in Figure 5.9.

Vibration

Another aircraft application involved the use of vibrator seats and shaker sticks. The old expression, "fly by the seat of your pants," originated at a point in aviation history where the airplanes were not equipped with altimeters. Inasmuch as they were flimsy vehicles compared with contemporary designs, the pilot could literally feel the elevation

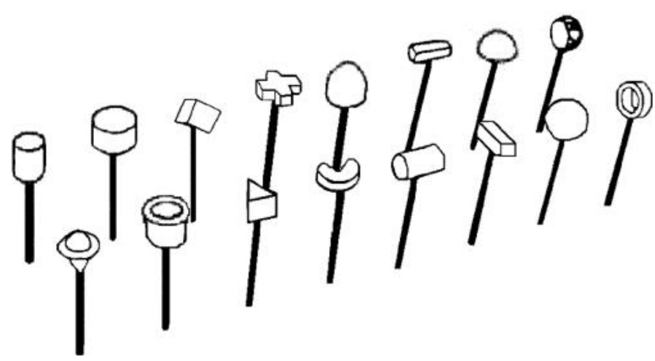

FIGURE 5.9
Maximally distinguishable knobs. (Reprinted from U.S. Department of Defense, *Department of Defense design criteria standard: Human engineering*, Author, Washington, DC, retrieved April 4, 2012 from http://www.assistdocs.com, 2012, in the public domain.)

and descent in his seat. As aircraft became more plush and insulated the human from the environment to a greater extent, it became necessary to put vibrators in the pilots' seats so they could feel the elevation. Of course, the same epoch of aircraft designs also included altimeters, but redundancy improved reliability, especially for pilots who had learned from the old school.

Cell phones that can be switched to vibrate in the operator's pocket instead of ringing audibly are now fairly common. They can give the operator a modicum of privacy when the message arrives.

Diving suits and other environmental suits insulate the operator from direct control of the environment, and there is often an advantage to building in synthetic sensory information so that the operator can react with the external environment in a manner more similar to direct contact. Similarly, virtual reality gloves (Chapter 11) replace the tactile sensations that are missing in a virtual environment.

Whereas knobs and vibrating phones convey simple pieces of information that does not change over time, vibrotactile tactile displays can convey more complex information that does change. Systems of tactile stimulators can be worn as objects strapped to the operator's thigh (Salzer, Oron-Gilad, Ronen, & Parmet, 2011), or in the form of belts and vests (Jones & Sarter, 2008). Information can be mapped to any of four aspects of a vibration—frequency, duration, amplitude or intensity, and location of the body. Sets of seven or eight signals that are readily distinguished and learned are not uncommon.

Vibrotactile displays can be viable additions to a PMS if the auditory and visual channels are overloaded. They can also add redundancy to visual and auditory signals. Some of the more common applications convey spatial information to the operator when the driver is moving and potentially disoriented, signaling the direction of an incoming event or object, and signaling the need to shift attention. A firm set of rules for making optimal connections between elements of information and aspects of vibration is not yet available, but one objective of on-going research to find those combinations (Jones & Sarter, 2008).

DISCUSSION QUESTIONS

1. A classic problem in human factors is to determine what information needs to be displayed. Of course, this question arises earlier in the design process than the choice of display type. How would you go about determining what information needs to be displayed, either aurally or visually?
2. Redundancy improves reliability. How would two modes of display help, and in what situations would an auditory display assist a visual display?
3. How many different types of noises does your desktop or laptop computer make? What is the system signaling? Are any of these sounds actually necessary? If so, which ones?
4. About 90% of the auditory alarms that go off in a hospital patient-care facility are false alarms. What is the impact of the false alarm rate? How might the false alarm rate be curbed?
5. A recent innovation in auditory displays is to present them in the form of a few bars of a popular song. Comment on the assets and limitations of this technique.
6. How many auditory signals come from your automobile? How many of them are necessary? Should others be added?

6

Cognition

This chapter considers the central principles of human thought processes that occur between the perception of stimuli and the execution of a control motion. The topics for this chapter are the organization of human memory, types of decisions, cognitive channel capacity and workload, automatization of cognitive processes, degrees of freedom in cognitive task organization, dynamic decisions and situational awareness, and the cognitive analysis of cognitive workload.

Organization of Human Memory

Figure 6.1 shows a schematic for the human memory process. At the first stage, an incoming stimulus is retained on the sensory register for a fraction of a second. The sensory register would be the retinal of the eye or the basilar membrane of the ear. This is just enough time for the remainder of the process to move into action if the human is paying attention to the stimulus.

A neural representation of the stimulus, or *memory trace*, starts to form when the initial stimulus is received. Once the memory trace is formed, the information item enters a two-stage process by which it could be committed to permanent memory or lost. It should be mentioned here that no direct correspondence has been made yet between the configuration of neural impulses and the thoughts that are produced by them, as we experience the thought. This unanswered question is known as the *binding problem*. In recent years, there has been some attempt to solve the binding problem by drawing on principles of quantum physics that are then combined with some clumsy notions of "consciousness." This approach has not yet paid off in any meaningful way (Koehler, 2011).

Short-Term Memory

The first stage after forming the memory trace is the short-term memory. The short-term memory can hold information for a range of 30 s to 2 min. During that time, the information is being processed for meaning. Although we do remember the original sights and sounds of events, the majority of our memories consist of the sense or meaning of what happened more so than a recording of the exact event in real time.

The further transition of an information item to long-term memory depends on rehearsal. Here the operator is repeating the information, probably silently, enough times to either retain the information long enough to use it, or for more in-depth processing for long-term memory. Anyone who has ever repeated a phone number to themselves often enough to actually get to the phone without writing the number down has experienced the first type of rehearsal. Anyone who has ever studied for an examination and gone over the material several times has experienced the second type of rehearsal.

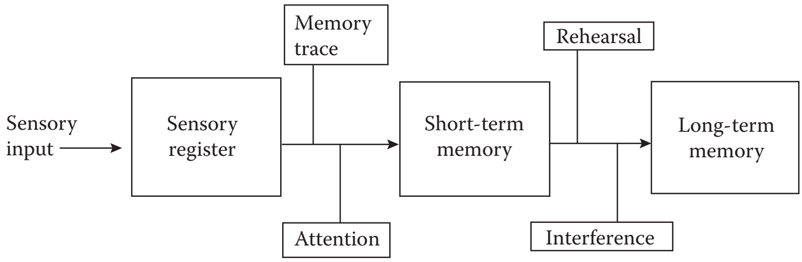

FIGURE 6.1
Memory schematic.

The transition from short-term to long-term memory can be interrupted, however, by interfering stimuli. There are two sources of interference: more incoming stimuli through the sensory register and short-term memory, and old information coming from the long-term memory. When a flow of incoming information occurs, the information items that occur at the beginning of the sequence are remembered best; this is known as a *primacy effect*. The information items occurring at the end of the series are remembered second best, so long as the opportunity for recall is relatively immediate; this is known as the *recency effect*. The information items in the middle of the sequence are at the greatest risk for being lost.

The primacy and recency effects were first discovered in experiments where human participants were asked to memorize lists of meaningless items. There were opportunities for timed rehearsals in those experiments. In real-world applications, there is little time for planned rehearsals, and the information is flowing in over time. The operators often lock onto the first items coming in for proper processing. At this point attention is diverted from newer incoming stimuli, which are likely to be lost if the machine system does not contain a memory storage facility of some sort. This special case of the primacy effect is known as an *anchoring effect*.

The short-term memory has a limited storage capacity in the amount of 7 plus or minus 2 chunks (Miller, 1956). A psychological chunk is not easily translated into information units such as bits and bytes. Rather, a chunk is better conceptualized as an intact information unit of some sort. It is also possible to cram more than 7 (or 9) units into short-term memory; people do remember longer lists of items. A process of chunking (verb) as depicted in Figure 6.2 accomplishes the short-term memory enhancement.

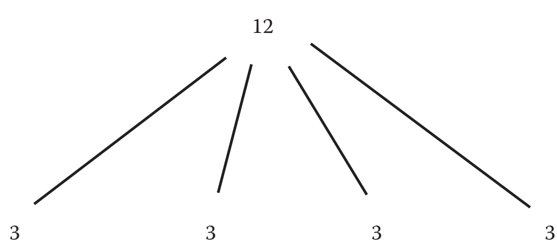

FIGURE 6.2
Chunking in short-term memory.

Long-Term Memory

The long-term memory has an indefinitely large storage capacity, and it can hold information for an indefinitely long period. Information that resides there has been processed for meaning during the rehearsal process of short-term memory. The meaning of "meaning" is subjective, as not every idea has equal meaning to everyone. Nonetheless, meaning can be interpreted as the number of connections an idea has to other ideas that are already in long-term memory (Kohonen, 1989).

Information can be retrieved from long-term memory on demand by means of a *retrieval cue*. A retrieval cue is a stimulus–response association between one idea element and other ideas with which it may be connected. Figure 6.3 describes a plausible configuration of idea elements in the long-term memory. Imagine that the dark circle in the shaded area denotes the target memory. A proper retrieval cue, perhaps a clear question, will activate the idea. Once the target idea is activated, adjacent ideas become somewhat activated as well, as depicted by the other circles in the shaded area. The ideas with secondary activation are linked to more ideas in the unshaded area, but the latter ideas do not become activated until the secondary ideas are more substantially activated themselves.

An important side effect of the retrieval cue process is that change in retrieval cues can produce a memory failure. It commonly happens in technical language, for instance, that a concept may be known by more than one name or word. If the speaker uses a synonym with which the listener is unfamiliar, the listener will not remember the idea when it is needed.

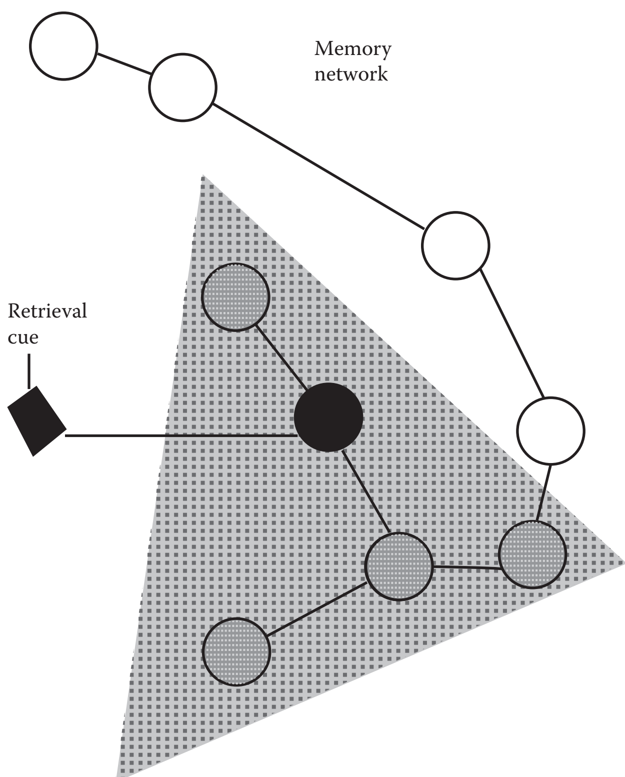

FIGURE 6.3
Retrieval in long-term memory.

Retrieval from long-term memory can be inhibited by interference. With anterograde interference, old learning may be so entrenched that it inhibits some element of new learning. With retrograde interference, new learning may inhibit the retrieval of the old learning. Again, relative meaning of the material that is being learned could account some anterograde interference effects. If one person already has learned one set of meanings, incoming new information that conflicts with the original in some way could produce interference, or might have very reduced incremental meaning, whereas someone who has not had the prior learning experience would not have the same experiences with the new material.

Types of Memory

There was some debate in earlier decades as to whether memory functions were localized to particular areas of the brain or whether memory was distributed throughout the brain. Both interpretations are now considered partially true. The hippocampus in the limbic system is active when any memory process is at work. Specific types of memory, however, such as verbal or pictorial memory, engage different specific regions of the cortex in addition to the hippocampus; mapping out the brain circuitry that supports this variety of memory functions is an ongoing challenge (Guastello, Nielson, & Ross, 2002).

There are several specific types of memory. There is *semantic memory*, where meanings of words are encoded and processed. The semantic memory makes use of categories, and the formation of categories. In a classic experiment (Bousfield, 1953), experimental participants were given a list of animals, plants, professions and names of people in random order. After they were given a period of time to memorize the list, they were asked to repeat the list items in any order they preferred. The head dump revealed that animals, plants, names, and professions were retrieved in clusters that corresponded to semantic categories and not in random order. Categorical thinking will be revisited in Chapter 12.

Propositional memory pertains to relationships among concepts. *Spatial memory* pertains to pictures and relationships. Propositional and spatial memory appear to be closely connected; perhaps for this reason it is not uncommon for people to draw diagrams of ideas.

Finally, there is *episodic memory*, which is the memory for events, and *procedural memory*, which is memory for sequences of one's actions. Procedural memory will be considered further in this chapter under the section on cognitive analysis of a person–machine system, where one of the objectives is to extract the content of an operator's procedural memory in order to understand how a task is performed and what resources are used to do so. Episodic and procedural memory are often aggregated as *implicit memory* to refer to the relative involuntary nature of memories we accumulate from experiences, as opposed to *explicit memory*, which is engaged when we deliberately try to remember something. Semantic, propositional, and spatial memory would be explicit forms of memory.

Working Memory

Working memory is "assumed to be a hierarchical system involving short-term memory, representational components, plus a general executive attentional component" (Kane & Engle, 2002, p. 638). It is analogous to random access model (RAM) in a computer, which is the amount of workspace available to manipulate programs and data input. It is considerably less than the total memory storage space, which is generally a static entity that does comparably little additional processing by itself; yet anything in it can be called into working memory (if it hasn't been forgotten) when given a viable retrieval cue.

The representational components are workspace modules that are associated with specific types of memory and cognitive functioning (Logie, 2011). The executive function recruits workspaces and the functions needed to perform a task. The executive function also maintains the focus of attention. When we tell someone, "Watch what you're doing!" we are recommending a re-engagement of the executive function, which might have lapsed due to fatigue or distraction by nontask events. Whereas human factors engineers examine properties of tasks that result in performance stability or decline under different types of load conditions, cognitive psychologists have sought to unravel the structure of working memory itself (Baddeley, 2003). Channel capacity, according to Logie (2011) is likely to be an emergent result of the working memory components that are required by a particular task. For more demanding tasks, the domain-general function of attentional control supplements the domain-specific resources. Furthermore, memory and processing are two different functions with different capacities, although they normally operate together to varying degrees.

Working memory capacity is usually measured by span tasks; in a common example, a test administrator would read sequences of numbers of varying length, and the respondent would have to remember the numbers that were spoken and repeat them back in order, or in reverse order. Span tasks can also involve verbal, pictorial, or episodic material (Conway, Kane, Bunting, Hambrick, & Engle, 2005; Guastello, Boeh, Shumaker, & Schimmels, 2012).

The relationship between working memory capacity and performance is somewhat enigmatic. Kane and Engle (2002) reported that "high spans engage attentional processing to achieve results while low spans rely on automatic processing. High spans are expected to be more taxed by a secondary task" (p. 641). McVay and Kane (2009) reported that working memory capacity was correlated with performance on a 45-min version of a task in that it kept the participants' minds from wandering, but capacity had little or no impact on performance on a 2-hr version of the task. This finding led Redick, Calvo, Gay, and Engle (2011) to conclude that time on task "does not appear critical to the working memory capacity relationship with task performance" (p. 322). In other words, working memory capacity was interpreted as relevant to the management of cognitive workload, but not to fatigue, which is more susceptible to time on task (Chapter 9).

Task Switching

Numerous studies, summarized in Andreadis and Quinlan (2010) and Rubinstein, Meyer, and Evans (2001), have shown that switching tasks incurs a *switching cost* in the form of lower response time. The time loss was traced to the activation of rule groups when anticipating the next task. Rule complexity adds to the switching cost, such that a switch from Task A to Task B might not be the same as vice versa. Some of the cost is attentional, and some of it is computational (Lorist et al., 2000).

Andreadis and Quinlan (2010) noted that experimental paradigms for task switching typically relied on predictable sequences of tasks, for example, blocks of trials that are all repeated tasks or alternating tasks. They found that predictable patterns incurred lower switching costs as did cuing the impending task and lengthening the lag time between the cue and the task. Ambiguous task cues increased switching costs, because both rule sets remained activated when the cue was presented. Such "crosstalk" between tasks was equivalent for both predictable and unpredictable sequences. When cues were ambiguous, the switching costs were actually less during the unpredictable conditions than under predictable conditions.

In the foregoing studies, task-switching regimes were planned by the experimenter. They might mimic conditions where task selection is controlled by automation reasonably well. Voluntary task switching could be a different matter entirely. Often people have several tasks to choose from and many of each to do in their work queues. Although involuntary task switching can induce fatigue (Hancock, 2007; Lorist et al., 2000), there is evidence that task switching can be a means for alleviating fatigue or boredom (Alves & Kelsey, 2010; Lorist & Faber, 2011).

Some logical questions at this stage are the following: (a) "What are the rules and patterns that people naturally engage for voluntary task switching?" (b) "How do task switching schemata shift in response to externally induced constraints on the task goals, such as task performance quotas?" In a recent study, 54 undergraduates performed 7 different computer-based cognitive tasks 7 producing sets of 49 responses. They worked under two instructional conditions; one required a quota of 7 examples of each task, and the other did not impose a task quota beyond doing 49 examples of the tasks in any combination (Guastello, Gorin et al., 2012). The sequences of task choices were analyzed using symbolic dynamics to extract types of patterns and pattern lengths. Patterns were then classified and compared with regard to Shannon entropy, topological entropy (related in principle to the type of turbulence measured by a Lyapunov exponent), number of task switches involved, and overall performance. The results indicated that similar but different patterns were generated under the two instructional conditions. There were four patterns in the task quota condition: "task-first," which was to do all examples of one task before moving on to the next task; "set-first," which was to do set of seven different tasks then repeat the set until finished; random selection of tasks, and a mixed strategy that showed some indications of task-first and set-first strategies. In the nonquota condition, the task-first strategy was replaced by the "favorite task" strategy, where the participant stuck for a while with a task that seemed likeable then switched to something else; sometimes there were two favorite tasks represented in the series. Set-first, random, and mixed strategies were also used in the nonquota condition, although the set-first strategy was adopted less often than in the quota condition.

Better performance was associated with task sequence choices that exhibited lower topological entropy, but only about half the participants adopted a low-entropy strategy. Both entropy metrics were correlated with the amount of voluntary task switching. There appeared to be a trade-off between switching costs and minimizing entropy. The set-first strategy actually produced the lowest amount of entropy because it involved the repetition of a long unchanging sequence, but it did involve the greatest number of switches. Task-first and favorite task produced the lowest number of switches, but not the lowest entropy in task patterns. The relevance of entropy metrics is explored further later in this chapter.

Fluid Intelligence

Individual differences in working memory capacity have been traced to individual differences in fluid intelligence. Some condensed history of intelligence would be helpful here. When the concept of the intelligence quotient (IQ) was first introduced by Binet in the 1890s, it was defined as a single construct. By 1930 there was a debate as to whether intelligence consisted of one general factor, *g*, or multiple independent factors. The arguments in favor of one conclusion or another were based on differences in the methods of factor analysis that researchers, primarily Spearman and Thurstone, were developing and using, the notion that different brain areas had different dedicated functions, and the validity of intelligence measurements for predicting behaviors of different types.

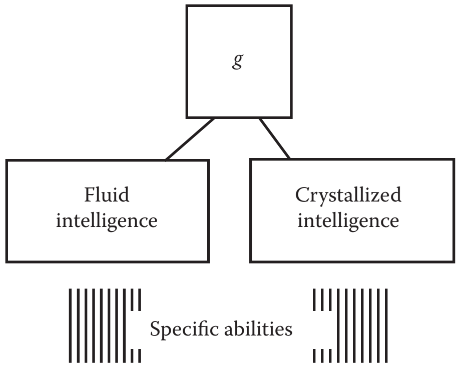

FIGURE 6.4
Hierarchical organization of mental abilities.

In the 1950s, Guilford (1967) introduced a different approach to defining intelligence, and a factor analytic strategy to go with it, that was based on different types of inputs, different types of operations on those inputs, and different types of outputs. The net result was a taxonomy of 120 factors of intelligence, each of which consisted of a unique combination of input, process, and output. One might parenthetically see a parallel here with the notions of input, process, and output in basic computer science. In any case, one important product of Guilford's work was the discovery of the difference between divergent thinking and convergent thinking. The former being more germane to creative thinking, whereas the latter was more germane to optimizing problems; this distinction is expanded further later in this chapter.

The next major development was introduced by Cattell (1971) who resolved the debate between *g* and multiple factors, and the omission of divergent thinking prior to Guilford in the form of a hierarchical theory of intelligence (Figure 6.4). At the top of the hierarchy was *g*, for general intelligence, which separated into two broad components, *crystallized* and *fluid* intelligence. Crystallized intelligence describes the mental abilities and knowledge bases that have been accumulated over time. Fluid intelligence involves the adaptation of mental schemata and knowledge to new situations; it includes the divergent thinking processes are involved in creative thinking (Hakstian & Cattell, 1978; Nusbaum & Silvia, 2011). The other more specific abilities that had already been identified in earlier intelligence research were then thought to form a third tier in the hierarchy (Hakstian & Cattell, 1978). That said, contemporary research now shows a robust relationship between working memory capacity and measures of fluid intelligence (Kane, Hambrick, & Conway, 2005).

Types of Decisions

There are several fundamental types of decisions that are relevant to human factors applications. Here they are grouped as simple binary decisions, optimizing decisions, and nonoptimizing decisions. Cognitive biases fall somewhere between optimizing and nonoptimizing decisions. Afterward we consider the production paradox and troubleshooting strategies where combinations of optimizing and nonoptimizing decisions are involved.

Simple Binary Decisions

Perhaps the most basic decision is the simple detection of a stimulus, such as a light or a sound. For the most part the dynamics of signal detection were covered in Chapter 3. It was also noted in Chapter 3, however, that some very complex decisions culminate in a binary decision. Is event *X* a member of category A, or isn't it? In both the simple and complex versions of signal detection, the accuracy of a set of decisions can be represented by the fourfold table for correct hits, correct rejections, misses, and false alarms. Inspection tasks are common examples of decisions that intensively involve signal detection.

Optimizing Decisions

Given a set of possible choices, which one will produce the best results? Any student who has taken a multiple-choice test has experience with this class of decisions. In the simpler examples of optimizing decisions, the thinker has all the information and reasoning processes available to determine which outcome will be optimal. In many real-world decisions, however, the thinker does not have all the information, but may be able to get the information if it were available along with the time to find it. In less fortunate situations, the thinker does not know what information is unknown but necessary. In flimsier situations, the thinker does not know what all the options are, let alone how to evaluate them.

Expectancy Theory

The foregoing compromises to optimality lead to a broad class of optimizing situations involving decisions under risk. Given the situation where there are defined options and rational behavior, optimal outcomes ($Max[O]$), and odds of an outcome coming to fruition ($Pr[O]$), a strictly rational decision takes the form

$$E[O] = Max[O] \times Pr[O], \tag{6.1}$$

where $E[O]$ is the expected outcome.

The concept represented in Equation 6.1 dates back to 17th-century economics (Vroom, 1964). It surfaced in psychology with cognitive learning theory (Tolman, 1932). The principle was that the rat *knows where* the cheese is. The notion that a rat could know anything ran counter to the basic tenets of the strict behaviorism of the time, which assumed nothing whatsoever about mental processes or structures. Strict behaviorism only considered the movement of people or animals through time and space, which, in principle, was observable to anyone who cared to observe the event or repeat the experiment.

The notion that a rat knew where anything was also ran counter to strict behaviorism because it assumed cognitive structures that Tolman named cognitive maps. *Cognitive maps* are mental representations of a physical environment that are embodied in the individual and which the individual uses to solve problems concerning the location of objects or one's personal location in space and how to move to another location. In the most pointed experiments, Tolman would first allow his rats to explore a radial maze without any reward. A radial maze has several walkways organized in a circle that all converge on a central area in the middle. He would then place cheese of different amounts in different fingers of the radial maze, and the fingers of the maze would be used at varying probabilities. The rat would be allowed to enter the maze from any of the openings, would run toward the middle, then take whatever left or right turns were necessary to go to the maze finger that had the greatest likelihood of containing the most cheese, as expressed by Equation 6.1. If there was no cheese in that target space, the rat would return and go to the second most likely location, and so on. If the rat entered the maze from different points, and the location of maximum cheese was the same in each case, the rat would need to make a different set of left or right turns to get the cheese. Because the rat was able to do so, the conclusion was that the rat had a cognitive map of the maze. Cognitive maps arise again in Chapter 14 on the topic of navigation through spaces.

Equation 6.1 would be (mentally) constructed for each option. The option that produces the best $E[O]$ wins. In situations with less than perfect information, however, one does not know the $Pr[O]$ and must guess it somehow. *Overconfidence bias* occurs when the individual overestimates the odds of success.

Prospect Theory

Equation 6.1 defines a strictly rational approach to optimization. Prospect theory (Kahneman & Tversky, 1979), however, found three common compromises to strict optimization in human decision making. One form of suboptimality occurs when the decision maker is willing to *pay for certainty*: The decision maker is inclined to accept lower $E[O]$ if there is a higher $Pr[O]$ associated with the option.

On the other hand, there are also conditions where risky behavior can be expected. If, in a series of investment decisions, the decision maker is accumulating a lot of losses, the tendency is to accept more risk in hopes that the large payoff will actually occur.

The second example of suboptimality is that losses weigh heavier than gains. Suppose Equation 6.1 were extended to include a cost for participation in an option:

$$E[O] = Max[O] \times Pr[O] - C[O] \tag{6.2}$$

and $C[O]$ was different for each option. Whereas strict rationality would simply make the subtraction in each case and look for the best $E[O]$, the suboptimal thinker would place greater weight on options that involved lower participation costs at the expense of expected outcomes.

Another version of the same bias is related to expected losses. Whereas the strictly rational decision maker would simply calculate Equation 6.1 for all the options, and make a subtraction for expected losses, and select the option with the highest net $E[O]$, the conventional thinker would regard \$500 lost as larger than \$500 gained. Thus, decisions would be biased toward loss aversion.

The third type of suboptimality is seen in insurance-buying behavior. Insurance is a protection against risks. Insurance costs are structured to reflect higher premiums in cases where the odds of having to cover a loss are high, and lower premiums in cases where the odds of a payoff are lower for the insurer. Hence, there is a deductible amount on many types of policies; the deductible amount represents an almost guaranteed loss. Smaller losses are much more likely than larger ones; the distribution of loss sizes follows an exponential distribution, The suboptimal insurance buyer, however, is inclined to protect against as much loss as possible, and thus chooses policies that cost more in the long run than the expected losses from out of pocket. The advisement would be, therefore, to buy a policy with as large a deductible as the buyer can afford, and protect against the losses of more devastating amounts.

One might rightly ask how investment decisions became entwined with human factors. One reason is that there are other types of decisions that follow the same basic framework of expected odds and maximum possible outcomes. Another is that the sheer volume of information that person needs to process to make some of these decisions has inspired the development of decision tools (software) that is meant to assist the decision maker.

Incomplete Information

When the decision makers do not know either or both *Max*[*O*] or *Pr*[*O*], they have to make something up for the missing amounts or reframe the problem in some way. In some cases, the missing information takes the form of a statistical distribution, which, in most cases, can provide information about *Pr*[*O*]. In other cases, probability distributions can support forecasts of what to expect over time. Consider the following problem suggested by Robert Gregson (personal communication, 2008):

> You are a medic who for the first time is doing volunteer work in Africa, and you get posted to a lonely hospital. When you arrive there are 10 dead bodies of Eboli virus victims in the hospital courtyard waiting to be buried. You have to order drug kits for possible *N* other patients who have not yet arrived. Kits are scarce and very expensive. How many kits do you order? How did you arrive at your estimate of *N*?

If the medic were to answer this question with the hopes of any degree of accuracy, it would be helpful to know whether the medic entered the situation at the beginning, middle, or end of an epidemic cycle. In many cases epidemic cycles start at low frequency, then the frequency builds to a peak and declines. The next question would be what sort of *Pr*[*O*] produced the 10 dead bodies in the courtyard, and how quickly did the 10 casualties accumulate?

Nonoptimization Decisions

Nonoptimizing decisions are not predicated on choosing a best option out of several options. In these contexts it matters only that a decision will fulfill an objective. Arguably, the nonoptimizing situations might involve subdecisions that have an optimization component to them.

Planning

Here the thinkers are defining a sequence of actions that are necessary to fulfill an objective. There may be more than one viable solution, and if there is sufficient time, alternatives can be explored. We often think of plans as cognition sequences where contingencies and options have been identified and selected in advance of any actions taken. There are times, however, where unexpected events occur, and either a Plan B goes into effect or the operator must make instant revisions on the fly.

In many cases, it is sufficient to say, "This will work." In emergency situations, where time is not a luxury, "minutely wrong decisions are better than no decisions at all" (Flin, Slaven, & Stewart, 1996, p. 272). Examples of the latter case are known as "go, no-go" decisions.

Predicting a Future State

One form of nonoptimizing decision is to determine what a system will be doing in the future based on knowledge of its situation at the present (Eberts & Salvendy, 1986). The actions that one might take in response to this prediction may involve an optimization component, but the prediction of the future state is, by itself, not an optimizing situation. Evaluating a situation and anticipating needs is thus one large group of nonoptimizing decisions that appear in both benign and emergency conditions.

In military operations, many courses of action are the result of a quick assessment of a situation and the selection of a single course of action that is known to work under the given circumstances. The decision maker must think through the intended operation first, however, to ascertain whether some plausible configuration of events would prevent the intended result (Klein, 1989). This thinking strategy is known as *recognition-primed decision making*, and could account for more than half of the decisions made in NASA flight simulators (Flin et al., 1996).

Divergent Thinking

Whereas optimization problems require the selection of the best option, divergent thinking, which characterizes the creative thinking process, involves the identification of many possible solutions or courses of action (Guilford, 1967). Divergent thinking is eminently valuable in the system design process (Mavrommati, 2001) and in unique situations where recognition-primed decisions are not available.

Two decades ago, however, divergent thinking resided outside the realm of decisions relevant to human factors. The advent of new types of computer programs that are intended to assist engineering design and creative exploration of options has brought creative thinking into another episode of human–machine interaction.

Other types of nonoptimizing decisions involve maintaining communications, delegating authority, and dealing with stress issues (Flin et al., 1996). The role of stress in human performance is addressed in Chapter 9.

Production Paradox

Somewhere among optimizing, nonoptimizing, and suboptimizing is the *production paradox* in technology choice (Carroll & Rosson, 1987). In this class of situations, humans have already become accustomed to their machine systems and learned how to obtain maximum production and reliability out of them. A new technology choice is introduced as a possible productivity enhancement. Although the new production system may indeed work out well, the users of the current system are reluctant to disrupt their well-worn and successful production strategies to learn the new system and make the usual learning mistakes along the way. The production paradox thus results in switching costs on a larger scale.

Thus, optimality may be different in the short run than in the long run. A person–machine system that is making a cold start-up may be better off with the new system. A system in operation, however, would have to reach a utility crossover point at which the expected gains from changing the system sufficiently outweigh the disruption and inconvenience involved.

Troubleshooting

Things break or stop working, and someone has to figure out why. Buck and Kantowitz (1983) identified some basic cognitive strategies for system troubleshooting, which may be assisted to greater or lesser extents by diagnostic tools and equipment.

Fault Isolation

A system consists of parts, and parts have functions. System parts may have many subparts with more specific functions. Diagnosticians who have a good working knowledge of a system's components can isolate a problem. The problem within the problem, however, is to determine how broadly or specifically the problem can be defined. A broad-level diagnosis may be quick to perform, but a detailed diagnosis may become extremely labor intensive and thus costly to repair before the repairs actually begin. Systems that are designed in a modular fashion can allow the replacement of chunks of the equipment without having to resort to a repair that is targeted to the smallest possible failing components.

"The car won't start" is not a very helpful problem statement. The noises that the car makes (or refuses to make) will help, however, to isolate the problem to a dead battery, a faulty ignition, or a faulty fuel-injection system. In this case the diagnostician is following a sequence of automotive system performances to isolate the failure. On the other hand, if the car starts and the "check engine" light goes on, another automotive subsystem is indicated, although the possibilities are still numerous.

Template Matching

Template matching is similar to recognition-primed decision making. Some user manuals for consumer products contain a chart of common faults with common fixes. They take the form "If fault X is present, try correction Y." Although troubleshooting pages that appear in the operating instruction for consumer products only address simple or peripheral repairs, a system diagnostician of greater skill may rely on the same kind of thinking but can reach a greater level of specificity in the diagnosis. This writer has been amazed more than once how many diagnostic windows exist on his computer, which only became known through a telephone call to the repair folks.

Statistical Template Matching

Sometimes an observed fault indicates more than one problem. The diagnostician then plays the odds. The most likely cause of the fault will be checked first, followed by the less likely possibilities. In such situations, it helps to have an odds table handy, mentally if nowhere else, which would be based on previous experience with the same or similar systems.

Sometimes the statement of one system fault does not provide enough information. It may be helpful to know what other parts of the system are still working. Based on that additional knowledge, the odds of one cause or another being responsible for the fault may shift.

Cognitive Workload

At one point in history, it was convenient to make statements such as, "The human information-processing capacity is 10 bits per second." This was convenient, but wrong, as it turned out. Rather, the number of bits per second that a person can process is highly dependent on the type of information that is involved as well as with the person's facility (based on knowledge and skill) for working with it. Thus, the relevant issues of human mental capacity must be considered in relative terms. Relativism, however, produced workload measurements that were incompatible across types of tasks (Lin & Cai, 2009; Morineau, Frénod, Blanche, & Tobin, 2009; Neerincx & Griffioen, 1996). For instance, cognitive workload for air traffic controllers is directly related to the number of planes taking off or landing in a given time period (Chatterji & Sridhar, 2001; Loft, Sanderson, Neal, & Mooij, 2007); such conclusions are not particularly generalizable to situations that do not involve airplanes. The following summarizes the implications of limited capacity and variable capacity theories of mental workload (Kantowitz, 1985).

Channels and Stages

Figure 6.5 depicts three basic configurations of cognitive channels and stages. The *serial process* is the simplest. There is a sequence of mental operations that usually begins with an input of some sort and ends with a product of some sort. That *parallel process* indicates two mental operations going on simultaneously. The two processes could be completely unrelated tasks, or they could be related such as processing the audio and visual cues of a situation while writing something down or clicking a mouse at an auspicious moment.

The *hybrid process* contains aspects of serial and parallel processes. Of importance, there is a bottleneck in the process whereby two mental streams must converge with their products at the same place in time. Bottlenecks are likely places where a process can become seriously slowed or derailed.

Although Figure 6.5 is intended to represent varieties of cognitive processes within one particular human, it can be used to depict production systems that involve multiple person–machine systems just as well. The latter constitute complex systems, which are considered in Chapter 13.

Limited Capacity Theory

The limited and variable capacity theories of cognitive capacity are considered next. According to Kantowitz's (1985) review, both views have support from cognitive experiments. According to limited capacity theory, there is a rigid fixed upper limit to the human mental channel capacity. The total capacity may be divided among primary, secondary, and even tertiary tasks or channels, where mental efforts are allocated among the tasks.

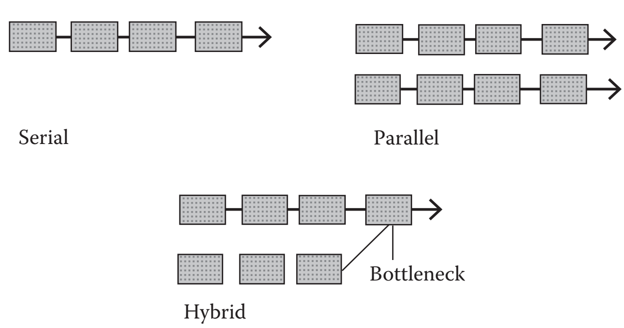

FIGURE 6.5
Configurations of channels and stages.

The allocation function is thought to function like a time-sharing computer; this is a metaphor that was held over from the days when mainframe computers were king. The mainframe would work on jobs that were submitted by many users at the same time. A certain amount of time would be allocated to processing each person's job. Run times for a job on a mainframe were enormously faster after midnight when most system users were not working.

In any case, if a person's total channel capacity was really 10 bits per second, and the primary task only required 6 bits per second, then 4 bits per second were left over for the secondary task. If the secondary task did not require the full 4 bits, then some small amount was left over for a tertiary task.

Task difficulty is defined in terms of the number of bits per second that a person can process. Note here that difficulty and complexity are different concepts. Complexity only refers to the number of channels and stages involved in a task. Difficulty pertains to transmission speed. If two easy tasks required a total processing bandwidth that was less than total human capacity, there would be no slowing of performance on either task. If one task was more difficult, however, slowing would be observed.

The limited capacity theory would identify two kinds of bottleneck conditions. In one case, if the human tried to process more jobs simultaneously than the total channel capacity would allow, any job in progress would be slowed. Another type of bottleneck point occurs where two parts of a parallel processes converge in a hybrid system. An added increment of processing time is related to the integration of the two flows of information in addition to what is required to process each one separately.

The implicit assumption of the limited channel capacity theory is that the limits to channel capacity are fixed by the nature of the channels. Naturally occurring channels in human cognitive processes may be related to the use of different brain functions, such as pictures versus words, or visual tracks versus auditory tracks. If the goal is to process pictures and words in two channels, the process might flow freely. A good picture could help the words to make more sense; hence, we illustrate textbooks. On the other hand, if the two channels of communication involve processing two conversations at once, the total information is crammed into a narrower channel and might not be processed well at all. Hence, we hear the familiar phrase, "One at a time, please! I can't listen to you both at the same time!" The colloquialism *information overload* applies to both types of situation.

Variable Capacity Theory

In variable capacity theory, processing space is allocated on an intentional basis. The operator sets priorities and thus allocates processing space to the incoming tasks in the desired order. The operator is then assumed to define tasks as those that must be done first, those that get some attention while the first task is really getting done, and those that benefit from a little downtime on the first two tasks.

Importantly, maximum channel capacity increases with demand. There are no firm rules as to when a bottleneck will occur, but an upper limit to the human's capacity can be reached eventually. Ralph, Gray, and Schoelles (2010) likened the phenomenon to squeezing a balloon: Constraints in one place produce stretching in another place, but one cannot predict exactly where in advance; eventually, too much constriction pops the balloon.

Another partial explanation favoring variable upper limits involves coping or resilience (Hancock & Warm, 1987, Harris, Hancock, & Harris, 2005; Hockey, 1997; Matthews & Campbell, 2009; Sheridan, 2008), but these processes are not sufficiently well defined yet.

Part of the conceptual problem is that there could be *many* possible adaptive responses that someone could make to control fatigue or excessive workload demands, some of which might be afforded in some situations but not in others. For example, people working under fatigue conditions often respond by slowing down their output, that is, reducing speed stress (Lorist & Faber, 2011). How many ways are there to regroup a task process? The question is explored later in this presentation in conjunction with the concept of degrees of freedom. Task switching is another source of possible strategies.

Resource Competition Model

According to Wickens (2002, 2008a), the decrement in performance that occurs when two or more tasks are performed simultaneously depends on whether the tasks draw on the same perceptual, cognitive, or psychomotor resources. Tasks are performed in stages, and if the stages are cascaded so that they do not absorb the same resources simultaneously, so much the better for performance overall. The relationship between task resources and stages is shown in Figure 6.6. This explanation explains how and when dual task or multitask performances can be effective. It does not yet explain, however, the variability in individual load capacity when a single, large intact task is performed and capacity limits could be reached. To attempt to answer questions about load in those situations one needs to look more deeply into the mental processes involved.

Multitasking

A good quantity of theoretical and applied research on cognitive workload utilizes the *dual-task methodology*. The usual objective here is to determine the workload impact of a particular task or to compare two alternative tasks. The first step is to define a task for the participants that all participants will perform. This task is intended to soak up a large amount of channel capacity so that the critical second task will push against the cognitive boundaries. The dependent measure is usually the response time or accuracy associated with the second task. Consider an example of dual-task methodology: Thackray and Touchstone (1991) gave their research participants a simulated air traffic control task as a primary task. The secondary task was to respond to another signal for of intrusions into the airspace. The intrusions were presented in two experimental conditions, flashing lights and colored lights. Flashing lights produced better detection over a prolonged period.

Multitasking is simply a case of doing multiple tasks at once, and has become a popular buzzword in recent years. One difference between multitasking and the dual-task methodology is that multitasking in the way most people do it does not clearly designate a

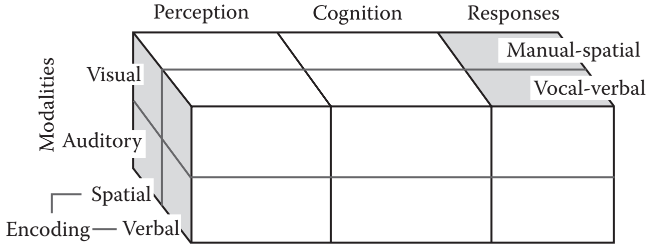

FIGURE 6.6
Resource competition model based on C. D. Wickens. (*Theoretical Issues in Ergonomics Science, 3*, 163, 2002).

primary or a secondary tasks. Another is that it is not so methodological. Any two tasks that need to be done can be combined into a multitasking event, and when is completed another task can be engaged to the extent possible. The key words, however, are "to the extent possible." There are limits to channel capacity, and the simultaneity of the two tasks can be problematic as well. If one task involves a flow of inputs over time that is slow and irregular, there is more opportunity to fit in a second task. If the input flow from one task is high and steady, however, there is less opportunity to work on the second task.

Success at multitasking often depends on the timing demands for the two tasks. De Pontbriand, Allender, and Doyle (2008) distinguish between explicit and implicit timing demands. *Explicit timing demands* are system-driven such that the pace of the inflow is controlled by the system and the person needs to respond to events quickly. This is again a scenario that does not favor performance on the second task. *Implicit timing demands* are set by the individual who finds a way to coordinate two tasks by managing the wait times on the two tasks. Sometimes it is possible to let work on one task pile up to a critical mass before switching tasks.

Task switching is thus a part of multitasking, and it can become more critical in cases where the operator must tend to the affairs of several machines simultaneously, especially when the need arises to shift focus from one type of machine to another. When the requirements of the two operations are very different, there is a greater demand on working memory as a result.

One of the more controversial scenarios for multitasking is found in the contemporary automobile where resource competition is high. According to Angell and Lee (2011, p. 3):

> In contrast with 1977—when the first vehicle to implement a microprocessor was introduced (controlling for a single function)—today's high-end vehicles contain 70 to 100 distributed computer processors that run more than 100 million lines of code, 20 million of which may be required for the navigation system alone …. And … drivers bring their own technology into the vehicle to use while driving: cell phones, smartphones, music players, and portable navigation systems, just to name a few.

Numerous studies published in the past decade have assessed the impact of distracting devices on driving. The key points that arise from the research are that (a) the nondriving tasks are visually demanding and sometimes cognitively demanding as well. (b) The distractions slow response times to suddenly appearing road hazards. (c) The drivers' attention becomes restricted such that they are most likely to miss events in their peripheral vision rather than in their central visual field (Horrey, 2011). The new technologies produce substantially different demands compared with the old-school distractions such as talking to a passenger, playing the radio, or lighting a cigarette with the in-vehicle lighter. Rather the focus of attention is often shifted away from the user-road interaction to the outside world that is not physically present (Strayer, Drews, & Johnston, 2003). The driving error rate associated with cell phone use while driving is comparable to the error rates associated with driving under the influence of alcohol at the legal-critical rate of 0.8% BAC (Strayer, Drews, & Croutch, 2006). Furthermore, hands-free telephone devices, while required in some states, do not mitigate the driving errors.

Text-messaging while driving is perhaps the most controversial distraction at present; 34 out of 50 U.S. states have laws prohibiting it, and 7 others have a partial ban. Of the other distractions beyond the cell phone, text messaging tops the list with a relative risk of a crash or near-crash in excess of 2300% while driving a truck compared with driving without using the device (Dingus, Hanowski, & Klauer, 2011). Heavy trucks require more attentional demand than standard automobiles. A comparable figure for texting while driving an automobile has not been consolidated yet, perhaps because of the legal ban in most areas. The use of cryptic, highly abbreviated, and often misspelled words, or "text-speak" puts an added demand on the person who is receiving and trying to understand the text (Head, Helton, Russell, & Neumann, 2012). Similarly, people do other questionable things behind the wheel: applying make-up and other personal grooming, or reading books, work papers, and newspapers while driving are good for a relative crash risk of 300%–450% (Dingus et al., 2011).

Automatization of Cognitive Processes

Telegraph Operation

Two of the first empirical studies that could be considered human factors studies pertained to the effect of training on the proficiency of telegraph operations (Bryan & Harter 1897, 1899). Telegraph operation required the use of Morse code to encode, transmit, and decode messages. Morse code required the operator to encode each letter of the alphabet into a system of long and short strikes on a telegraph key. The same messages would eventually be sent by telephone or telegrams that used the standard alphabet a couple of technology revolutions later, but in the days of the telegraph, channel bandwidth was very narrow if the message had to transmit a long distance.

Bryan and Harter's (1897, 1899) results are redrawn in Figure 6.7. Their first finding was that the speed by which a message could be sent was generally faster than the speed with which it could be decoded. Decoding required writing the message down by hand, and a skilled operator was often required to keep an entire sentence in short-term memory while

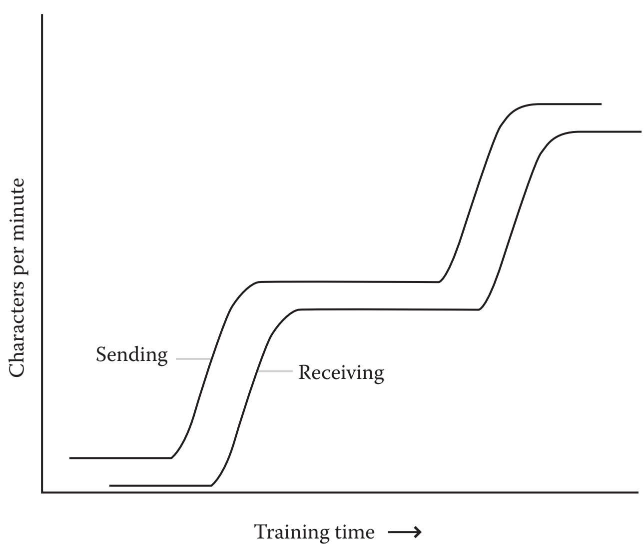

FIGURE 6.7
Automatization of telegraph operation, a cognitive task.

writing it down. Their second basic finding pertained to the bumps shown in the learning curves in Figure 6.4, which are not consistent and gradual. At first the operators would send and receive messages on a letter-by-letter basis. After a period, they could mentally manipulate entire words, beginning with the familiar words such as *train*, *arriving*, *late*, and so on. Additional training did not seem to improve for a while until suddenly the operators could manipulate entire phrases and short sentences.

Automatization was the general principle that emerged from Bryan and Harter's (1897, 1899) studies. Cognitive processes that may be separate in the early phases of training merge together into a flow of mental operation where the separate phases are not distinct to either the performer or the observer. Kantowitz (1985), in his review of research on channels and stages, also noted that stages of learning progress might not always be as distinct as Figure 6.5 would indicate. Nonetheless, we encounter at least one difference between novice and expert system users: Experts can work in larger chunks.

Controlled Processes

Ackerman's (1987) review of the literature connecting abilities to training results showed that some basic mental abilities and job-relevant knowledge bases did predict performance in the early stages of training, but some of them became less relevant as greater levels of training took over. Importantly, as learning progressed, task performance separated into automatic processes and controlled processes that still required problem solving that is specific to the situation. Later research showed that the executive function is more engaged in the early stages of learning; it then phases out to varying extents as learning progresses (Chein & Schneider, 2012).

Bryan and Harter's (1897, 1899) skill-acquisition curves indicate clear phase shifts in the underlying mental activity. Phase shifts in turn indicate that a self-organization process is taking place. According to Juarrero (1999; see also Van Orden, Holden, & Turvey, 2003), self-organization of mental activity is propelled by the intentionality of the mental activities. The thinker has a goal. A sequence of actions is taken, and feedback is received for each action that informs a "go, no-go" decision as to whether the next action in the sequence should be undertaken.

The foregoing analysis of automatization assumes that a visible control action (e.g., tapping the telegraph key) is taking place. Otherwise it would be very difficult to observe the mental process directly. Control actions are covered in Chapter 7, but for now it is only necessary to anticipate that control actions are behavioral results of the mental operations that are implicated in a particular work process.

Recognition Primed Decision Making

In military operations, many courses of action require a quick assessment of a situation and the selection of a single course of action that is known to work under the given circumstances. The decision maker must think through the intended operation first, however, to ascertain whether some plausible configuration of events would prevent the intended result (Klein, 1989). This thinking strategy is known as *recognition-primed decision making* (RPD), and could account for more than half of the decisions made in NASA flight simulators (Flin et al., 1996). In principle RPDs can eliminate time lost to discussion and debates, particularly in emergencies and promote the use of learned optimal responses (Bond & Cooper, 2006; Smith & Dowell, 2000).

The use of RPDs constitutes another distinction between experts from novices. The experts would have a larger inventory of experiences to draw upon, and can quickly evaluate a situation and pull out a response. The effectiveness of RPD is limited by the *coherence* of the problem situation. Put simply, clear problems more often lead to clear answers. Another limitation is that RPD is a mental shortcut where a solution is identified without fully analyzing the problem. The decision maker might in fact be trained to jump to incorrect conclusions.

The RPD is practically applied in the development of training programs, which should be organized in a three-step process. The first step is a cognitive task analysis of the specific situations that require training. The goal of the analysis is to isolate the knowledge base and decision processes of experts who have the greatest proficiency at the tasks. Further expansions of styles and methods for cognitive task analysis appear later in this chapter. The second step is to organize the learning objectives into modules so that elementary modules are presented first and followed by modules requiring a greater level of integration. The third step is to train the trainees and evaluate their own performance (Straszewski, 2004).

One salient application of the RPD technique resulted in a military training program for landmine detection equipment (Staszewski, 2004, 2006). The earlier generation of equipment worked well for detecting the earlier generation of metallic mines, but it performed poorly against the newer generation of landmines that replaced most of the metal with plastic and were often smaller in size. A design for detection equipment was developed with a new sensor system at a cost of $38M over nine years, but first-run tests with operators who were already trained on the older equipment showed that the new system was not appreciably better than the old one for detecting low-metallic mines. The solution was to redevelop the training program for the new equipment based on RPD principles. Results for the training treatment group showed an average accuracy of 85%–95% correct detection of low-metallic mines depending on the mine design compared with 15%–25% accuracy for the control group. Significant gains were recorded for the detection of metallic mines as well. Of course, one should bear in mind that a 5% miss rate could be deadly several times over in the course of a day's minesweeping operation.

Degrees of Freedom in Cognitive Task Organization

Self-organizing dynamics typically take the form of information flows among the subsystems. The concept of degrees of freedom was first introduced in conjunction with physical movements (Bernstein, 1967; Marken, 1991; Rosenbaum, Slotta, Vaughn, & Plandon, 1991; Turvey, 1990), and explains how fixed and variable upper limits to cognitive channel capacity are both viable. In any particular complex movement, each limb of the body is capable of moving in a limited number of ways, and the movements made by one limb restrict or facilitate movement by other limbs. For this reason, we do not walk by stepping both feet forward simultaneously, for instance. More generically, degrees of freedom are the number of component parts, such as muscles or neural networks, that could function differently to produce the final performance result. The notion of internally connected nodes of movement is substantially simpler and more efficient than assuming that all elements of movement are controlled by a central executive function (Turvey, 1990); see Figure 6.8.

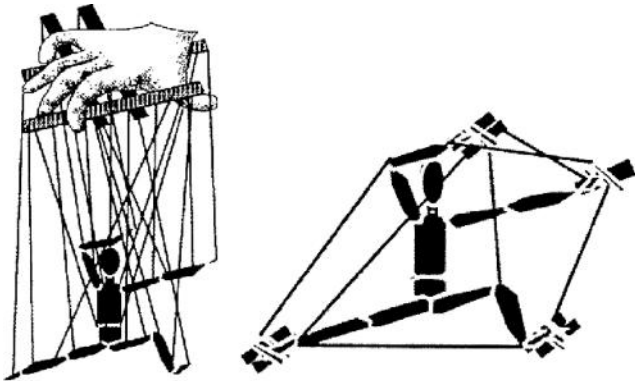

FIGURE 6.8
Degrees of freedom in (left) an executive-controlled and (right) a self-organized system. (Reprinted from M. T. Turvey, *American Psychologist*, *45*, 939, 1990. With permission of the American Psychological Association.)

When a movement is in its earliest stages of learning, the individual explores several optional combinations; but once learning consolidates, the movement combinations gravitate toward conserving degrees of freedom, which is essentially a path of least resistance (Hong, 2010). The gravitation is actually a self-organization dynamic involving muscles and neural pathways. Residual variability in the movement persists, however, to facilitate new adaptive responses. Substantial changes in goals or demands produce a phase shift in the motor movements, which are observed as discontinuous changes in the sense of catastrophe models. Cognitive behaviors are thought to operate on similar principles with regard to stages of schematic development, the role of executive functions, and the principle of conserving degrees of freedom (Hollis, Kloos, & Van Orden, 2009).

Although the full complement of possible degrees of freedom is not known at present, the current thinking about the structure of working memory and the various mental abilities should provide numerous clues regarding what could be involved in any particular task. Cognitive components would be involved in combination with psychomotor skills. Yet even simple tasks that require little cognitive input, such as reaching and grasping an object could involve testing possible neurological pathways before the individual locks on to a pathway that is most functional; this phase transition is important in rehabilitation contexts especially when connecting intention to actual movement (Nathan, Guastello, Prost, & Jeutter, 2012).

Cognition is often tied to action, particularly in situations that concern human factors. Cognitive psychologists would introduce the constructs of embedded and embodied at this juncture. A cognitive process is *embedded* in the sense that the individual is interacting with the environment when the process is occurring, either through manipulation of objects or navigation through space in search of some goal object or condition. The process is *embodied* in two respects: One is that natural cognitive processes, thought to be the outgrowth of evolution and which have the limits of various types, act automatically in what Gibson (1979) called invariant processes. The other aspect of embodiment is that the mental operations are in fact connected to the body that move through the physical world, and the body's available movements play a role in the formation of a behavioral schema that one can observe. Thus, the possible degrees of freedom in a cognitive process span the whole perception-action sequence. Here one might look to resource theory (Wickens, 2002) to unpack the channels and stages that could be arranged and rearranged.

When overload results from reaching a fixed upper limit of cognitive channel capacity, the sudden decline in performance would be the simple result of hitting a barrier. As such there would be little room for the elasticity associated with bifurcation effects or variable upper limits. If variable upper limits were operating, however, the principle of conserving degrees of freedom would have a few implications: In the case of adding tasks or new demands to existing tasks, a change in one cognitive–behavioral motion would impact on the other motions in the system or sequence. If it were possible to conserve degrees of freedom further, a phase shift in the cognition–action sequence would result in a catastrophic effect on performance. Bifurcation variables (see Chapter 2) would predispose the individual to produce the appropriate adaptation or remain unchanged. An example of an adaptation might occur when an increased demand for visual search results in a shift from an exhaustive search schema to an optimized first-terminating strategy (Townsend & Wegner, 2004). Fatigue could induce strategy changes also, a point that is explored in Chapter 9.

A similar process of juggling degrees of freedom occurs and is probably more observable when the individual plans sequences of discrete tasks in the context of a longer work period. Changing priorities could be involved. Task sequencing is essentially rule-governed, although not completely so. Again if there are N tasks there are $N - 1$ opportunities for a task selection with possibly multiple options at each switch point. The use of rules for task selection, nonetheless, reduces the number of degrees of freedom as a task sequence become more automatic. A simulation study by Walker and Dooley (1999) showed that a system is more stable in the sense of consistently high performance to the extent that the system does not have to respond to error corrections, interruptions, or disruptions of input flows. A complex adaptive system would be successful at responding to such events, but would have a few other rules (schemata) in place to do so.

Dynamic Decisions and Situation Awareness

Dynamic Decisions

Dynamic decisions involve a series of decisions that are not independent of each other, a problem situation that changes either autonomously or by virtue of the person–system interaction, and decisions that are made in real time (Brehmer, 2005, p. 77). The time-phased inflow of information induces dynamics that increase the complexity of the decision situation. Currently we know that time pressure, feedback delays, and reliability of incoming information place demands on the human operator that affect their performance (Brehmer, 1987; Jobidon, Rousseau, & Breton, 2005; Omodei, McLennan, & Wearing, 2005).

The computer programs that are typically used to generate scenarios for the study of dynamic decisions are alternatively known as "scaled worlds" or "low-fidelity simulations" (Schifflett, Elliott, Salas, & Coovert, 2004). As such there is a reduced concern for the realism of the peripheral features of the scenarios and a strong emphasis on the psychological constructs that the experimenter wants to assess. Realism is thus regarded as relative to the research objectives (Cooke & Shope, 2004). The systems lend themselves to reprogramming for desired experimental conditions.

There has been some expressed concern, however, whether the unreliability of the performance measures that have been used in research on dynamic decisions, which typically consist of a single number at the end of the simulation, is undermining attempts to test conventional hypotheses such as the relationship between general intelligence and performance (Brehmer, 2005; Elg, 2005). NDS theory would suggest here that the apparent unreliability of simulator performance measures could be related to the time-phased nature of the task and might not be a psychometric problem at all. As with other forms of individual and group learning, chaotic behavior occurs before the self-organization and stabilization at the levels of neural networks, individual behavior, and group work performance (Guastello, Bock, Caldwell, & Bond, 2005). Unlike the typical learning experiments, however, the specific decisions within dynamic decision sets are not independent of each other. Choices made in an early stage can affect options and utilities of options later on. The latter point is particularly true when the actions of one operator affect the actions of others in what could, or should, be done in a coordinated fashion.

Small world simulations are usually designed to be *opaque*, meaning that the individual working the simulation is not given explicit instructions regarding the rule structure or mental model embodied in the computer program. Rather, the individuals must figure all those things out for themselves. This form of learning is known in other contexts as *implicit learning*, which is a form of learning that takes place while the individual is explicitly learning to do something else (Frensch & Runger, 2003; Seger, 1994). Although the trainees might be less attentive to the implicit learning process than they are to explicit learning goals, they are acquiring knowledge about processes that are instrumental to executing the more obvious goals. One group of processes that arises often enough involves the control of dynamic systems.

Control of Dynamic Systems

One of the more classic simulations that involves opaque rules, implicit learning, and dynamic decisions is the beer distribution simulation by Sterman (1988). The participants' tasks were to place orders, receive deliveries from breweries, receive orders from customers, and make deliveries, all without overflowing the warehouse or running out of stock. Unpredictable events occurred meanwhile, such as "strikes," "transportation problems," and "demand shifts." Respondents' beer inventories were chaotic over time, and only 11% of the players were capable of maintaining a beer inventory between the two limits for the duration of the game.

The low rates of effectiveness in the beer simulation were interpreted as indicating that the cognitive skills needed to control chaos in real-world applications was in short supply (Guastello, 2002). Other studies with different experimental designs were leading to the same conclusion (Guastello, 2002; Heath, 2002). Gonzalez, Vanyukov, and Martin (2005) later extracted four principles of small world simulations—complexity, dynamics, dynamic complexity, and opaqueness—which is tantamount to saying "chaos that no one explained to you yet." The elusive skill apparently does not arise from greater working memory capacity. Lerch and Harter (2001) found that people with high working memory capacity tended to focus on a small range of the decision space very carefully, and not consider the full range of possibilities; their attention was overly focused and not sufficiently broad.

Sterman (1989), however, explained the results as a misinterpretation of feedback. Apparently, there is some opportunity for misinterpretation as Atkins, Wood, and Rutgers (2002) isolated three different types of feedback, that might not be readily distinguished by the operator: *process feedback*, which could take the form of a list of mistakes the operator made during a shift; *system feedback*, which is the function of the whole system including collateral events, and *outcome feedback*, which more clearly reflects the results of actions made by the operator to the extent that clarity is actually possible. An independent line of research shows that when feedback is controlling behavior in a continuous ongoing process, delays in feedback can produce a great deal of volatility in performance that would otherwise be relatively steady (Matsumoto & Szidarovszky, 2012).

Situation Awareness

Effective control of dynamic systems appears to require an accurate and sufficiently complex assessment of the situation. Thus, *situation awareness* evolved as a system design concept in its own right (Endsley, Bolte & Jones, 2003) and was defined as, "the perception of the elements in the environment … the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p. 36). It is a good example of the whole being greater than the sum of its parts. Although it is an important matter to design information displays well, the right *combination* of displays and controls gives the operator a sufficiently complete sense of everything meaningful that is going on in a form that makes immediate sense. As a very elementary example, the earliest automobiles did not have fuel gauges, but someone got the idea that a fuel gauge would be very helpful to the operator. Similarly, the earlier airplane pilots flew "by the seat of their pants" because they did not have an altimeter. How did these displays, even if crudely designed, help the operators gain awareness of their vehicle's situation and their own needed actions? Recent thinking on the nature of situation awareness recognizes it as both an individual and a group or team process. For present purposes, the rendition of situation awareness in this chapter is confined to individual-level experience. Group or team processes are explored in Chapter 13.

Endsley et al. (2003) elucidated eight principles of display design that contribute to situation awareness. The first principle is to organize displays around *"the operator's major goals*, rather than presenting it in a way that is technology-oriented—displayed based on the sensors and systems that generated the information" (p. 83, emphasis added). This is the principle behind smart displays covered in Chapter 4.

The second principle is to *present information directly*, rather than require the operator to make calculations to determine the needed nugget of information. The third principle is to provide *assistance for system projections*. This recommendation is essentially the principle behind predictive displays, which are thought to be of particular value to the less experienced operators.

The next three principles involve the global and local aspects of system operation. The fourth principle is to *display the global system status*. Operators could become too focused on information pertaining to limited portions of the system's functionality and could miss the big picture. The fifth principle, however, is to *support global-local trade-offs* where the emphasis is not letting the big picture overwhelm the recognition of more specific information that requires action. The sixth principle is to *support perception–action schemata*. Displayed information should, whenever possible, translate directly into control actions.

The last two principles recognize the limits of operators' cognitive capacity. The seventh principle, *taking advantage of parallel processing capability* recommends not overwhelming the visual system but instead making use of tactile or auditory displays. The last recommendation is to *filter information judiciously*. On the one hand, reducing information to the essentials lightens the demand on working memory capacity. On the other hand, dropping out the wrong details could have a negative impact on situational awareness.

Ultimately situation awareness is a psychological state rather than a property of the contributing displays (Stanton, Salmon, Walker, & Jenkins, 2010). Salmon et al. (2008) reported that different author teams in the topic area do not quite agree on the definition of situational awareness, with some emphasizing the tracking of information as it changes over time and combining it into working memory, while others (notably Smith & Hancock, 2005) emphasizing the embedded nature of the interaction between the individual and the environment. The keyword "situation awareness" has blossomed in journal publications during the 1994–2008 period, with nearly 10 times as many publications occurring in areas outside of human factors as within it (Patrick & Morgan, 2010), and it seems the basic concept has been co-opted in different ways.

The psychological components involve three levels of abstraction: (a) the elementary information appearing in displays, (b) the perception–action schemata at the middle range, and the global picture of the system's status, (c) the means of controlling it and the ability to envision future status that result from the control actions (Endsley et al., 2003; Patrick & Morgan, 2010). The optimal sequence of control actions could be as important as having the schemata themselves in place. Nevertheless, Endsley et al. (2003) developed a system for measuring situation awareness and evaluating the system by assessing operators' knowledge at all three levels of abstraction when given some test scenarios. The inferences from the evaluation of situational awareness are tied to the design of the contributing displays and possible needs for design modifications. The evaluation of awareness is not meant to be the same as an evaluation of system performance, however, and sometimes there is a conflict between the two evaluations (Wickens, 2008b). The use of smart displays and other elements of automation could serve to reduce workload, but there is, in principle, a critical point where the benefits of workload reduction counteract actual situation awareness (Patrick & Morgan, 2010; Wickens, 2008b).

The literature on situational awareness appears to assume that the functional mental model of the situation remains fixed while the particular situations within the model are changing. Such an assumption might work well in a relatively closed system. A more open system might not be so convenient, however. Sharma and Ivancevic (2010) considered the possibility that too much information could distort the global evaluation of situational awareness by channeling attention toward two or more possible global states. Their analysis was predicated on a computer simulation rather than a real-world example, but the bifurcation of awareness into multiple states is a distinct possibility. The uncertain associated with fast-moving and changing information could contribute to the fractioning of awareness.

Arecchi (2011) considered the scenarios where the switching of a mental model might be more sudden than gradual. Figure 6.9 shows a set of three mental models that are represented as distinct hills on a rugged landscape. Rugged landscape models are usually drawn to show high-fitness areas as peaks. Once a hypothetical organism arrives on a peak, there is little entropy of motion within the area, but it takes a substantial amount to hop to the next peak. The short peak in the diagram is a starting point for a mental model that is supported by existing information that continues to flow into the system. More information does not necessarily support the existing model or offer a better one until there is an insight on the part of the human. The insight involves the construction of *meaning* or new meaning. The construction of meaning and the switch to a new peak is not algorithmic in nature and thus does not lend itself to automation; it is a human thought process nonetheless.

The two tall peaks in the diagram represent another critical feature of the insight process—which new model of the situation do we choose? Arecchi characterizes this

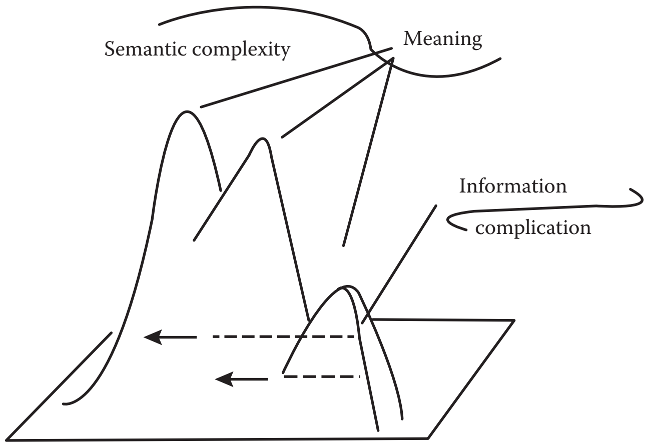

FIGURE 6.9
Rugged landscape model for insight. (Reprinted from F. T. Arecchi, *Nonlinear Dynamics, Psychology, and Life Sciences, 15*, 366, 2011. With permission of the Society for Chaos Theory in Psychology & Life Sciences.)

question as a Bayesean decision process. Given Condition X, what are the odds that Model A or Model B would be appropriate? The same question would characterize the shift from A to B or vice versa. Importantly, he further characterized such a distinct shift in the mental model as a paradigm shift. In the context of human–system operations, a paradigm shift would result in different control actions taking different priorities or new ones being invented. In deeper cases, it could reconfigure the smart portions of the smart displays. Although human factors examples that fit Arecchi's model have not yet been reported, the rugged landscape scenario stands, nonetheless, as a viable result of the behavior of complex systems when meaning is applied to increased information.

The Complex Adaptive System

The central ideas behind the complex adaptive system (CAS; Gell-Mann, 1994) combine recognition-primed decision making with principles of self-organization and degrees of freedom. A schema in Gell-Mann's lexicon is not really very different from the way the word was used in psychology decades previously. It is a perception-action module. What makes a CAS adaptive is its ability to receive and interpret new stimuli that could alter the schemata or propel the development of new ones.

The processes observed at the level of individual organisms can be observed at the group and organizational levels of analysis as well (Dooley, 1997, 2004). The inevitable next question is how the CAS concept informs us differently about fire-fighting, military operations, and other high-stakes circumstances. To be adaptive, the individual or group needs good sources of incoming information and a process of *sensemaking*. Endsley's (1995) "situational awareness" concept is really about the incoming information, insofar as it can be harnessed in human–computer interfaces and smart displays in the optimal combination, but it does not directly address the process by which the humans make sense out of it. Weick (2005) has probably gone the furthest to tackle the facet of sensemaking directly. An important principle for making any sense is to understand the interconnectedness of the various aspects of the environment and the agents within them. In essence, complex systems need to be understood as such (McDaniel & Driebe, 2005).

Cognitive Analysis of a Person–Machine System

Several writers noted quite some time ago (Eberts & Salvendy, 1986; Neerincx & Griffioen, 1996; Sanders, 1991) that the technology changes, mostly in the form of automation, shifted the demands on the human from physical demands to cognitive demands. More of the work that would be conducted by machines falls under the control of a smaller number of humans. Other forms of automation can have the impact of just moving the work around, and not actually making the cognitive load lighter (Tenner, 1996). The disadvantage to the trend in automation is that mental work capacity can reach its maximum more often than it ever did before. Thus, techniques for assessing mental workload would seem like a good idea.

Neerincx and Griffioen (1996) recommended that systems engineers take the totality of cognitive tasks performed by the operator into consideration, and not just focus on single tasks out of context. Tasks must be understood in a hierarchical context. The upper level context would take the form of a job description, such as the types mentioned here already. The next level of analysis would focus on particular tasks within the job description with the intent of describing the workload for each task. The third major step would be to construct a time chart showing the probable times and circumstances where a mental overload occurs. Obviously, corrective actions could involve revising the single task, but it may involve reorganizing the contributing tasks to prevent the transient times of mental overload, which may have a critical impact on system performance or safety. Thus, this section of the chapter starts with techniques for broader-level job descriptions, and then narrows the focus to the cognitive demands and workload assessments.

Job Descriptions

Job descriptions are building blocks of many personnel functions, most of which fall outside the realm of this book. For present purposes, however, it would be valuable to consider two common types of job description briefly and to reflect on the type of information they might provide. In both cases, cognitive processes and physical actions are both involved.

Functional Job Analysis

Functional job analyses consist of clearly written paragraphs that detail the worker's actions. The *Dictionary of Occupational Titles* (U.S. Department of Labor, 1977, with updated versions currently available online) utilizes this type of job analysis in an inventory of all known jobs that exist in the United States. The textual definitions are tagged with a hexadecimal numerical code. The first three numbers to the left of the decimal indicate an industry category. The three numbers to the right of the decimal indicate scales for data, things, and people.

The body of the definition usually consists of two main parts: a lead statement and a number of task element statements. The lead statement summarizes the entire occupation. It offers essential information such as: worker actions; the objective or purpose of the worker actions; machines, tools, equipment, or work aids used by the worker; materials used; products made; subject matter dealt with or services rendered; and instructions followed or judgments made.

Task element statements indicate the specific tasks the worker performs to accomplish the overall job purpose described in the lead statements. The sentences in the example beginning with "Turns handwheel … ," "Turns screws … ," "Sharpens doctor … ," "Aligns doctor … ," "Dips color … ," and so on are all task element statements. They indicate how workers actually carry out their duties.

The scales for data, things, and people (Table 6.1) denote the level of sophistication associated with each of the three categories. Fine and Getkate (1995) introduced four more scales to accompany functional job descriptions. The four scales pertain to Worker Instructions, Reasoning, Math, and Language. A job that rates a low level on Worker Instructions is one where all the goals and details are presented to the worker for execution. A job that rates a high level on Worker Instructions is one where the needs are presented to the worker, and the worker must figure out what information, tools, equipment, and communications are necessary to fill the need and to get the job done.

Reasoning ranges from simple algorithms, in which the worker executes the same steps in a specified order for each example of the task, to mental operations that require technical

TABLE 6.1

Dictionary of Occupational Titles Scales

Use of Data	
0	Synthesizing (integrating analysis of data to discover facts and/or trends)
1	Coordinating (determining time, place, and sequence of operations or actions)
2	Analyzing (examining and evaluating data)
3	Compiling (gathering, collating, or classifying information about data, people, things)
4	Computing (performing arithmetic operations and reporting on and/or carrying out a prescribed action in relation to them; does not include counting)
5	Copying (transcribing, entering, or posting data)
Interacting with People	
0	
0	Mentoring (advising, counseling, and/or guiding others)
1	Negotiating (working with others to arrive jointly at a decision, conclusion, or solutions)
2	Instructing (teaching subject matter to others, or training others)
3	Supervising (determining work procedures, maintaining harmony, promoting efficiency)
4	Diverting (amusing others)
5	Persuading (influencing others in favor of a product, service, or point of view)
6	Speaking–signaling (including giving assignments and/or directions to helpers)
7	Serving (immediately responding to the needs of requests of people or animals)
8	Taking instructions–helping (no responsibility involved; includes nonlearning helpers)
Use of Things	
0	Setting up (adjusting machines or equipment to prepare them to perform their functions)
1	Precision working (involves use of body members or tools and considerable judgment)
2	Operating–controlling (setting up and adjusting the machines as work progresses)
3	Driving–operating (involves controlling a machine that must be steered or guided)
4	Manipulating (using body members, tools, or devices and involves some judgment)
5	Tending (starting, stopping, and observing machines and equipment; little judgment)
6	Feeding-off bearing (inserting, throwing, dumping, or placing materials in or removing from machines or equipment that are automatic or tended by other workers)
7	Handling (using body members, hand tools, and/or special devices to work, move, or carry objects or materials); involves little or no latitude for judgment

reasoning and a large proportion of controlled process thinking. Math ranges from simple counting to advanced calculus. Language ranges from pointing to signs and checking badges to the writing and interpretation of cutting-edge scientific material.

Task-Based Job Analysis

Task-based job analysis involves making an inventory of tasks that a group of workers might be performing, and then asking workers how often they perform the particular task. The responses are factor analyzed to produce groupings of tasks that are likely to go together. Each worker is then likely to be performing one group of tasks to a greater extent than other groups of tasks.

The task-based analysis does assume that a master list of tasks can be compiled, and that all the tasks are defined to the same level of specificity. It may be necessary to follow up with a cognitive analysis to decompose each task into parts and examine the mental challenges in each part of the job.

Benchmark Jobs

A benchmark job is one that exists in many organizations, such that the management from the different organizations can compare notes as to how the job is done, and what combinations of human input and technologies work best under what conditions. In such circumstances, management may be comparing person–machine systems as whole units, and it may be up to the human factors engineer to figure out why certain combinations of people and machines work as well as they do (or do not).

Cognitive Task Analysis

Cognitive task analysis shifts the focus from the global definition of the job to the contents of the more specific tasks that comprise a job. Different tasks make different demands, and there was a useful insight years ago that the more enriched jobs—those that were more motivating overall—actually required a variety of skills and competencies (Hackman & Oldham, 1976). Other aspects of job enrichment are doing the whole task instead of only part of it, having sufficient autonomy to make job-related decisions instead of having to ask permission to respond to situations in any unique way, doing a job that has some impact on the work of other people, and deriving feedback about the results of one's efforts from the job itself. Although the old wisdom still holds true, the focus of attention here is on the cognitive demands.

To complicate matters further, the unique association between an individual and a cluster of tasks is a lot less prevalent in today's workforce than it was in the manufacturing-centered economy of the not-too-distant past, as job and cognitive task analysts have noted (Grant & Parker, 2009; Lesgold, 2000). Rather it is more often the case that tasks are performed by teams, although individuals often play unique roles within teams, and could be members of multiple project teams. Thus, once again, because the focus of this chapter is on individual cognitive processes, it is necessary to filter the various cognitive task analysis strategies that have been advanced over the years to those that speak to the individual experience, with emphasis on possible designs for person–system interaction and to represent some of the more distinctive varieties of approaches. A review of twenty *review* studies on cognitive task analysis by Schraagen, Chipman, and Shute (2000) was very useful in this endeavor.

Cognitive Inventory

Olson (1987) composed a useful list of cognitive functions that should be addressed in a cognitive analysis of a person–machine system. The basic functions include perceptual and motor skills, rule-based decision making, and the opportunities for analytic thinking and problem solving.

The first step in the analysis is to define the goals of the system and the tasks that are nested within it. Cognitive operations are identified in the second stage of the analysis. At the simplest level, the worker is simply moving information around. At the second level, information is transformed from one form to another. At the third level, information is processed through changes in format, but not in content. At the fourth level, information is interpreted; here the information is placed in a new format that adds some value to the data to facilitate the extraction of meaning. At the fifth level, data are analyzed so that new information is extracted. At the sixth level, a new information base is created.

Human assets and limitations are considered in the third stage of cognitive analysis. What are the important visual and auditory perceptions in the work situation, and are they taxing the extremes of human capability? What are the demands on motor skills, response time, and overall action? Response time and motor skills are considered further in Chapter 7. What are the demands on short- or long-term memory?

Hierarchy of Rules, Skills, and Knowledge

The contents of Table 6.1 illustrates a range of cognitive involvements ranging from simple to complex, which forms the basis of an approach advocated by Rasmussen, Pejtersen, and Goldstein (1994). At the simplest level, the operator follows simple rules—essentially perception–action schemata—reapeatedly and not much else. Skills fall in the middle of the spectrum. The word "skills" denotes individual differences in proficiency that comes with experience and practice. Skills combine with a variety of potentially applicable rules to execute tasks of a much greater level of complexity overall.

Knowledge sits on top of the hierarchy and informs the individual of what to do when novel circumstances arise, particularly when planning and design issues are involved. Although knowledge without some supporting skills does not usually go very far, the level of abstraction that is associated with a sound knowledge base can make an individual very effective at a variety of complex and demanding tasks that draw upon a particular knowledge base. As Figure 6.9 shows, insight in problem solving requires more than an accumulation of information. It requires a system of meanings that can reconfigure to accommodate a new context.

Hierarchy of Goals

The first instances of techniques for specifying systems of goals date back to the early 20th century (Adams, Rogers, & Fisk, 2012). The objective was to specify goals from the top-down, finally arriving at the simplest goals that it was possible to define. Although this style of goal setting has seen many incarnations in management practice, it tends to be light on specifying actions that will attain the goals, and even lighter on cognitive content.

A more actionable technique for human factors purposes is the method of goals, operators, methods, and selection rules (GOMS; Kieras, 1988, 2004). This method would take a task and decompose it into the most elementary actions with the idea of programming a computer to do part of the job. This goal of automating a process, unfortunately, requires an analysis that is as microscopic in perspective as the code, and it neither communicates a big picture to humans that might require one nor directly addresses cognitive processes in the action sequences. Adams et al. (2012) gave a pointed example for a short-order cook: The task goal hierarchy consisted of reading the order, preparing each ingredient to be used as part of another ingredient base, planning cooking time so the whole order will be ready at the same time, and so on. (Fortunately for some programmers, cooking is relatively modular if one looks at it the right way.) GOMS models were, nonetheless, friendly to LISREL and related symbolic computer languages that were fashionable in the day.

Ecological Task Approach

The ecological task approach (ETA; Kirlik, 2006) is so-named because it follows Gibson's (1979) basic premises of ecological psychology fairly closely. In this perspective, the task undergoing analysis is embedded in the environment more explicitly. Key questions for a task analysis would include areas of expertise, cues, and strategies used, and some consideration for why the task could be difficult sometimes or for some people. Expertise would involve all the forms of knowledge, skill, and rules considered by others, although additional areas of knowledge or skill that involve noticing changes in one's working environment would count also. For cues and strategies, Adams et al.'s (2012) example of the short-order cook makes a good comparison against GOMS: The cook would anticipate larger volumes of popular orders, plan the ingredients accordingly, juggle cooking times, and respond to the overall rowdiness of the hungry crowd to minimize their wait time. The difficulties often revolve around managing multiple orders simultaneously, staying calm, and keeping track of all the customers' antics.

Think-Aloud Technique

It would be tempting to forgo the job descriptions, scales, and laundry lists of cognitive functions and simply ask the operators what they are doing. Unfortunately, the verbal reports are likely to be incomplete and not necessarily in the actual order of the real functions. The experts in the group, who might be best equipped to tell the story, are also those whose cognitive processes are the most automatic. Thus, Card, Moran, and Newell (1983) developed the think-aloud technique for analyzing the cognitive processes in a human–computer interaction. The basic concept can apply to other types of systems as well.

The think-aloud technique asks the operators to explain what they are doing three times: before doing it, while doing it, and after doing it. Surprisingly, the three reports may come out a bit differently. Operators may initially stick a couple of steps together and take them apart on a later account. Some parts of the process may have been left out of the first report, but the actions themselves can bring some of the important aspects of the work to mind.

One should bear in mind that the process of verbalization itself might introduce some distortions (Olson, 1987). If the task places a high demand on channel capacity, talking will slow down the execution of the task. Similarly, talking may interrupt the automatic processes and disrupt both the thinking process and the report.

On the other hand, if the task places a low demand on channel capacity, talking will function as a form of rehearsal, and thus enhance the transition of information from short-term memory to long-term memory compared with normal working conditions. If the task is performed in an environment where there is a great deal of noise, especially verbal noise from other people, verbalization of the task aids performance. This form of verbalization is called *shadowing*. In other words, talking to yourself helps.

Cognitive Workload Analysis

As mentioned earlier, the first attempts to quantify cognitive workload in terms of bits and bytes was quickly disbanded when it became clear that quantifications that were framed from a machine's perspective were mostly irrelevant to the human experience and processing capabilities. An analogous phenomenon exists in NDS theory regarding the quantification of complexity. Numerous metrics for quantifying complexity exist, but the questionable one at present is *algorithmic complexity*. A graphic image (which could be a picture of anything to represent anything) is complex to the extent that its GIF (compressed) file is large (minus the bytes associated with the heading). The lack of generalizability from machine language to all other systems where content matters should be apparent. Sprott (2003) writes,

> Whereas chaos is best quantified by the Lyapunov exponent, there is no universal measure … [or] definition of complexity. When an initially disordered system self-organizes, it becomes less complex since its entropy decreases … although you could also say that a completely random pattern has low complexity since it conveys little information. In this view, the complexity should increase and then decrease again as a system changes from completely random to highly ordered (p. 410).

Further within the confines of NDS, *complicated* would refer to many channels and stages of cognitive processing, but those components could operate in a predictable mechanistic fashion. A person within a complex system, however, is susceptible to influences from other subsystems and makes an impact on other elements (people) within the system. Similarly, complicated and complex cognitive operations within the person can be distinguished the same way. Importantly, cognitive systems are self-organized, and there is little reason to believe that the self-organization of elementary circuits is the same across individuals (Hollis et al., 2009), although it has been obvious in other ways that there is enough similarity to draw some useful conclusions about mental processes. Furthermore, without some similarity of process and structuring of content, people would not be able to talk to each other! Similarly, the carriers of workload have not been identified beyond what has been said already about channels, stages, automatization, the structure of working memory, and resource competition.

At the other extreme of things that do not work well for cognitive load assessment are the overly aggregated indices of workload such as the number of airplanes hovering over an airport, or the number of customers at a restaurant. Those measurements are very relative to a situation and are not informative for assessing workload in other contexts or for comparing person–machine systems for their relative impact on load. There are three classes of solutions available, nonetheless—behavioral, subjective, and physiological—which have been more viable.

Behavioral Indicators

The objectivity of behavioral indicators has always been attractive to psychological researchers. In a workload experiment, the participants would perform tasks under different conditions, and the dependent measures would be performance quantity or quality, error rate, or response time. One would then look for conditions that produced significant decreases in performance or response time or increases in error rates. The dual task methodology was developed for just this purpose. Although performance on the task may be situationally dependent, *changes* in performance and response time are more generalizable.

After testing enough task situations, it is possible to draw conclusions about system attributes that produce performance deficits, which is often the goal of the experimentation anyway.

The downside of behavioral indicators, however, is that they might not always indicate the unseen part of the process. People can juggle their available degrees of freedom in task performance to compensate for changes in load and keep their performance consistent as a result. Learning and automatization can play a role in keeping performance buoyant while workload is increasing (Hockey, 1997; Szalma & Teo, 2012). Thus, other indicators are often used.

Subjective Indicators

The most widely used subjective indicator of workload is the NASA Task Load Index (TLX) form (Hart & Staveland, 1988). Although the form is not always used in the same manner in all applications, a commonly used version consists of six 20-point scales. At critical points in the test situation the research participants mark their ratings of mental demand, physical demand, temporal demand, performance (How successful were you in accomplishing what you were asked to do?), effort (How hard did you have to work to accomplish your level of performance?), and frustration (How insecure, discouraged irritated, stressed, and annoyed were you?). The extreme ends of the scale are simply labeled "very low" and "very high" without any other anchors in the middle.

The TLX has good reliability, with Cronbach's alpha values of .81–.83 for internal consistency across the 6-item survey (Braarund, 2001). It has good sensitivity to experimentally induced changes in workload (Dey & Mann, 2010), and it correlates as expected with physiological measurements of workload (Funke, Knott, Salas, Pavlas, & Strang, 2012). Some variations on the scale format involved the use of 10-point of 100-point scales or a system of weighted scales. Neither the weighting system nor the numerical ranges on the scale seem to have made any difference in experimental results worth considering further.

The scale has also been used in a computer-delivered format rather than paper-and-pencil format, which seemed to make sense for evaluating experiences with a computer program. It appeared, however, that the computer-delivered format produced higher workload ratings than the paper-and-pencil version (Noyes & Bruneau, 2007).

Physiological Indicators

Electroencephalograms (EEGs) are measurements in electrical conductance of the brain that are taken from electrodes attached to the skull. Up to 64 electrode channels have been usually used, but modern high-density equipment can accommodate 128 site readings. EEGs have been useful for decades for diagnostic purposes or general study of the brain that can be done unobtrusively with humans. The chief limitations of the EEG medium are that (a) they are topical readings and are not likely to register functions from deeper within the brain, and (b) the persons under study have limited mobility with all the wires attaches to their heads connected to a machine that reads the data. Current wireless technology, however, has solved the second problem so that research participants only need to wear a light skull cap containing the sensors while performing tasks that might involve moving above or driving a vehicle. Military personnel can wear the sensor caps under their helmets while on maneuvers. EEG signals are apparently sensitive enough to make the question of deeper brain activity less of a concern to researchers studying workload.

The P300 wave is the signal that is most sensitive to changes in workload. Sensory readings are taken from the parietal lobe, and there is a lag of 300 msec between the onset of a stimulus and the electrical responses of interest. Sensor readings in the parietal-to-frontal and parietal-to-temporal areas have also been shown to activate differently in response to mental processes (McEvoy, Smith, & Gervins, 1998; Schmorrow & Stanney, 2008; Verbaten, Huyben & Kemner, 1997). The P300 signal increases in amplitude when a stimulus produces an attention demand, and decreases in amplitude when working memory load is higher (DiNocera, Ferlazzo, & Gentilomo, 1996; Fowler, 1994; Gontier et al., 2007; McEvoy et al., 1998; Morgan, Klein, Boehm, Shapiro, & Linden, 2008; Pratt, Willoughby, & Swick, 2011).

Alpha and beta waves, which are actually wave shapes, are also under investigation for usable responses to workload demands. Beta waves are prominent in a waking and alert state. They are more "tightly wound" than the alpha waves (Figure 6.10), which are prominent in a relaxed state and often just prior to falling asleep. Deeper stages of sleep produce yet other wave forms. The latest approaches to workload analysis involve observing the interplay between alpha and beta waves, where the alpha waves signal cognitive fatigue (Desari, Crowe, Ling, Zhu, & Ding, 2010; Schmorrow & Stanney, 2008). The P300 also shows variability over time on sustained attention tasks; Smallwood, Beach, Schooler, and Handy (2008) showed that subjective indicators, error incidence, and P300 amplitudes varied together over time leading to the conclusion that operators' level of awareness, attention, or focus naturally varies over time.

Augmented Cognition

The success with the response of the P300 wave and the alpha–beta distinction led to a broader plan of augmented cognition and adaptive interfaces. Augmented cognition takes the form of an adaptive paradigm and an integrative paradigm (Schmorrow & Stanney, 2008). In the adaptive paradigm, the system senses elevated workload and goes into automated mode. The automated mode turns off when workload returns to normal. The switching between automated and normal modes makes sense if the human performs better than the machine under normal circumstances. Otherwise, it would make better sense to design a better interface, and the need would go away.

The integrative paradigm would program the machine to detect stimuli that are high demand, for example, when the operator is analyzing or scanning a set of photographs.

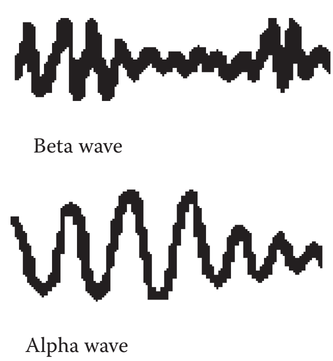

FIGURE 6.10
Examples of beta and alpha waves.

The machine would return high-demand stimuli to the operator for a second inspection. Importantly, the demand is detected from physiological data rather than from analyses of the stimuli themselves.

There are at least three unanswered questions at this juncture: (a) Given that attention and the P300 indications thereof wax and wane over time, how would the machine discern between a stimulus that might have received less than enough attention and normal fluctuation? (b) Is the technological vision here to test all tasks for P300 fluctuations and then use the knowledge some other way, or is it to have every operator everywhere wear a skullcap all day long so a machine can prod them into increased levels of attention? (c) EEGs contain a lot of unexplained variance, which is often characterized as "noise." Data analysis techniques for nonlinear dynamics have needed to address noise issues in physiological data, and algorithms have been developed to minimize the influence of noise on the data used to characterize dynamics whithin a time series (Guastello & Gregson, 2011; Shelhamer, 2007). How does the noise factor affect the reliability of the machine portion of a human–machine system that contains an augmented cognition component?

DISCUSSION QUESTIONS

1. Your car will not start. What do you observe that tells you how to localize the problem?
2. How would you build a computer program to assist with a classic (pick your favorite) optimization problem?
3. *Multitasking* is a contemporary management buzzword. What is it supposed to mean? How might you use multitasking efficiently in cases where: (a) you are sitting at a desk or computer workstation as your primary physical setting, and (b) you are required to move around and to do things other than press computer keys.
4. Is it possible to say, "In every nonoptimizing problem there is an optimizing problem"? Consider a nonoptimizing problem that you know something about. Does optimization play a role in any part of the thinking?
5. Consider texting while driving and your answer to the previous question. What are the implications for product liability suits? How about wrongful death suits?
6. The National Traffic and Highway Safety Association (NTHSA) currently sets an advisory limit of 30 characters for incoming text messages to a driver for nondriving information. Nondriving information would include messages from friends (who might be unaware that the recipient is driving) or coupons from advertisers. Is the 30-character limit reasonable? If not, what limit would you recommend instead? How did you come to your decision?
7. An individual who might be sympathetic to the information technology industries states that the question above regarding the 30-character limit is the wrong question to ask. Rather, we need to consider the trade-off between increased productivity and the risks of accidents. What do you think of this idea? Whose productivity are we talking about?
8. One industry observer suggests that the reason why drivers overstep their capacity limits while using nondriving technologies while driving is because they fantasize they are expert NASCAR drivers who seem capable of doing amazing things behind the wheel. Is this an accurate characterization of the conventional road user or the NASCAR driver?

9. Let's revisit the question about the virus kits: When you arrive at the rural hospital there are 10 dead bodies of Eboli virus victims in the hospital courtyard waiting to be buried. You have to order drug kits for possible *N* other patients who have not yet arrived. Kits are scarce and very expensive. How many kits do you order? How did you arrive at your estimate of *N*?
10. The TLX appears to produce higher ratings of workload when it is delivered through a computer application rather than by paper and pencil. Why do you suppose that might be the case?
11. Wireless EEG hats are promising tools for research on cognitive workload, but how would the findings from that type of research be implemented? Could it be done without the continued use of EEG hats? Do you think the day will come when everyone will be required to wear one all day every day? And do you think that idea would be popular?
12. How does noise affect the efficacy of augmented cognition systems?

7

Psychomotor Skill and Controls

This chapter focuses on control actions and the human responses that are most proximal to them. The first section describes the major issues that affect human response times, and concludes with an inventory of psychomotor skills; many of the latter extend beyond pushing buttons or turning knobs. Types of tactile machine controls are considered next, followed by issues pertaining to the complexity of the control actions themselves. Voice control is considered in the final section of the chapter. More controllers that are particularly germane to computer systems are covered in Chapter 11.

Reaction or Response Time

The reaction time between a stimulus and a response was a critical dependent measure in the earliest days of experimental psychology (Cattell, 1886). The thinking at the time was that the length of the delay between stimulus and response signified the complexity of the mental process that was taking place. Although there was a large element of truth to that supposition, we now know that some tasks are more complex or difficult than other tasks, and the number of bits per second of specific material that a person can process will greatly affect the reaction time. Indeed, we now use *response time* instead of *reaction time* to signify the totality of mental processes that must be occurring. *Reaction time* is now reserved for situations involving strictly psychomotor responses; the acronym RT is used here for either purpose. Response time is a critical outcome in HFE today, although it is sometimes known as *movement time* (MT) to signify human motions that are psychomotor in principle, but could involve more actions that are more complex than simply pushing a button.

Donders' RT

The first concept of RT that is relevant to HFE, and still relevant today, traces back to Donders in the late 19th century (Kantowitz & Sorkin, 1983). Donders contributed a format for an RT experiment that allows the experimenter to separate RT into two components: identification time and selection time.

The experiment requires three separate measurements. In the first, the human is presented with one stimulus and must make one possible response when the stimulus occurs; the time elapsed is $RT(A)$. In the second, the human is presented with two or more stimuli and the human must choose among two or more possible responses to the stimuli as they occur; the time elapsed is $RT(B)$. In the third, the human is presented with two or more stimuli, but only one stimulus maps onto a response; the time elapsed is $RT(C)$. Stimulus identification time is $RT(C) - RT(A)$. Response selection time is $RT(B) - RT(C)$.

In practical application, an experiment would involve a representative sample of people who are presented with many examples of stimuli in the A, B, and C conditions. The data analysis would compute averages for each person in the A, B, and C conditions. Afterward one could examine the statistical distributions for identification and response selection times, where there is one final observation for each person. Alternatively one could examine whether different stimulus and response combinations are producing markedly different RTs in the A, B, and C conditions. One might then aggregate the RT measures according to the type of stimuli or responses, make the statistical comparisons, and thus determine whether there are some aspects of the display-control (stimulus–response) set overall that need to be improved.

Type of Stimuli

There are several groups of determinants of RT. One is the type of stimuli that are involved. For instance, RT increases in a linear fashion as the amount of information increases if the information is composed of digits. RT increases in a logarithmic fashion as the amount of information increases if the information is composed of other types of information (Kantowitz & Sorkin, 1983).

Considering the scope of the possibilities for display design and content, the logarithmic relationship between RT and quantity of information is the general rule, and the linear function for digits is the special case. Note also that the clean relationships between quantity of information and RT that were reported in the experiments that Kantowitz and Sorkin (1983) summarized characterized information as the number of choices facing the operator. This operationalization of information stuck close to the Shannon (1948) definition. Cluttered displays, which might not contain any more useful information than an uncluttered version of the same display, induce larger RTs because of the time required to search for information (Yeh et al., 2003).

Stimulus–Response Compatibility

Consider the knob and dial combination in Figure 7.1. If the operator wanted to move the pointer from its present position to the location to the right, in which direction should

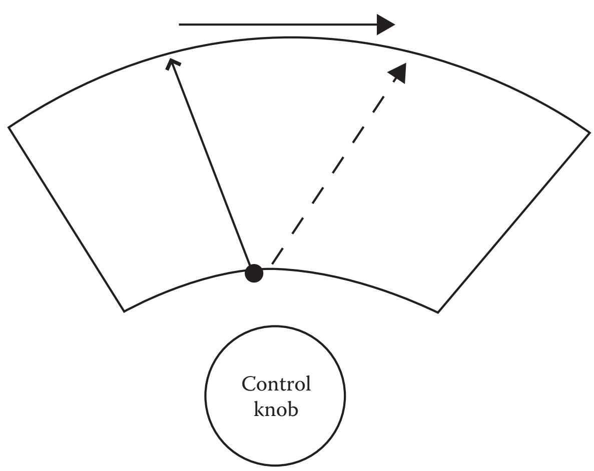

FIGURE 7.1
Stimulus–response compatibility: Which way should the knob be turned to move the pointer to the right?

the knob be turned? The knob would be turned to the right in a compatible situation. The knob would be turned to the left in an inversely compatible situation. Outboard motors on small craft are common examples of inverse compatibility. If the driver wants the boat to turn to the right, the tiller arm must be turned to the left.

In a completely incompatible situation, the knob in Figure 7.1 might not be a knob at all. It might be a pushbutton that could be pushed one or more times to move the pointer a small number of increments to the right. To move it to the left, one would need another button, or a toggle button with an up arrow and a down arrow drawn on the button. All other aspects of the situation being equal, RT is fastest when the stimulus and the response are compatible (Fitts & Seeger, 1953). RT would be second-fastest, but by a significant margin, when the stimulus and response are inversely compatible. Incompatible stimulus–response combinations produce the slowest RTs.

Population Stereotypes

A *population stereotype* is an expectation within a population of users that a system should be controlled in a particular way. The expectation would come from experience with different but similar situations. For instance, in a kitchen or bathroom or anywhere else, is the faucet handle for the hot water on the left or on the right? Imagine your surprise if you stopped at a cheap motel one night, tried to take a shower, and found that the opposite was true.

Similarly, if you wanted to screw a screw into a wall or into a device of some sort, which way would you turn it? Imagine my surprise when I encountered a toy with a screw that was lathed in the opposite direction. If a naive individual had not suggested, "Try turning it the other way," I would not have found the solution for a very long time.

In a more serious example, suppose two aircraft were approaching each other head-on. Which way should they turn? The correct answer here is to turn to the right; the convention is an outgrowth of maritime law for two ships approaching head-on, dating back a couple of centuries. Unfortunately, Berringer (1978) discovered that only 85% of professional pilots in a simulator turned right. The others turned left, and the mistake was unrelated to any of the pilots being left-handed. Consider the implications here when two pilots are involved: there is a 28% chance of a collision because one of them did not turn to the right, and a 2% chance of no collision occurring because they both made the wrong move.

There were also comparison groups of nonpilot students in the experiment, of whom 75% turned to the right when faced with a head-on collision. The conclusions were that a population stereotype existed among pilots and nonpilots and that safety in the skies would improve with training on this particular maneuver.

Learning and Skill Acquisition

Skill Acquisition

RT improves with practice. The curve shown in Figure 7.2 is an example of a skill acquisition curve in which the RT declines as the number of trials progresses. It is actually an inverse learning curve. Instead of showing an increased amount of learning over time, up to an asymptotic level, learning here is characterized as a decrease in performance time.

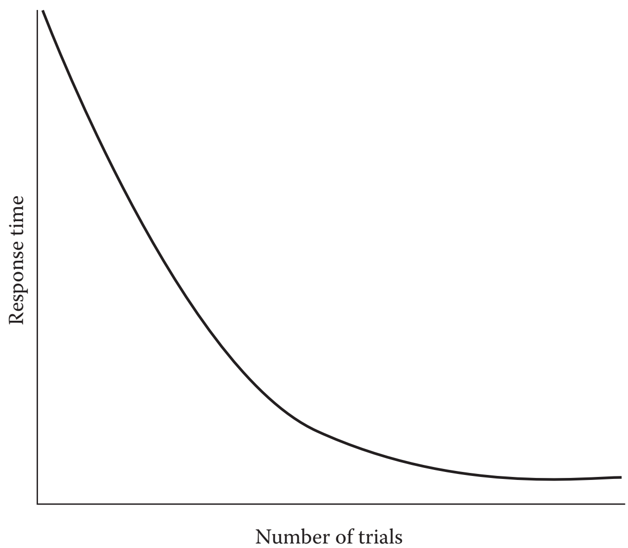

FIGURE 7.2
Skill learning curve.

Not evident from the curve is another long-standing relationship in learning theory: spaced trials produce better learning than massed trials. In other words, if one were to deliver 100 trials in a learning set, better results would be obtained if the 100 trials were spaced out over several days than if they were delivered all in one session.

The equation of the skill acquisition curve of the type shown in Figure 7.2 is

$$RT = k \log(P) + a, \tag{7.1}$$

where k and a are constants and P is the number of trials or units of practice. It can also be characterized as a power law relationship:

$$RT = aP^{-b}, \tag{7.2}$$

where b is the shape parameter for the power law function and signifies the rate by which training speeds up response time; a is a constant once again.

In the earlier times of human–computer interaction a good many commands (control functions) were typed into the system by a keyboard. Some thought was given to the brevity of commands under the thinking that shorter was better. McKendree and Anderson (1987) modeled the situation as

$$RT = \text{physical time} + \text{mental time} = nkP^{-a} + cmP^{-b}. \tag{7.3}$$

In this case, n = number of keystrokes, k = time per keystroke, P = units of practice once again, m = mental time, c is a constant; a and b are separate speed-up exponents associated with physical time and mental time, respectively.

Models such as Equations 7.1, 7.2, and 7.3 can be used to compare possible stimulus-control designs. One of the more generalizable findings from the keystroke research, however, was that fewer keystrokes are not always better. Rather, if the commands are too brief, the commands become ambiguous and errors in control operation or longer think times are induced (Landauer, Galotti, & Hartwell, 1983).

Dynamics of Learning Processes

Another point that is not evident from Figure 7.2 is how much variability exists within the learning curve; it is not as smooth as it is traditionally shown in textbooks. The nonlinear dynamics of learning can follow one of two basic patterns depending on one's interest and emphasis. The current thinking is that learning processes are chaotic in the early states before reaching the asymptote where self-organization occurs (Hoyert, 1992). The neurological explanation is that neural firing patterns are themselves chaotic in the early phases of learning while the brain is testing out possible synaptic pathways. Once learning has progressed sufficiently, the brain locks onto a particularly pathway to use consistently (Minelli, 2009; Skarda and Freeman, 1987). The variability and reduction thereof can be explained as the result of the number of degrees of freedom inherent in the neural pathways or neural–motor combinations that could comprise a motion (Hong, 2010, Chapter 6). The system gravitates toward minimizing the degrees of freedom during the learning process.

The second dynamic principle involves the cusp catastrophe model. If we extend the baseline of the learning curve (Figure 7.3, left) prior to the onset of the learning trials, two stable states are apparent; according to Frey and Sears (1978) hysteresis exists between learning and extinction curves cannot be explained otherwise. Different inflections in learning curves can be explained as a cusp bifurcation manifold (Guastello et al., 2005) as shown in Figure 7.3 (right). The cusp model for the learning process would be

$$dy/dt = y^3 - by - a, \tag{7.4}$$

where control parameter a (asymmetry, governing proximity to the sudden jump) is the ability of a person or the number of learning trials, and control parameter b (bifurcation, size of the sudden jump) would be the difference between treatment and control groups,

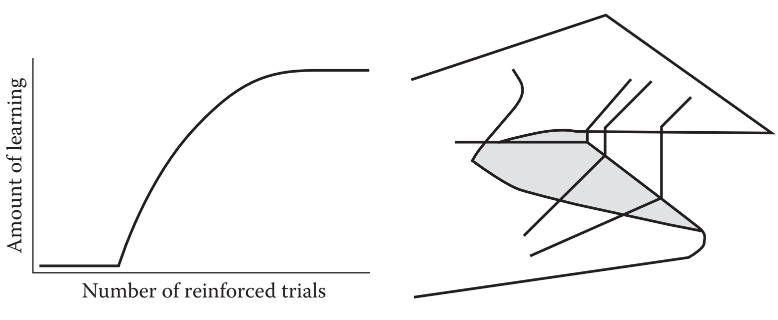

FIGURE 7.3
Cusp catastrophe model for learning events.

motivation, or differences in schedules of reinforcement, or any other variable that would contribute to making some learning curves stronger or steeper than others.

The cusp model is particularly good for training and program evaluation. If a statistical cusp effect turns out to be better than the next best alternative linear model it would denote all the features associated with a cusp model. Here the idea of *stable* end states adds a desirable feature to program evaluation: We want stable improvements to behavior targets, not simply statistically significant differences. "Stable" does not mean "without variability," however. A bit of variability is necessary if it will ever be possible for the person, group, or organization to attain greater levels of performance (Abbott et al., 2005; Mayer-Kress, Newell, & Liu, 2009). Figure 7.4 illustrates the dynamics of performance improvement. The person, group, or organization encounters a new task that cannot be readily assimilated into old or crystallized learning. With practice the new learning is attained, and the level of hysteresis across the cusp manifold increases with repeated new challenges.

Speed–Accuracy Trade-Off

"Haste makes waste," as the old saying goes. More specifically people can work faster than they usually do without making more mistakes, but only up to a critical point. After the critical point the error rates increase dramatically, as shown in Figure 7.5. The engineering strategy would be to find the critical point that is located just before the sharp increase in errors and set the work pace to that point.

It is noteworthy that the *S* shape of the curve in Figure 7.5 denotes a cusp catastrophe function. Work speed would function as the asymmetry parameter; speed brings the system closer to the critical point. The bifurcation factor would be one that promotes larger or smaller increases in errors. One possible bifurcation factor would be the number of

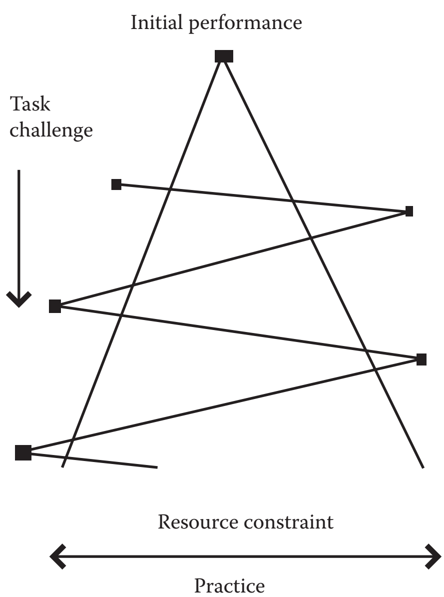

FIGURE 7.4
Hysteresis during performance improvement.

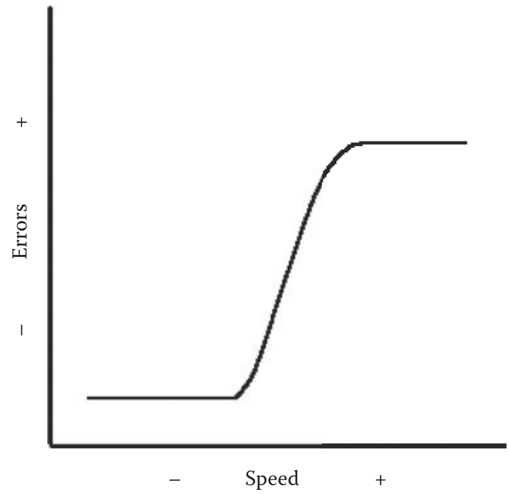

FIGURE 7.5
Speed–accuracy trade-off.

opportunities for error that are inherent in the work situation, such as a complex task with several steps in the process. This is the principle of degrees of freedom again. Another possible bifurcation factor would be the standard of accuracy that constitutes a meaningful error. In the process of manufacturing television sets, for instance, the system can tolerate 40 missed welds out of 10,000. That same standard of accuracy would be completely unacceptable for the assembly of guided missile systems.

There is evidence favoring the existence of multiple performance processes that explain work performance under conditions of different speeds. Signal detection tasks offer the opportunity for two types of errors, misses and false alarms. It is also well known that the errors can be readily biased in favor of one error or another by the rewards or penalties associated with them. Balci et al. (2011) designed an experiment that induced participants to make a trade-off between accuracy and rewards. They found that bias toward accuracy dominated the early stages of learning and performance and a bias toward higher overall rewards dominated later on.

In a visual search example, Zenger and Fahle (1997) manipulated load by increasing the search target area to produce a pressure on response time. Missed targets were more common than false alarms, and they were traceable to a speed–accuracy trade-off.

Szalma and Teo (2012) constructed a vigilance task that varied the number of screens containing possible targets and the speed of screen presentation. For that set of conditions, there was a linear decline in accuracy as a function of speed for low-demand content, but the classic breakpoint was observed as the number of screens increased from 1 to 8. False alarms declined with increasing speed for the 8-screen condition only.

According to Koelega, Brinkman, Hendriks, and Verbaten (1989), research prior to their study had shown that the density of targets and the memory load inherent in the task affected whether error rates set in during slow- or fast-working conditions. They assessed performance on four vigilance tasks that varied in memory demand and by type of stimuli for speed, accuracy, electrodermal response, and subjective ratings. A principal components analysis of the measurements produced two components that indicated that speed and accuracy could result from two different underlying cognitive processes. The outcomes of working too slowly are revisited in Chapter 9.

Taxonomy of Psychomotor Skills

Fleishman (1954) contributed a taxonomy of psychomotor skills that was based on a wide range of tasks performed by military personnel. A large sample of personnel performed the full range of tasks, and the taxonomy of 11 basic skills was produced from the factor analysis, as follows.

Control precision refers to the precise setting of a control, such as the calibration controller shown later in this chapter or the control–display combination in Figure 4.3 (upper left). *Multilimb coordination* involves the use of arms and legs simultaneously, such as one might do while playing the drums.

Response orientation is a person's first response to an alarm signal, such as the stimulus identification time in the basic Donders experiment. *Reaction time* is a separate psychomotor measurement that was discussed here already, except that the reactions were localized to the use of hands. *Speed of arm movement* is another distinct psychomotor skill.

Rate control involves following a moving object at the same speed as the object's, and changing direction when the object changes direction. Several types of tracking tasks fall into this category; the pursuit rotor task is an example. The operator must maintain contact, usually with a stick or other handheld device, with a sensor that is mounted on a wheel. The wheel spins at different speeds and changes direction. The criterion would be the length of time the operator can maintain pursuit without breaking contact.

Manual dexterity involves the production of complex motions with one's hands, such as the use of some of the more challenging hand tools. Similarly, *finger dexterity* involves the production of complex motions with one's fingers; tying a knot would be an example. *Arm–hand steadiness* occurs in tasks requiring the combined use of arms and hands. *Wrist–finger speed* involves the combined use of the wrist and finger in repeated motions, such as tapping keys on a keyboard. The last psychomotor skill on the list is *aiming*; this skill might be expressed in the use of firearms or in threading a needle.

Types of Manual Controls

Kantowitz and Sorkin (1983) characterized controls as discrete versus continuous and linear versus rotary. A discrete control is comparable to an on–off switch, or a similar switch with a fixed and small number of options. Examples would include pushbuttons and toggle switches. Some prime examples appear in Figures 7.6 through 7.8. The example in Figure 7.6 is one of the first pushbutton control designs that were introduced in the earliest part of the 20th century. It required two buttons to do its job, one for "on" and one for "off." When "on" was depressed, the "off" button popped out ready for the operator's next move.

The earliest light switches in homes worked like the example in Figure 7.6. The home unit often required less physical force than the controllers as some industrial systems did at this time. This design for light switches was eventually replaced by the two-pole toggle switch that has been in common use for at least the past 50 years. A different style of toggle switch appears in Figure 7.7, which requires the operator to turn the handle from one setting to another. This design is less likely to be swiped up or down, left or right, into the wrong position accidentally because of the arc of motion required to change the setting.

Figure 7.8 depicts two other forms of pushbutton that were encountered on an emergency override panel in an electrical power plant. The top-level control system was operated

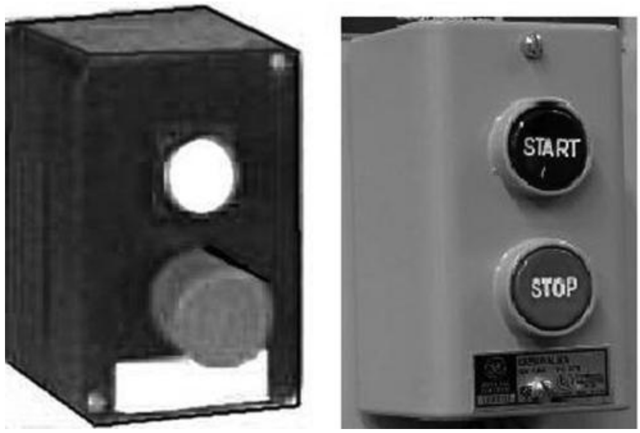

FIGURE 7.6
Early pushbutton design (left) and its modern counterpart (right).

FIGURE 7.7
Discrete control with two states.

from a computer terminal. The button on the left in Figure 7.8 is an example of a button that has been protected against accidental operation with a collar around the button; the finger pushing the button has to push below the outer surface. The button on the right is an emergency stop button. Note that the shape of the emergency button allows the palm of a hand to slam against the surface of the button at a moment's notice.

Figure 7.9 depicts a pushbutton panel in which the controls are hierarchically organized. The operator selects a mode. The name of the selected mode appears in the LED display at the top along with the numeric setting associated with it. The operator may then adjust

FIGURE 7.8
Pushbutton with protective collar and pushbutton designed for emergency access.

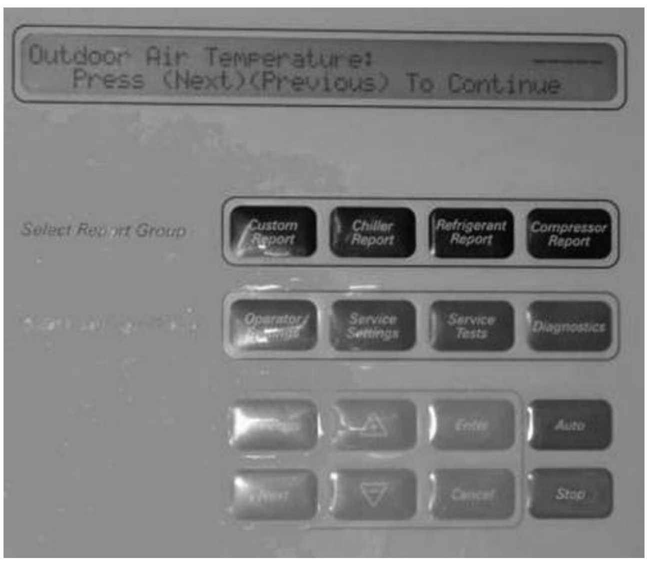

FIGURE 7.9
Discrete controls for a set of hierarchically organized functions with numerical settings.

the setting using the up and down buttons. The up and down buttons are located on the last two rows of buttons, second button from the left. The designers saw fit to put the up button above the down button.

Figure 7.10 looks like a common keyboard with additional function buttons. A keyboard is a familiar example of a system of discrete controls. This particular example comes from

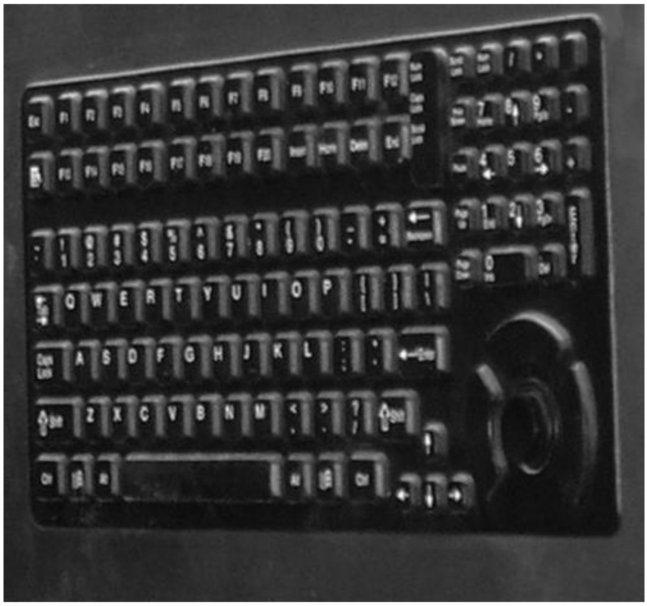

FIGURE 7.10
Wall-mounted keyboard with a generous supply of function buttons, numeric keypad, and position controller in the lower right portion of the panel.

an HVAC system in a municipal building complex. The top-level controls appear on computer terminals, which the humans monitor on a regular basis; control operations can be made from this point. The button panel shown in Figure 7.9 is mounted on one of the chiller devices, and changes to the behavior of that particular unit can be made on the unit itself. The keyboard in Figure 7.10 is an intermediate level of control between the device-specific controls and the global computer controller. At this point the operator can intervene into many aspects of the system's operation from the keyboard panel. This control panel was once the top-level controller, but it was eventually superseded by the computer system with a graphic user interface (GUI; also see Chapter 11).

Continuous controls typically take the form of knobs and slideswitches. A knob would be a rotary control because it turns. A slideswitch would do the same job as a knob, but it would require movement along a vertical or horizontal path. Figure 7.11 depicts a virtual slideswitch on a computer interface. It gives the appearance of an actual slideswitch although it is actually controlled less directly by a mouse. Continuous control motions in many cases can be entered through a keyboard: position the cursor, hit the desired keys, and then press "ENTER." The example in Figure 7.9 allows for operation through either the virtual slideswitch or the keyboard using the white panel directly below the slide.

The device shown in Figure 7.12 is also a rotary controller of the continuous type. It calibrates the flow of steam through a complex pipe system in a building complex. It was installed in the 1935 era and was still in use at the time the photograph was taken for this book.

The development of computer-based technologies brought a number of additional control types that are considered in Chapter 11. The buttons considered thus far have a real-time function, in which a flow of electricity starts or stops. In computer-based technologies, buttons shown on computer screens are only virtual representations of familiar

FIGURE 7.11
Virtual slideswitch emulates a real-world control.

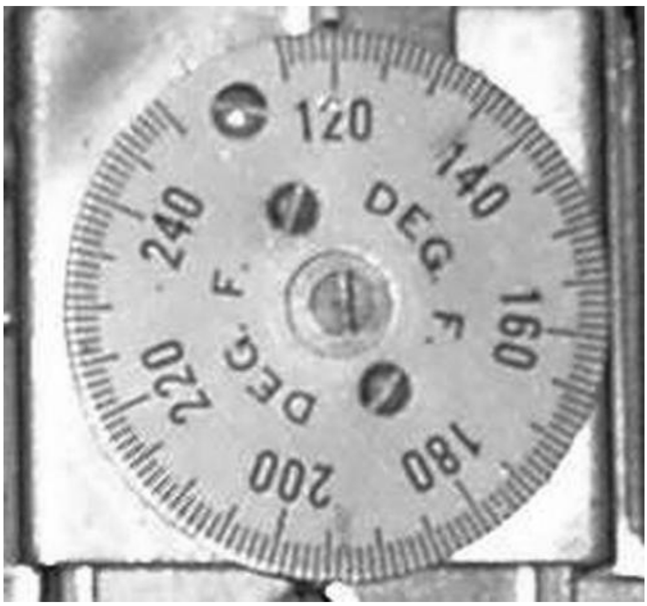

FIGURE 7.12
Continuous calibration control.

controls. The virtual nature of the control led to some newer varieties of controls including touch screens, gestural interfaces, and eye-tracking signals.

Multidimensional Controls

The most common example of a multidimensional control is one that moves an object around a two-dimensional space. The joystick in Figure 7.13 belongs to a security camera system that allows the operator to adjust the camera remotely. A panel of additional buttons is used in conjunction with the joystick, several of which pertain to zooming or

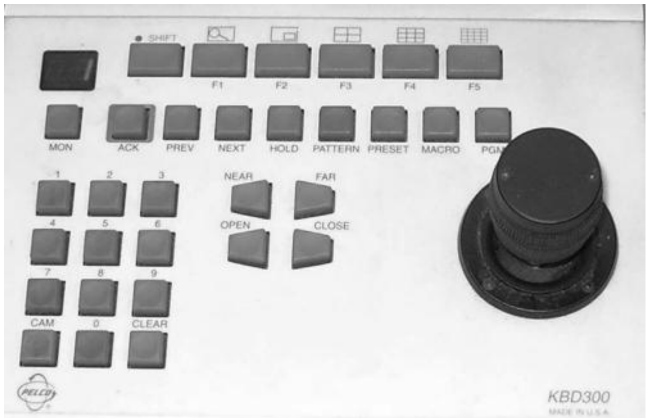

FIGURE 7.13
The joystick is a 2-D control for positioning a security camera. It is shown here with a set of discrete function buttons associated with the camera system.

camera selection. Two-dimensional buttons can do the same job in a continuous fashion. Up-, down-, left-, and right-arrow keys can often be substituted for a joystick or 2-D button; in this case, the movement of the object has been discretized.

The HFE wisdom in this case is that a two-dimensional control should be employed if it corresponds to a two-dimensional display. If the two intended parameters of control are depicted as one-dimensional displays, two one-dimensional controls should be used (Kantowitz & Sorkin, 1983). Figure 7.14 shows three renditions of a common two-dimensional control. The example on the left only navigates a menu with the enter (select)

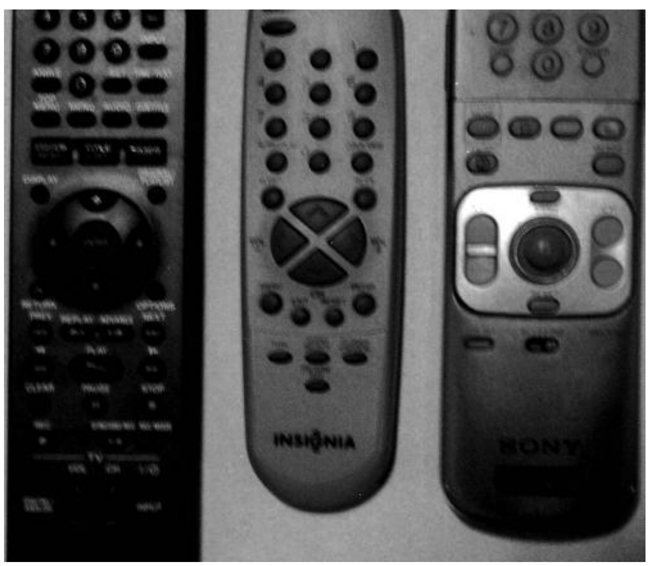

FIGURE 7.14
Two-dimensional controls for remote control of televisions.

button in the middle; the four points on the circle are analogous to the up, down, left, and right arrow keys on a common computer keyboard. The example in the middle uses the four-directional configuration to manipulate the two controls that are most often used on a television—channel (station) and volume; this control configuration runs counter to the standing wisdom because volume and channel are really independent functions. In the example on the right, also for a television, volume and channel are two separate one-dimensional controls. Fortunately, these are low-consequence household entertainment devices, and the user acclimates to the device quickly.

Another way to induce multidimensionality in a control system is by creating hierarchical arrangements of control functions using modes. If one controller has set the system into Mode A, the remainder of the buttons will perform functions, and generate displays, pertaining to Mode A. If the mode controller is set to Mode B, the remaining buttons will perform their usual functions, but with respect to Mode B. The danger here is that the operator must be fully aware of the mode in operation. If the operator imagines that the system is in Mode A, but it is really in Mode B, the correct information will not be introduced to Mode A and new wrong information will be introduced to Mode B.

Display design is critical to helping the operators keep track of the mode that is in operation. In one surprising and fatal airline crash, the cockpit crew wanted to insert information to the automatic control system that signified an angle of descent, measured in degrees. They did not notice that they were in the wrong mode and entered the numerals that they wanted, which were read by the system as speed of descent, measured in km per second. The only difference in the display for the two modes was the placement of an automatic decimal point, which the operators did not notice (Straub, 2001). The aircraft collided with the side of a mountain instead of flying over the mountain and gently descended to the airport that was located on the other side of it.

The cargo arm on the space shuttle requires a 3-D controller. The arm is manipulated from within the life-support pod while the action is taking place outside. The arm is manipulated through 3-D space until it reaches the grapple connection on the cargo object (perhaps a satellite). A second controller opens and shuts the grapple mechanism, and the operation of the arm in 3-D space continues to relocate the object in the desired position.

Size

Historically, large-size controls were large because they moved large physical forces. Their size was dictated by the physical configuration of the equipment. A sense of importance was thus attached to size. Figure 7.15 depicts an early 20th-century wheel that is now located in the basement of an electrical power plant. It is still available for use if the layers of automatic control fail to function.

The device shown in Figure 7.16 is a piece of equipment that converts physical activity to signals that the computer upstairs can process and display for the operator. Note that it needs to be precisely set and calibrated, and it has controls and a display of its own. Errors at this stage in the process of information flow can upset what is recorded and interpreted later on.

Size continues to have symbolic value with current technologies, where larger is more visible and probably more important to the system in some fashion. Size also matters if the operator is required to push a button or flip a switch when approaching from a distance, especially in an emergency. For a panel of pushbuttons or switches, each one should be large enough and separated far enough apart from the other buttons and switches to allow for accurate operation. The width of the human finger is a limiting size requirement. The relationship between size and distance is expanded further in this chapter in the context of Fitts' law.

FIGURE 7.15
Large controls like this one control large mechanical forces.

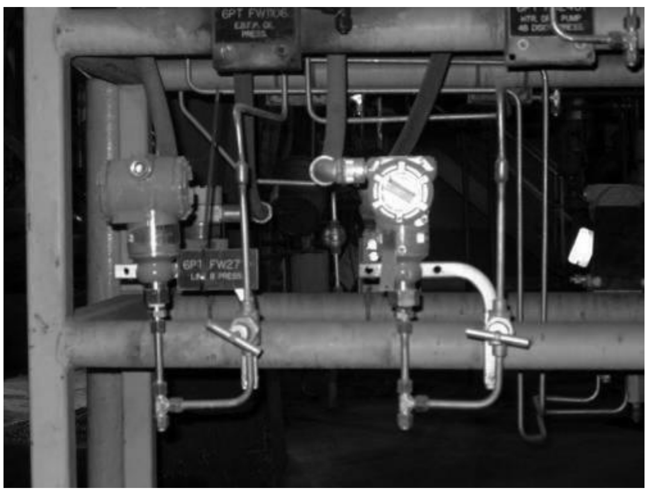

FIGURE 7.16
Equipment that translates a physical activity of a system into something that a computer can process.

Touch screens have been replacing button panels on a variety of devices for work-related and leisure technology products. Common examples include banking kiosks, cell phones, and tablets. According to Sesto, Irwin, Chen, Chourasia, and Wiegman (2012), some agencies such as ANSI have issued recommended minimum buttons sizes ranging from 9 to 22 mm, based on performance studies or anthropometric tables for the size of human fingers. Sesto et al. compared forces applied to buttons of sizes ranging from 10 to 30 mm to determine if size or spacing between the buttons affected impulse or dwell time; they also investigated how people with motor deficits would respond differently to the touch screen panels. The experimental touch screens was equipped with force detection sensors under the surface.

Impulse was the cumulative amount of pressure applied over time for the duration of the button-push. *Dwell time* was the amount of time the finger spends on the button before releasing it. It should be noted here that some contemporary button designs utilize dwell time to move the control into another control mode. Another variation is to use dwell time to adjust the amount of the control motion, such as the buttons in Figure 7.14 for adjusting the volume of a television set.

The results of the study (Sesto et al., 2012) showed that the force measures were about 6% harder or longer for the smaller size buttons. Button spacing did not show an effect for the range they considered. People with motor impairments tended to apply more force than unimpaired people. Overall the participants applied about four times as much force on the buttons than was actually necessary to activate the control. The added force was thought to be related in part to participants' experiences with a broader range of buttons. Real buttons have a stroke length that is a lot longer than what is needed for a virtual button on a touch screen. Some button controls have built-in levels of resistance that vary. Thus, the forces that might be reasonably applied to some types of buttons might be overkill for touch screen buttons.

Shape

Designers might exploit the shape of controls to induce greater tactile feedback to the operator that indicates that the operator used the correct control. Control operations that are performed in low-visibility conditions, such as underwater, would benefit from shape cues on the controllers (Carter, 1978). If both the auditory and visual systems are overloaded, shape cues on the joystick handles can make marked improvements in operator accuracy.

In another classic example, numerous pilot errors on World War II military aircraft were resolved by tactile displays. Importantly, not all types of error occurred in all types of aircraft in use at the time, which was a clue that cockpit design was contributing to the error rates. For instance, a common error in some aircraft was raising the landing gear after the plane had landed. The problem was solved by Alphonse Chapanis. Roscoe (2006) reported retrospectively:

> Chapanis realized…that the problem could be resolved by coding the shapes and modes of operation of controls. As an immediate wartime fix, a small, rubber-tired wheel was attached to the end of the wheel control and a small wedge-shaped end to the flap control on several types of airplanes; the pilots and copilots of the modified planes stopped retracting their wheels after landing. When the war was over, these mnemonically shape-coded wheel and flap controls were standardized worldwide, as were the tactually discriminable heads of the power control levers found in conventional airplanes today (p. 12).

Seminara, Gonzales, and Parsons (1977) and Norman (1988) reported a do-it-yourself retrofit of controllers in a nuclear power plant. The operators were required to reach for two important controls and wanted to avoid confusing them. They replaced the indistinguishable control handles with beer tap handles, which had different shapes and were much larger than the standard issue equipment.

Space of Controls

The total amount of action resulting from possible control movements is known as the space of the control. The system is capable of a minimum and maximum output. A good control design will allow for a range of control settings (i.e., with a continuous control) in which the full range of action is evenly distributed across the full movement of a knob or slideswitch (Kantowitz & Sorkin, 1983).

A common violation of this principle occurred years ago in the design of volume (loudness) controls on guitar amplifiers and stereo equipment. The knob would show calibrations from 1 to 10, but the effective action of the control was contained in a shorter range, for example, 1 to 4. Values 5 through 10 did not do appreciably much. In an equally common sales tactic the prospective purchaser would be told, "This is how it sounds on 4. We can't play it any louder in the store." The statement was true in a way: You could not play it any louder anywhere else either. Imagine the surprise and delight of this writer who, after becoming accustomed to this design flaw over a couple of decades, encountered a knob control that was actually properly designed.

Labels

Kantowitz and Sorkin (1983) identified four rules for the use of labels on controls, besides the obvious basic notion of labeling the controls. First, the labels should be located in a consistent position relative to the controls: always above the control, or always below or to the side of the control. Shifts in the positioning of labels can mislead the operator. It is sometimes possible to engrave the label on the control itself. Here one runs the risk of the lettering wearing off with extended use.

The second rule is to avoid abstract symbols. Clear verbal labels are better. A true tale of two motorists illustrates the problem. Motorist A had owned a German-made automobile for decades when the event occurred around 1970. His next door neighbor, Motorist B, had owned a U.S.-made automobile for as long a time. Motorist A was planning an extended trip to Europe and asked B if he would start A's car once a week or so and take it for a ride just to keep it running. B agreed, but when B eventually got behind the wheel of A's car, he found he could not interpret any of the glyphs on the controls and thus could not get it out of the garage. Eventually, A returned from the trip and rented a car to travel from the airport to his home. The car rental desk gave him a Buick convertible to drive, which he managed to start and get onto the expressway. Motorist A could not interpret most of the glyphs he saw either and ended up driving home with the top down on a cold and windy day in second gear.

The third rule is to use color coding if and only if the colors have meaning. Decorative use of color only introduces irrelevant information. One strategy for colorizing a control panel might use one color for controls that pertain to one group of functions, and use a different color for controls that pertain to a different group of functions. Another strategy might use an unremarkable black button for most of the controls, but a special green or red control for "start" and "emergency stop" functions.

The fourth rule is to be sure there is sufficient lighting around the control panel to see the labels. Otherwise the labels are pointless. On the other hand, shape cues could mitigate poor visibility, as mentioned earlier.

Resistance

Controls are often built with a small amount of resistance to the operator's motion. The resistance is sometimes a result of the physical forces that are involved. At other times resistance is deliberately introduced to protect the device from accidental operation or from damage during operation (Kantowitz & Sorkin, 1983).

There are four types of resistance. Static resistance gives the operator a bit of a counterforce in the early milliseconds of operation, but operates freely afterward. Elastic resistance feels like a spring-loaded knob; once the control has turned to a desired setting, it reverts to its initial position as soon as the operator releases it.

Viscous resistance is proportional to the speed of motion. The dashpot on a utility door allows the door to close at a preset speed once the user has set it in motion. The objective is to prevent the noise or damage that could arise from slamming the door or the operator's failure to give a manual door enough force to close properly.

Inertial resistance is proportional to the acceleration of motion. This utility prevents unwanted consequences from sudden starts. The accelerator pedals of many makes of automobile were redesigned in the mid-1970s to prevent "jackrabbit starts," which wasted fuel. In the older design, the metal bar behind the pedal that was ultimately connected to the engine was suspended from the floor. In the later designs the same metal bar was suspended from above. As a result, if the operator "put the pedal to the metal" the physical forces would translate into fewer forces to the carburetor.

Control Panels

The same rules for good organization that apply to a panel of displays also apply to a panel of controls. Control panels should be organized by their frequency of use and their order of use. Controls should be grouped by function. Controls that are used in combination should appear together. Linking analysis may be useful for determining the foregoing patterns.

A control panel for a power plant dating back to the mid-1970s appears in Figure 7.17. It is not possible to tell from the picture to what extent the currently known rules for control panel organization have been observed. The sheer morass of controls and displays, however, strongly suggests that a better organization of the panel design would greatly reduce operator error and response time. Notably, the connection between the displays and control is not readily apparent. Operators would need to retain a mental map for which displays on the vertical back connect to displays on the horizontal panel.

In newer systems, functions such as those shown in Figure 7.17 are controlled from a computer terminal. The program's screens are organized hierarchically by function. In the

FIGURE 7.17
Power plant control room, mid-1970s vintage.

typical hierarchical organization of screens, the top-level screen shows broadly defined situations and content, and the screens further down the sequence become progressively more specific in control. With a windowing interface it is possible in principle to use different combinations of functions together as desired. Figure 7.18 depicts a control station for a liquid and gas pipeline operation. The operator is apparently controlling 11 screens' worth of material simultaneously. Industry standards are now being developed to assess the cognitive load on the operators during specific phases of operation and to design systems to facilitate their decisions regarding how to mitigate an impending emergency (McCallum & Richard, 2008; U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration, 2012).

The controls in the panel shown in Figure 7.19 are primarily grouped by function. In this particular case, several buttons in a function group are used together. The buttons within a function are colored differently to enhance the segmentation of the layout. The visual displays for this digital sound synthesizer are modest. Little red status lights appear above the control button. Alphanumeric information appears in a central LED display.

FIGURE 7.18
Control room for pipeline operation, late-2000s vintage. (Reprinted from U.S. Department of Defense, *Department of Defense design criteria standard: Human engineering*, Author, Washington, DC, retrieved April 4, 2012 from http://www.assistdocs.com, 2012, in the public domain.)

FIGURE 7.19
Control panel for a sound synthesizer, late-1980s vintage. The box with the white frame indicates the location of a central LED display.

Feedback and Control

The operation of machine controls should result in feedback to the operator concerning the results of the control action. The feedback can be visual, auditory, or tactile. Visual feedback can come from three sources: a display indicating that something has happened as a result of the control action, the observation of one's finger pushing the button on the control panel, and the effect of the system as one watches it do what it is supposed to do. A common form of auditory feedback is the tonal status display that signals to the operator that a control has been received. Auditory feedback can also take the form of a change in the sound of the machine itself, for example, an engine turning on or off, or the sound of some other mechanism starting to move.

Tactile displays are more commonly encountered as feedback from control operation. Diving suits and other environmental suits insulate the operator from direct control of the environment, and there is often an advantage to building in synthetic sensory information so that the operator can react with the external environment in a manner more similar to direct contact. A mechanical form of tactile feedback delivery occurs when an operator types on a keyboard. Note that the surface of the keys is slightly beveled to accommodate the curvature of the fingertips. The keys are spaced so that an operator can feel the edges. As a result, some errors can be corrected when the operator notices that a keystroke did not *feel* correct.

Open and Closed Loops

Feedback can be positive or negative in character. Negative feedback tells the operator to reduce or stop the control action. Its purpose is to hold the status of a system to a fixed point; a thermostat is a prototypical example. Positive feedback tells the operator to increase the quantity of a control motion. If there is no negative loop kicking in, the control action and the result escalate to an indeterminate point. Other sorts of dynamics could occur under strictly positive conditions such as the destabilization of the system (Jagacinski & Flach, 2003). A diagram of an open loop system is not fundamentally different from the serial processing schema shown in Figure 6.5.

The feedback patterns, or *loops*, can be characterized as open or closed. In an open loop system, the operator exerts a control motion, and the effect is carried out without the effect of feedback changing the operator's actions in any way. In a closed system, information from visual, auditory, or tactile displays affects the operator's motions, plan of action, or hypothesis concerning the nature of the problem that needs attention. A diagram of an open loop system is not fundamentally different from the diagram of a self-organized system in Figure 2.6, except that loops can be inserted at several points to signify changes in psychomotor motions, plans, or hypotheses.

Fitts' Law

Fitts' law is one of the more basic descriptions of control response time. An operator's response is conceptualized in a physical space where the finger, hand, or tool is applied to a button, control, or target location (Fitts, 1954). It is a precise movement toward a specific target in space, where movement time (*MT*) is a function of distance to the target (*D*) times the width of the target (*W*):

$$MT = a + b \log(2D/W), \tag{7.5}$$

where a and b are constants. The constants are calculated through linear regression analysis. In practical application, researchers modify Equation 7.5 to include separate regression weights for distance and target width (Liao & Johnson, 2004):

$$MT = a + b \log(D) + c \log(W). \tag{7.6}$$

The additional regression weight accounts for some individual differences, either amount of people or among types of tasks, in how the two facets go together.

Kantowitz and Sorkin (1983) noted that a flow of information occurs between the operator and target as the finger approaches. The operator gauges the distance between the target and the current location of the finger, and adjusts the direction of the moving finger to hit the target. This simple-sounding task can be challenging for cluttered control designs and tiny buttons. In those cases, the operator is required to process greater amounts of information.

The speed of finger movement changes from the early phase of control motion to the later phase. Movement is relatively fast in the early stage, but as the finger gets closer to the target area, more information needs to be processed and the speed of movement slows down. Woodworth (1899), who defined the two velocities as *impulse amplitude* and *impulse duration*, first reported this change in movement speed. The phenomenon was later reframed as a simple nonlinear model in which the square of movement time was proportional to the distance between the finger and the target:

$$MT^2 = a + bD. \tag{7.7}$$

The nonlinear model was also known as bang-bang theory (Jagacinski & Flach, 2003; Kantowitz & Sorkin, 1983).

If one wanted to fix a movement time to a desired level and do the same for the distance between the operator and the target, the target would have to have a critical width (W_c) to meet the distance and movement time objective. According to Schmidt, Zelaznik, Hawkins, Frank, and Quinn (1979; also Jagacinski & Flach, 2003), the relationship would be

$$W_c = k\, D/MT. \tag{7.8}$$

Fitts' law and the bang-bang theory assume that a very simple movement is occurring. If the movement involves multiple submovements, Fitts' law generalizes to

$$MT = k_1 + k_2\, (D/W)^{1/n}, \tag{7.9}$$

where n is the number of submovements (Jagacinski & Flach, 2003; Meyer, Smith, Kornblum, Abrams, & Wright, 1990).

The foregoing models do not consider the weight of the operator's arm while making a movement. It is not unlikely that most of us take the weight of our body parts for granted and simply know how to move them. On the other hand, the arm might not be just an arm, but an arm plus a tool, such as a pair of pliers, welding device, chainsaw, or rocket launcher. According to Flach, Guisinger, and Robinson (1996), the amount of force

(F) required to make a control motion is proportional to a function of the distance and movement time required:

$$F = kAD/MT^2, \tag{7.10}$$

where k is once again a proportionality constant, and A is the weight of the arm.

Jagacinski and Flach (2003) likened the effect of F to the effect of water pressure in a dripping faucet. At the lowest levels of flow, the drip is periodic and irritating. With greater force, the drips are aperiodic or chaotic over time. With still greater force we no longer have a dripping faucet, but running water instead. The analogy here is that insufficient force will produce an ineffective control motion or at least a poorly executed one.

Fitts' law can also be applied to the operation of foot pedals. In the 1990s, many drivers of the Chrysler Jeep Cherokee reported cases of unintended acceleration. In other sport utility vehicles and automobiles more generally, the brake pedal is located to the right of the centerline of the steering wheel, as it is meant to be operated by the right foot. In the Cherokee, the accelerator and brake pedals were located too far to the left, with the brake pedal left of the centerline, so drivers hit the accelerator instead of the brake. Here the situation was a combination of too long a distance to the brake pedal and the population stereotype for where the brake is supposed to be located (Lidwell, Holden, & Butler, 2011).

The computation of distance across three dimensions is actually an extrapolation of the calculation of distance of objects across a two-dimensional plane. The computation for a two-dimensional orientation is the application of the Pythagorean theorem:

$$D_{a,b} = [(a_1 - a_0)^2 + (b_1 - b_0)^2]^{1/2}, \tag{7.11}$$

where a_0 and b_0 are the coordinates of the starting position and a_1 and b_1 are the coordinates of the landing position. For three dimensions,

$$D_{a,b,c} = [D_{a,b}{}^2 + (c_1 - c_0)^2]^{1/2}. \tag{7.12}$$

The transition in the computation of distance from Equations 7.11 through 7.12 can be extrapolated to four or more dimensions, but the need has not arisen in concrete applications to control systems. It has been used, however, with multivariate statistics when the goal is to detect an outlier on the totality of measurement dimensions that were involved.

Liao and Johnson (2004) assessed the accuracy of Fitts' law for 3-D arrays of objects that are portrayed on a 2-D screen rather than through a stereoptic viewing system. One of the difficulties of the 3-D application of Fitts' law is that target size and distance are confounded visually by virtue of the size–distance constancy in perception. Thus, objects that vary along the depth axis could be more difficult to reach accurately. Liao and Johnson also considered the possible advantage of placing a dropline on the display for orienting an object on the horizontal axis. Their example appears in Figure 7.20. Their results showed that Fitts' law accurately represented the movement time data for different experimental subjects, and for movements along the X, Y, and Z axes, which were assessed separately. The R^2 coefficients for degree of fit between Equation 7.6 and the data were all equal to or greater than .90. The droplines produced a small but consistent effect on reducing movement time, but did not affect the degree of fit between the equation and the data.

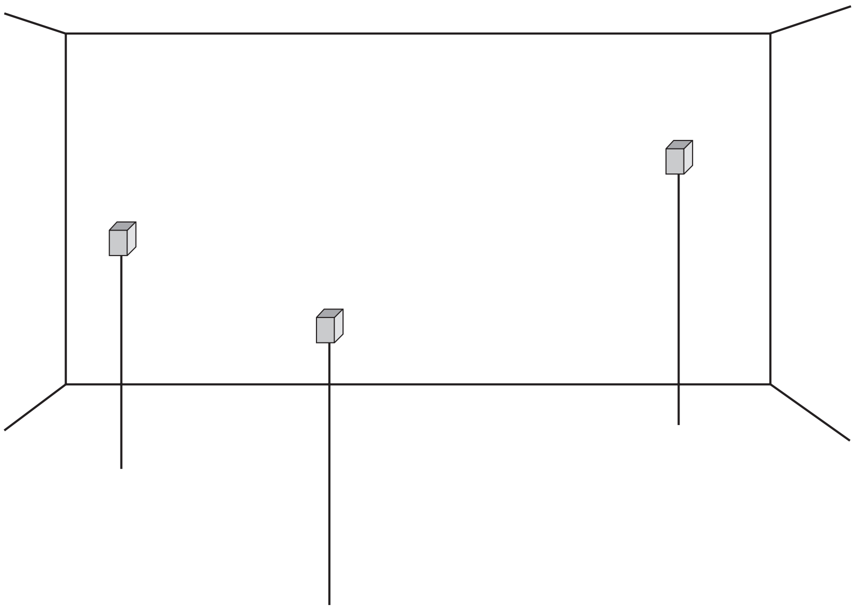

FIGURE 7.20
Droplines on 3-D display.

Motor Control

It is evident from the discussion of Fitts' law that if the analysis involves anything but the simplest movements, the mathematical description of those movements becomes complex in a hurry. Complex movements involve coordination between body movements. The timing and phasing of each movement is at least as important as the movements themselves. The acquisition of motor coordination for a given task involves a process of self-organization among the motor movements. The exact pattern of movements varies with the specific task, the range of environmental constraints under which the task is performed, and the demand for stability in the behavioral outcome that is desired (Turvey, 1990).

Motor coordination is an example of a broader class of synchronization phenomena. Synchronization phenomena require three basic components: oscillating behaviors, feedback, or information flow among the oscillators, and a mechanism by which the information flows affect the speed of the oscillator that is receiving the information (Strogatz, 2003). In the simple task of walking, our left and right legs are making reciprocal movements at precisely the correct speed; one can only imagine the result of the left leg stepping consistently faster than the right leg. In the less simple task of juggling (e.g., one person, two hands, three rubber balls), the two hands are tossing balls in the air to a consistent height and pattern; the balls are caught and tossed back into the air when the speed of gravity returns them to the vicinity of the hand.

Motor control is an underresearched topic in psychology and in human factors more specifically (Guastello, 2006; Rosenbaum, 2005), although there have been some new developments in rehabilitation science, particularly where concepts from nonlinear dynamic systems have been involved. Human factors engineers, and perhaps the information technology industries more generally, have not shown any shyness toward simplifying motor skill requirements in the interest of improving human performance speed, quantity, and accuracy. Machines often replace old skills. High-skill tasks, such as those found in the graphic arts, have been reduced often to two motor skills—point and click. The commercial interests that might be served by explaining and enhancing motor skills have been served better by minimizing the demand for those skills.

It is possible, nonetheless, that the industrial trends are now reversing. One of the latest frontiers in the computer industry is haptic perception and motor control. The perception questions revolve around the means by which touch sensations are translated into mental representations of objects. The motor control aspect attempts to describe what the person does as a result of those mental representations, bearing in mind that there is a continuous two-way flow of information between the person and the external object while the event is taking place. Virtual reality programming requires modeling information about simple-looking tasks such as how one bounces a rubber ball off a wall and catches it (Burdea & Coiffet, 2003). Motor control is, furthermore, strongly influenced by the otolithic system, which is in turn strongly influenced by gravity, which tells us which way is up. Simple physical tasks that might be taken for granted on Earth become scrambled in outer space, as the astronauts acclimate to new relationships among visual cues, motor movements, and disrupted touch sensations (Bock, Fowler, & Comfort, 2001).

Walking

Walking is more complex than it looks to the untrained eye, and the tools of nonlinear dynamics have been useful for gaining insight into the process. Most notably, there is a lot of variability within and between people as measured by gait camera systems in biomedical engineering laboratories. Gait cameras would capture horizontal hip swing, vertical hip swing, floor clearance, leg muscle contraction, and stride length. Abnormalities along any of these parameters are often detected by the camera more readily than by the person walking (Postman, Hermens, De Vries, Koopman, & Eisma, 1997).

The amount of variability in a gait sample taken from a single individual could reflect a level of abnormality, but variability in the healthy or normal case is not random. The healthy level is chaotic in much the same way as variability is found in skilled performance more generally (Meyer-Kress et al., 2009; Stergiou, Harbourne, & Cavanaugh, 2006). The loss of variability should be considered cause for alarm. Loss of variability means loss of adaptive capability, for example, to various irregularities in walking surface conditions or other environmental conditions. Thus, the loss of variability should not be equated with "stability" (Stergiou & Decker, 2011). Similarly, excessive variability that meets the analytic definition of random, however, strongly suggests that neural pathways have not stabilized or have been severely disrupted. The more common rehabilitation challenge, however, is to overcome the loss of variability.

The task of rehabilitation, therefore, is to restore the proper levels of variability. Here is where some potential for developing interesting person–machine systems can be found. The approximate entropy metric (Pincus, 1991; Pincus & Goldberger, 1994) for nonlinear time series analysis is presently the favored option for capturing *deterministic* variability and separating it from noise, given the demands of the data that are produced in a gait laboratory (Katsavelis, Mukherjee, Decker, & Stergiou, 2010a,b; Kurz, Markopoulou, & Stergiou, 2010). In principle, by increasing the level of variability, additional degrees of freedom in the person's movement have also been induced (Vaillancourt & Newell, 2000).

In one promising system, Katsavelis et al. (2010a,b) devised a virtual reality display that a person would watch while walking on a tread mill. The idea was that by varying the optic flow on the visual display, the visual–motor feedback channels would induce improved variability in gait. The technique produced the desired result in the laboratory. The long-term efficacy of the technique for rehabilitation has yet to be determined.

Reaching and Grasping

Stroke survivors can have any of a number of cognitive or psychomotor deficits. For many, the simple task of reaching and grasping a common object is a challenge. Thus, robotic therapies are current being developed to help retrain the cognitive–motor system. The essential idea of the therapy is that the patient engages in reaching exercises while attached to an electromechanical device (Figure 7.21). The robot matches the human movement as it transpires against a mathematical model of normal movement and guides the arm when it gets off-course. The theoretical challenges are currently: the development of appropriate mathematical models for the reaching movement, the fact that most prototypical person–machine systems capture the reach portion of the movement but not the grasp, and the definition of the mathematical model that coordinates the opening and closing of the hand with the reach portion of the movement. Part of the problem was resolved in principle by the development of a sensor glove that is fitted at the hand's end of the arm controller (Nathan, Johnson & McGuire, 2009).

Mathematical models for movement are developed in the laboratory by attaching sensors to the arm and hand that transmit information to a recording device that produces digitized data for the movement along X, Y, and Z axes of motion. One of the candidate mathematical models for the reach segment of the motion is the *minimum jerk*, which is a set of three equations for X, Y, and Z axes, all of which are fifth-order polynomial functions of time to complete the motion (Flash & Hogan, 1985). Another candidate is a seventh-order polynomial function of time that operates along one axis of motion (Amirabdollahian, Loureiro, & Harwin, 2002).

The latest candidate model reframed the movement as an NDS process where position at time 2 is a function of position at time 1, in contrast to earlier models that used time as an external variable (Guastello, Nathan, & Johnson, 2009). A movement cycle that involved reaching for a cup of water, putting it to one's mouth, and replacing it to the table was divided into four "events": reach, lift, pull to mouth, replace. According to their operating theory (p. 106):

FIGURE 7.21
Robotic therapy device. (Reprinted from K. J. Wisneski and M. J. Johnson, *Journal for Neuroengineering and Rehabilitation*, *7*, 2007, under the terms of the Creative Commons Attribution License.)

> The perception of the goal object triggers the movement. The visual information stream provides a continuous feedback flow that directs the hand to the object. The grasp response is coordinated to both the visual stream and arm movement to open the hand when it is in the near vicinity of the object and close it again upon reaching the object. Tactile and visual feedback at the last step signals the human or robot that the goal has been reached, and triggers the next movement…Attractors can be conceptualized in two forms. In the situated form, the object [of the grasping motion] constitutes the attractor. The perception–action sequences are organized around the capture of the object. In the embodied form, the attractor is the stabilized pattern of behavior that is, for the most part, insensitive to variations in irrelevant environmental cues (Ijspeert, Nakanishi, & Shaal, 2003; Nolfi & Floreano, 2000).

The models themselves were Lyapunov structures:

$$\begin{aligned} x_2 &= a\exp(bx_1) + c\exp(dy_1) \\ y_2 &= a\exp(by_1) + c\exp(dx_1) \\ z_2 &= a\exp(bz_1) + c\exp(dx_1) \end{aligned} \tag{7.13}$$

where x, y, z are positions in the Cartesian space coordinates, and a, b, c, and d are nonlinear regression weights (Guastello et al., 2009, p. 107). As the results turned out, motion along any one axis required input components along two spatial axes. The two exponents captured two different rates of movement simultaneously, which is consistent with the two-velocity feature of Fitts' law. The Lyapunov model provided a more accurate fit to the data than the other contenders, and with fewer parameters (Figure 7.22). The R^2 coefficients for goodness of fit ranged from .940 to .999 for the various tasks and events that were studied.

The Lyapunov structure was also useful for developing the first mathematical model for the grasp phase of movement. When the arm extends toward an object, there is a critical point located at approximately two thirds of the path toward the object when the hand

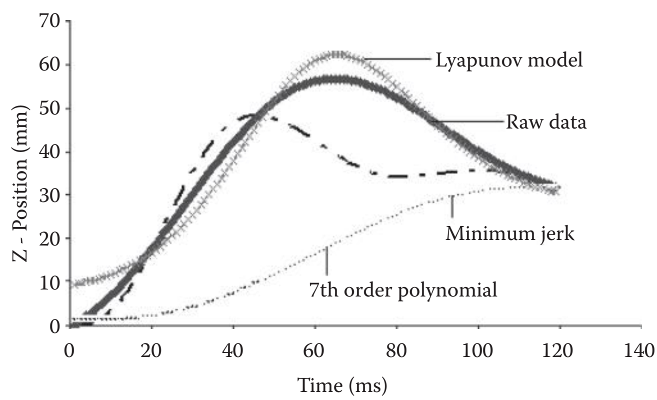

FIGURE 7.22
Relative accuracy of Lyapunov, minimum jerk, and seventh-order polynomial models for arm movement. (Reprinted from S. J. Guastello, D. E. Nathan, and M. J. Johnson, *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*, 114, 2009, with permission of the Society for Chaos Theory in Psychology & Life Sciences.)

opens up. The goal of the model is to predict the critical point in order to send a signal to the patient's nervous system if the hand is not opening properly:

$$g_2 = a \exp(bg_1) + c \exp(dx_1), \tag{7.14}$$

where g is the aperture of the hand, x is the position along the x axis, and a, b, c, and d are nonlinear regression weights (Guastello et al., 2009).

The method for signaling the nervousness has not been developed yet. The long-term effectiveness of any of the robotic therapy modalities to patient recovery is still to be determined.

Aiming

Fitts' law provides an accurate prediction of movement time if the operator's actions are also accurate at a level of 96% or greater. Wijnants, Bosman, Hasselman, Cox, and Van Orden (2009) considered the case of much lower levels of accuracy, which one could expect during the early stages of a learning process. They found that movement time had considerable variability under conditions of low accuracy, but the variability is deterministic in nature such that it can be quantified using fractal metrics. They found that the distribution of movement times fit a $1/f$ distribution, and the degree of fit improved over blocks of trials of an aiming task. Attractor strength increased with practice. Sample entropy, which is a close relative of approximate entropy, decreased over blocks of trials. The underlying dynamics that produce the self-organization of the perception and action patterns is thought to be similar to those described in conjunction with arm movements.

Order of Controls

Controls have different effects over time, and part of the learning process is to learn the behavior of the system once the command has been given to it. Figure 7.23 compares some common control functions as they transpire over time.

A *zero-order control* is comparable to the effect of adding a constant to a column of numbers. The control motion in the upper left of Figure 7.16 is probably turning a system, or a function of a system, on or off. A *first-order control* exhibits an effect that is proportional to time; hence the relationship between the control output and time is linear. First-order controls display velocity; if we drive a car at 50 kph for a half-hour, we will have traveled 25 km.

A *second-order control* exhibits acceleration. The middle panels of Figure 7.23 show two types of acceleration. The growth curve could be the result of a positive feedback effect, and could signify a system that is heading out of control in some fashion. The decay curve could be the result of a negative feedback effect that could be induced to hold the system at a constant level of output.

The accelerator pedal on an automobile certainly does accelerate the vehicle, but it is also used to maintain a constant speed. This type of control is actually an acceleration-assisted control. It has a first-order component built into it as well as the second-order component.

Third-order controls involve a change in acceleration, which we would experience as a sudden jerk motion. As mentioned previously, accelerator pedals on automobiles were redesigned years ago to dampen jackrabbit starts, or sudden jerk motions. Jerks are unfriendly to many systems, but they might be needed in others when an extra initial force is needed to start the system, but not for long afterward.

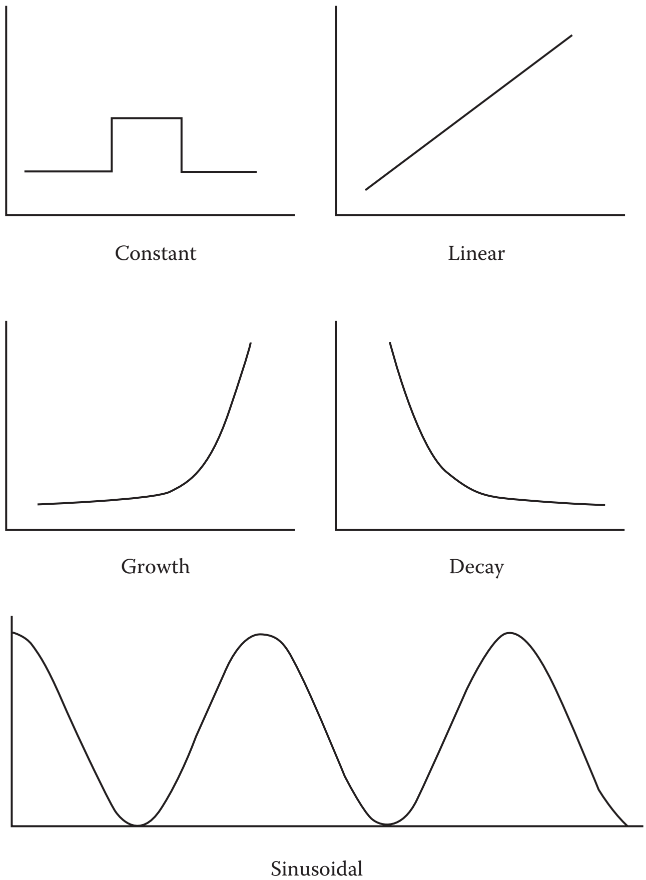

FIGURE 7.23
Effects of some basic controllers as they occur over time.

Fourth-order controls are rare. One example occurs in ships, however, in the mechanism that changes the heading of the ship while the ship is in motion. The position of the ship during the control sequence can be imagined as a 4-D function of time. One dimension is a change in heading in degrees off the bow. The second is the rate of change in the rudder position, which actually steers the ship. The ship is already moving, but if the ship is maintaining a constant speed going around in a (partial) circle, it is actually accelerating. The movement associated with this type of control is complex, and the mechanism is comparably slow. A predictive display would thus be needed to inform the operator if the ship will be headed in the correct direction once the mechanism has completed its work.

The sinusoidal function in the lower portion of Figure 7.23 has the potential for third- or fourth-order control functions. The operator may wish to adjust the signal that is displayed in the oscilloscope by its frequency or amplitude. Frequency is rate over time, or velocity. It could be made to accelerate over time if the operator wanted to change the frequency over time, for example, change the pitch of a sound. The amplitude may be held constant, or it can be modulated also at a rate of acceleration.

Chaotic Controllers

Some control functions are obviously simple, whereas others are just as obviously complex. Which is better? According to Ashby's (1956) law of requisite variety, a controller must be at least as complex as the system it is meant to control. *Complex* in this case refers to the number of system states or combinations of states that the system could produce. This important rule brings us to the topic of chaotic controllers.

It is unlikely that the operator would wish to induce a chaotic effect into a system, unless it is done as a means of controlling a system that is already unstable. More often than not, the operator would want to convert the chaotic and unstable behavior into something stable and predictable. At present there are four basic approaches to controlling chaos: anticipation and contrary motion, adding more instability, forcing periodicity, and manipulating control variables that are known to induce dynamic effects in the system (Guastello, 2002).

Anticipation

There are three groups of studies that address some aspect of human ability to track complex dynamic processes. In one series of experiments (Ward & West, 1998), human participants were presented with numeric series that were generated from Equation 2.4. The participants' task was to predict the next value in the series. Correlations between actual and predicted values ranged from .55 to .85. Decision makers were found to have engaged in two heuristics that assisted their forecasts. They ignored the tiny differences in numerical values that were associated with extreme decimal places; thus the gross swings in actual values were more important to their forecasts than the fine details. The second heuristic was based on the forecasts of sequential pairs of actual values. Thus, the intuition of a sequential point in the series depends on the direction and size of the previous movement in the actual values.

In a similar type of study, Heath (2002) determined that participants who were required to forecast an event from a series of chaotic data did a better job than a control group that was asked to make predictions from a randomized data set. The studies showed that people can intuit a deterministic process that underlies an apparently random sequence of events, although they vary markedly in their success rates. Other studies involving the management of a complex system illustrated similar variability in this skill (Guastello, 2002).

Another use of the anticipation principle is to include chaotic controllers in the machine's functionality, leaving a simpler task for the operator. For instance, one of the leading causes of death in agriculture is tractors tipping over on their drivers. The problem is related to both the slope of the hill and irregularity of ground; the latter has a fractal or chaotic character to it. Roll bars can be factory installed on the tractors or as retrofits. Roll bars address the major problem of the slope, but not the instabilities coming from the land under the wheel. Thus, mechanisms in the wheels have been developed that can interpret the incoming pattern in the land and induce a correction to the wheel–axle assembly (Sakei, 2001). The relative contribution of instability from the land is much greater in light excavating equipment. No data are currently available on how widespread the use of chaotic controllers in excavation equipment happens to be.

Adding Instability

The ideal of controlling chaos with chaos was introduced in the early 1990s (Breeden, Dinkelacker, & Hubler, 1990; Jackson, 1991a,b), but the idea has not been developed very far for person–machine systems at the present time. This type of chaotic controller works counterintuitively by first adding a small amount of low-dimensional noise into the system. The reasoning is that the amount of sensitivity to initial conditions is not uniform throughout the attractor space. Sensitivity is greatest in the outer edge of the basin of the attractor and least in the center, compared with the middle regions of the attractor space. The level of sensitive dependence may be affected by the attractor's proximity to another attractor. Adding noise to the system allows the attractor to expand to its fullest range. It is essentially a variation of stochastic resonance, except that what was noise is now something that resembles noise but is a deterministic process instead. The controller can then respond in one of two ways. One option is to apply a filter that allows a special range of values to pass through. The second option is to mimic the chaotic motion of a point, predict where it is going, and respond in a strategic manner.

One application that was devised involved the development of a special shoe for people who had difficulty walking due to motor impairment. As mentioned previously, their gait typically suffers from insufficient variability. The prototype shoe would apply vibration to the foot with the intent of encouraging greater variability and adaptation of movements by the wearer (Priplata, 2006). The efficacy of the design is still under study (Collins, 2008).

Periodic Entrainment

The method of controlling chaos through periodic entrainment (Ott, Grebogi, & Yorke, 1990; Ott et al., 1994), or OGY, is based on the relationship between oscillators and chaos: Three coupled oscillations are minimally sufficient to produce chaotic time series. Many economic and biological events exhibit periodic behavior, or are the result of coupled periodic oscillators. Although not all of the possible combinations of oscillations do produce chaos, it is still a viable analytic strategy, nonetheless, to decompose chaos into its contributing oscillating functions.

One common application of periodic entrainment is batching tasks. If an operator receives a task of a particular type on a random or chaotic basis, and probably mixed up with tasks of different types, the operator would let tasks of a given type pile up to a critical mass before going to work on them. Then once the batch is completed, move on to a batch of a different type of task. Another variation is to wait until a specified time in the work period to do whatever is in a batch. Batching has the effect of minimizing task switching and all the loss of momentum that goes with it.

The OGY method has some limitations in situations where its computations are to be taken literally to model behaviors of real, rather than simulated, systems. The principle problem is its susceptibility to disruption from noise. An external shock will be carried through subsequent iterations of a chaotic system. The disruptions become magnified and the tracking of the stable manifold becomes impaired. For this reason, the literal application of the OGY method may give way to other methods in practical settings.

Use of Control Parameters

Equation 2.4 received a great deal of attention in complexity studies because its one control parameter governs changes in system behavior from fixed to periodic, aperiodic, and chaotic states. Although there is no guarantee that a real chaotic system can be controlled through all its dynamic states by just one parameter, it is intriguing, nonetheless, that a wide range of behavior patterns could be controlled by a very few well-chosen parameters.

Catastrophes, which comprise another group of dynamic processes, are often controlled by two parameters.

Voice Control

Voice-controlled systems are often sought for disabled persons (Noyes & Frankish, 1992) and for situations where the operator's hands and eyes are overtaxed such as piloting a military aircraft (Corker, Cramer, Henry, & Smoot, 1992; Reising & Curry, 1987) or driving an automobile with a navigation system or cell phone (Zaidel & Noy, 1997). Voice controllers also have some potential for distinguishing users' voices when the voice is intended as a locking or security system. Although advances in signal processing have made it possible for systems to recognize single words with sufficient accuracy, there is substantial within-person variation in vocal production, perhaps resulting from stress, mood, or ambient distraction that could compromise the effectiveness of voice-controlled locking systems.

The typical system uses the word as the unit of analysis rather than the phoneme. Systems tend to discriminate best among words when the system vocabulary is small, and when the speaker is given an instruction to something specific, such as, "If your system is a dial-up system, say 'dial-up.' " In this telephone application, the voice control is substituting for pushing a button on the touchtone phone (Baber & Noyes, 2001). The limited vocabulary compensates for the wide range of voice registers that the system is anticipating. Fortunately, voice-control users automatically make their speech commands more concise than they might do when talking to another human (Amalberti, Carbonell, & Falzon, 1993).

Noise on the channel and noise in the operator's environment induce two sources of error that limit voice controllers to office, home, or noise-controlled environments. If the system requires individual users to enter template commands in their own voice, the templates should be entered while the user is in the environment in which the voice-controlled device is going to be used (Baber & Noyes, 2001).

Studies have started to appear that compare voice-controlled systems against other control modalities. In one application involving the disclosure of potentially sensitive medical information in a survey, respondents disclosed more information using an interactive voice response system than they did using a web-based survey (Evans & Kortum, 2010). Here one could say, however, that the user was controlled *by* the system rather than controlling the system. In a different type of system in which the users were accessing weather information in an air traffic control environment, voice control compared unfavorably against wand (stylus) and touch-screen versions of the system with regard to exploration, gathering, and recall of information (Dang, Tavanti, Rankin, & Cooper, 2009); the users found the slow rapport time for the voice controlled system frustrating. In a military environment, there are advantages to having multiple control modes available as some communication objectives are more compatible with some control systems than others (Walker, Stanton, Jenkins, Salmon, & Rafferty, 2010); digital voice input seems to work well enough for transmitting *data*, but information that is supposed to convey *meaning* is communicated better through conventional person-to-person analog systems such as radios and telephones.

DISCUSSION QUESTIONS

1. Take another look at the control–display combination shown in the upper left of Figure 4.3. The positions on the pointers in the display are set with the use of the knobs appearing below it. How would you assess the compatibility of the control with the display?
2. Consider some tasks where error rates increase dramatically with work speed. What are the consequences of the errors? What characteristics of the situations promote large or small opportunities for error?
3. What types of psychomotor skill are involved in driving an automobile?
4. What are some good methods for preventing controls from accidental operation? A few methods have been mentioned in this chapter. Can you think of some others?
5. Several types of discrete controls were introduced in this chapter. Describe the assets and limitations of each. How would you determine which is the best one to use? (Hint: Pick the right tool for the right job.)
6. The notion of redundancy improving reliability was introduced in Chapter 2. How many applications of this concept could pertain to the design of control systems?
7. Some researchers have considered the possibility that people could become fatigued by pushing too many buttons in the course of a day. Is this really a meaningful concern? Why?
8. How do astronauts determine which direction is "up"?
9. One science fiction fantasy that could become a reality soon is the "smart house." Imagine a house or apartment that is voice-controlled so that lights and security systems would turn on and off, appliances would operate, and heat and air conditioning would adjust if you spoke the magic words. Would such a system really be effective?

8

Anthropometry and Workspace Design

Anthropometry is the study of all measurements of the human body and the uses of this information. It naturally progresses to workspace design and some related issues in biomechanics. This chapter extends the basic ideas of human measurement to physical abilities such as strength and balance.

Body Measurements

The study of bodily measurements began with some of the earliest uses of descriptive statistics by Adolph Quetelet (1976–1874; Howell, 1992). Importantly, Quetelet discovered that the measurements were normally distributed. Although the prevalence of the normal distribution in biological and psychological measurement can easily be taken for granted, the finding gives statisticians a great deal of direction for handling and interpreting large populations of measurements and samples taken from theoretical populations.

Figure 8.1 shows several aspects of bodily measurements that are used for common workspace design problems: operator standing, operator's reach envelope on a flat workbench, and operator's seated area. The arrows mark commonly needed measurements. For the most part, workspace engineers can rely on tables compiled in known documents such as U.S. Army (1979) and Roebuck (1995) for U.S. military and civilian populations, Pheasant (1996) and Pheasant and Haslegrave (2005) for British populations of varying ages, and PeopleSize (Open Ergonomics, 2000) for a compendium of international samples. In addition to the linear measurements denoted in Figure 8.1, circumference measurements such as the wrist and forearm, are also included in some tabular sources.

Some occupational groups could differ from the general population as a result of the people who are attracted to the job and how they are selected. Electric utility workers, for example, tend to run larger than the average population (Marklin, Saginus, Seeley, & Freier, 2010). The U.S. military has its own norms tables that need to be utilized when supplying any equipment to the military (Department of Defense, 1991). A web search of *DOD-HDBK-743* would turn up anthropometric standards for the military of some other countries.

Note that most tables provide measurements for unclothed people. People will be wearing clothes when the system is put into use, however. Pheasant (1996) recommended adding 25 mm to each linear dimension for males to compensate for shoes, and 50 mm for females. Other adjustments should be made for the thickness of clothing, as the situation requires. Swimwear and winter clothing will have different effects obviously.

If the measurement goals can be accomplished from a single measurement from a single population source, the engineering only needs some simple statistics: the mean, the standard deviation, and a chart of percentiles associated with each measurement. Systems are typically engineered for the central 90% of the distribution, that is, for the range between

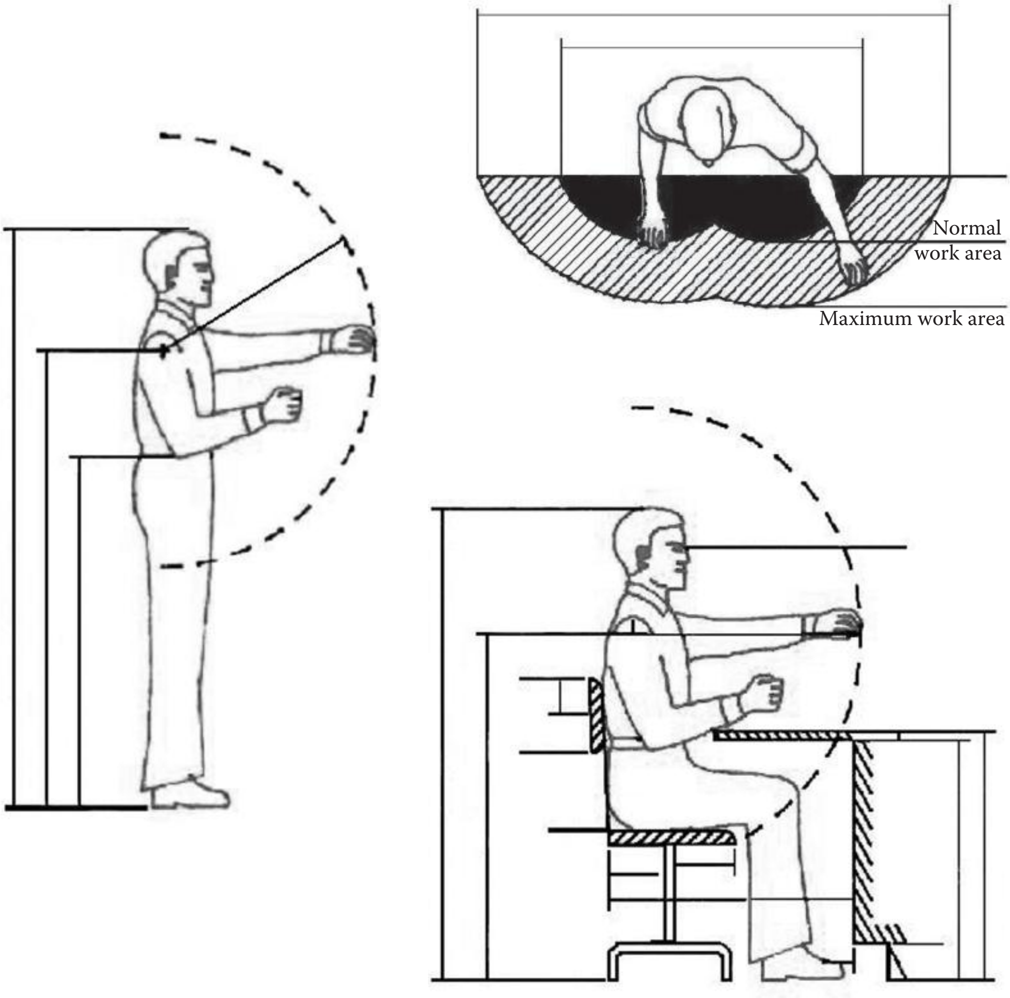

FIGURE 8.1
Three common postures for anthropometric measurement.

the 5th and 95th percentiles. If the goal of the design is to make sure that operators have at least enough space, the design process ends here. If too much space is a potential problem, such as objects being placed out of a person's reach or a driver not being able to reach the pedals of a vehicle, some flexibility for adjustments in seating or object positioning need to be built into the design.

If a composite of populations is involved, the engineer needs to introduce some statistical sampling techniques that draw from the measurements of specific populations. For instance, in the early days of the space program (late 1950s), astronauts' capsules were designed for a highly selected population of 45 men who were candidates for space flight (Roebuck, 1995). Space flight crews of the late 1990s and beyond include women, civilians, and people of other national origins. The range of body movements has also changed from constricted seating to spaces that support work that involves movement around the spacecraft and exercise.

If the goal of anthropometric design involves multiple measurements on a human, the designers should take another step to ensure a proper workspace. On the other hand, sizes of body parts are correlated. If one part is larger, the other measurements are larger in proportion. On the other hand, the correlations are imperfect and sometimes low. Thus, the correlations among the measurements are important information. If the set of required measurements were completely unrelated, one could determine the bounds of a workspace by simply looking at the 5th and 95th percentiles of each desired measurement. If the relevant population contains males and females, which it almost always does in recent decades, the design range is for the 5th percentile female and the 95% male, which is wider (Marklin et al., 2010). Inasmuch as statistical independence is seldom the case, reliance on simple percentile data could produce a design that is too large or too small for the real central 90% of the population. One way to get around the problem of correlated measurements is to collect one's own data from a representative sample of the user population; local population data would capture with all the ethnic variability represented in the sample.

Bodies in Motion

The discussion of body dimensions thus far has been predicated on the assumption that little physical movement is actually involved. In the real world, however, tools and work objects must be manipulated in many ways. Figure 8.2 depicts a system for determining a work envelope in 3-D space (Roebuck, Kroemer, & Thomson, 1975). The worker is positioned in a space that is fitted with movable poles in all directions. The worker then acts out the likely work tasks, moving the poles sufficiently out of the way to accomplish the actions. The amount of protrusion of each pole into the workspace can be measured. The data for each pole can be compiled and tabled as for other anthropometric measures.

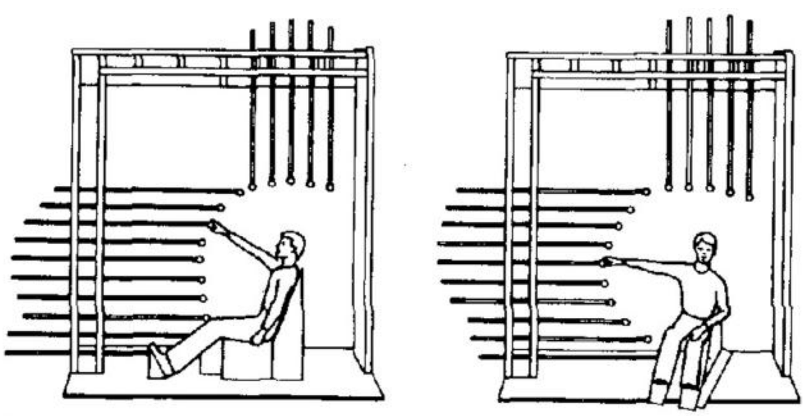

FIGURE 8.2
Equipment for determining a 3-D work envelope. (J. A. Roebuck Jr., K. Kroemer, and W. Thomson: *Engineering Anthropometry Methods*, p. 72, 1975. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

Iterative Design

Once the space requirements have been determined, the next phase of the design is to build the prototype workspace. The iterative design concept that was mentioned in earlier chapters for controls and displays is applicable here as well. The engineer would build a mock-up of the workspace with inexpensive materials. Any controls and displays that would be part of the workstation do not need to be functional at this point. Small, medium, and large people should stand or sit in the mock-up and go through the motions of work activities. Designers should then determine whether activities could be executed comfortably, whether the users can reach the controls or tools, and so forth. If the workspace is an environmental pod such as the driving compartment of a motor vehicle, the designers should determine whether users can enter and exit quickly and comfortably and whether there is enough head clearance.

Various components of the system tend to interact. Once the suitable space for the operator has been determined, the next questions are the following: Can control and display panels be fit in comfortably? Does the operator need to reach over one control in order to grab another? Is accidental activation of one of the controls likely? Can the displays be made large enough relative to the viewing distance?

If the system were expected to move through space and time, it would be helpful to use mannequins instead of people for testing some aspects of the system's integrity. Hence, a few crash test dummies would be worthy investments. Mannequins can be built with internal sensors that detect impact on various parts of the body. Small annoyances such as rubbing or chafing can be detected with simulations that assess the results of extended use of the system as well as traumatic impact.

As each layer of questions is answered, the next layer needs to be addressed and tested before the design can be considered finished. Although human factors research and development may seem laborious and expensive in some respects, it is a far better investment to place the resources into a proper design than in manufacturing a product that needs to be redesigned to compensate for user dissatisfaction.

Safety and Related Concerns

The previous section of the chapter considered the engineering of a single workspace. In this section we consider the juxtaposition of one workspace with another, or with hazardous objects positioned nearby. The soldier in Figure 8.3 has a workspace problem. Can you figure out what it is?

Machine Guards

Blades, other sharp edges, protruding structures, and pinchpoints on machines are all examples where a person–machine contact should be avoided. Figure 8.4 illustrates two hazardous configurations where two machine rollers have turned inward (based on Smith, 1977). If a hand or piece of clothing becomes caught, it could drag the operator into the machine, and the odds of a significant injury would be very high. Another form of pinchpoint occurs where a drive chain connects to a sprocket; an example appears on almost any bicycle. Machines with pinchpoints should be equipped with guards to prevent the

FIGURE 8.3
A workspace design problem. This picture circulated the Internet without any identifiable source.

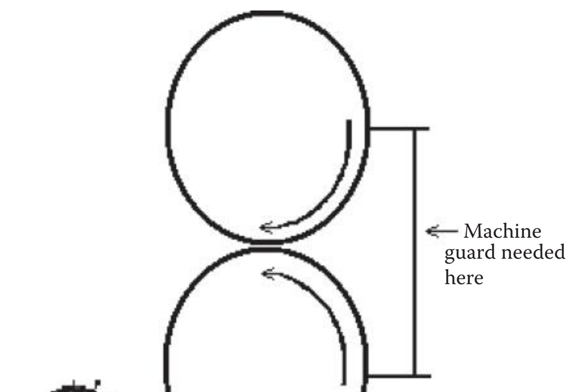

FIGURE 8.4
Machine configurations illustrating pinchpoints.

operator from inadvertently entering the mechanism. For equipment that contains a number of pinchpoints, a good solution might be to construct a guard that surrounds a large portion of the mechanism (Haigh, 1977).

Figure 8.4 also illustrates what would happen if the operator should accidentally slide under the machine or try to do so while performing maintenance on it. Safety, comfort, and practical amounts of workspace are just as important for maintenance processes as they are in regular machine operation.

Emergency shutoff switches are as important as machine guards to workspace design. Emergency switches should be reachable from any operator position relative to the machine. In the case of maintenance functions, it is standard practice to turn off motors before servicing the machine. There must also be sufficient notice given to operators of machines in the immediate area that someone is performing maintenance and that the system should not be turned on. In some factories it has been possible to install lockout devices on a cluster of nearby machines to prevent any one of them from operating during a maintenance task.

Overcrowding

The previous sections of this chapter outlined the necessary steps for giving the works just enough space. Repeat the steps too many times, however, and something dismal happens, as shown in Figure 8.5. Crowding is stressful. This form of stress is considered in the next chapter.

Confined Spaces

"Confined spaces" is a formal term given by NIOSH to denote a type of space that workers are only expected to enter on a very occasional basis. Their options for entry and exit are very limited, and they are usually bedeviled by insecure standing platforms, structural instability, insufficient oxygen and highly toxic air pollutants. Examples include grain silos, trenches, water and chemical tanks, wells, manure pits, pipelines, and sewers.

FIGURE 8.5
Tight workspaces contribute to crowded and stressful conditions. This picture circulated the Internet without any identifiable source.

The Centers for Disease Control and Prevention (2009) reports that confined spaces are responsible for about 92 fatalities per year in the United States.

NIOSH (2011) has a system of guidelines for working in different types of confined spaces. Some of the more generalizable ones, which appear to have been overlooked in fatal situations include: (a) have a written plan for entering, exiting, and working in the confined spaces; (b) equip the workspace with lifelines and harnesses; (c) test air quality before entering; (d) ventilate the space from contaminants; (g) supply oxygen masks (self-rescuers) on site; (h) have an emergency rescue plan developed that should be communicated to local fire and emergency medical teams.

Physical Abilities

This section describes some other human measurements that are especially relevant to physically demanding work. Fleishman (1964) identified nine physical abilities factors. The objective of the study was to establish physical fitness norms for military personnel and to define training requirements for them. Some of the concepts transfer well to civilian industrial settings.

Strength

Static or *isotonic strength* is the amount of force in pounds per square inch (or metric equivalent) that a person can exert on a target. The measurement does not actually involve moving the target. Static strength is measured by a dynamometer; some simple examples for measuring arm, back, and leg strength appear in Figure 8.6 (from Chaffin, Herrin, Keyserling, & Foulke, 1977). For the measurement of arm strength, the person stands upright, holding a metal bar. The bar is attached by a chain to a load cell that is mounted on the floor. The posture and chain length allow only arm strength to register on the load cell when the person pulls upward. The load cell is attached to a digital meter that displays the strength measurement. The body position and chain length for back strength and leg strength are also shown in the diagram.

Dynamometers are also available for other strength sources, such as vertical pull and handgrips. The importance of handgrip strength is well known to anyone who has ever tried to open a jar of pickles. From an engineering point of view, norms for handgrip strength from a relevant population are also important when designing mechanisms such as motorcycle clutches. Popular models can vary by a factor of 2 in their requirements for user strength. These differences in strength requirements are large enough to produce a schism in the markets for motorcycles, such that different subpopulations gravitate toward different models.

Dynamic or *isometric strength* describes the ability to move objects. It is calculated in terms of weights and distances and is the human equivalent of horsepower as it is applied to engines and motors.

Contemporary norms in factories restrict the weight of objects moved by humans to 50 lb (or 25 kg) without the use of mechanical lifting devices. This is a guideline rather than a rule, and it is an impossible objective to maintain in some circumstances, particularly in cases where the physical environment is not conducive to positioning the lifting devices. Roofers in the construction industry are often able to load packages of roofing materials

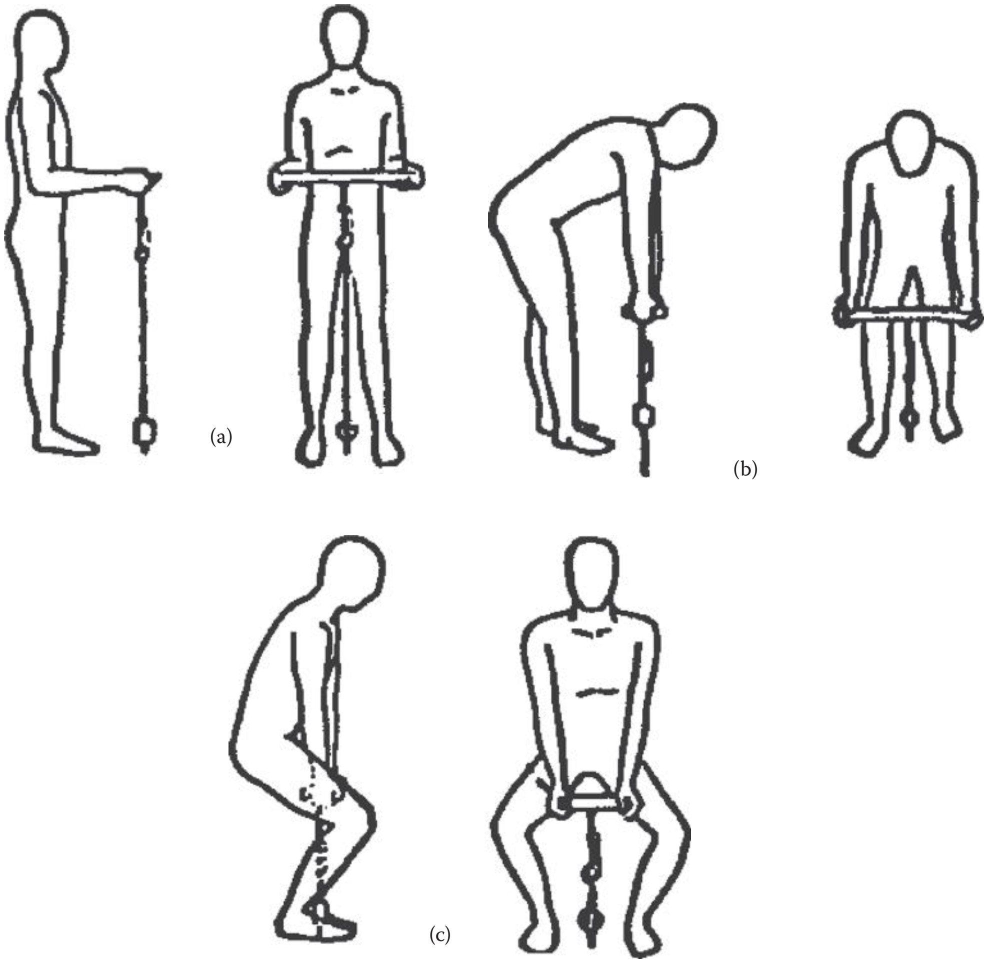

FIGURE 8.6
Dynamometer positions for (a) arm, (b) back, and (c) leg strength measurement. (From D. B. Chaffin, G. D. Herrin, W. M. Keyserling, and J. A. Foulke, *Pre-Employment Strength Testing in Selecting Workers for Materials Handling Jobs*, Report No. CDC 99-74-62, National Institute for Occupational Safety and Health, Physiology and Ergonomics Branch, Cincinnati, OH, 1977, in the public domain.)

to desired locations on a roof with a crane, but the 100-lb (50-kg) packages must be moved short distances manually often enough. Firemen are typically trained or screened for their ability to carry a 150-lb (75-kg) body out of a burning building. People seated in the rows of airplanes adjacent to the emergency exits are informed that the door that they must remove weighs 65 lb (30 kg).

Explosive strength is the ability to deliver exceptional force over a brief period. It is analogous to sprinting versus running, but often with a sledgehammer or other tool. In this writer's experience in the steel industry, emergency conditions arose occasionally where the operators had to break off a machine fitting with a sledgehammer before an explosion occurred. Two people working at explosive proportions for 2 to 5 min could get the job done, but they reported being exhausted for the next 20 min.

Trunk strength is the limited ability associated with the stomach muscle. The best indicator of this ability is the number of situps a person can perform.

Flexibility

Extent flexibility is the ability to extend one's limbs as far as they can reach. *Dynamic flexibility* is a similar ability associated with repeated flexing.

Body Coordination and Equilibrium

Gross body coordination involves tasks requiring two hands or hands and feet. Some example tasks would involve packaging and manipulating boxes. Drummers in musical ensembles typically require extensive coordination of hands and feet, and some have become famous for their exceptional ability.

The ability to maintain *gross body equilibrium*, or *balance*, is typically measured in the rail walk test, the apparatus for which is shown in Figure 8.7. A 2″ × 4″ × 8′ piece of wood (standard construction material) is firmly lodged to the floor with the narrow side up. The person being tested walks the plank three times forward, three times backward, and three times sideways. The score is the number of times the person walks the length of the whole beam without stepping off. The rail walk test tends to produce higher scores for men than for women on the average because the body center of gravity is lower to the ground for women than for men.

Stamina

The last physical ability in Fleishman's (1964) taxonomy is stamina. Stamina is the ability to withstand physical effort over a long period. It is the opposite of fatigue, which is considered as a separate issue in the next chapter.

FIGURE 8.7
Apparatus for a rail walk test.

Lean Body Mass

Lean body mass did not appear as a physical ability in Fleishman's (1964) taxonomy, although it has relevance to physical fitness and work capacity. Pound for pound, muscle is more useful than fat. Given equal strength, people who are higher in body fat are more likely to fatigue to greater extent (Guastello & McGee, 1987). The dynamics of fatigue are expanded further in Chapter 9.

The original method for measuring percentage of body fat was to weigh the individuals then submerge them in a task of water, record the amount of water displacement, and then calculate the person's density. (And if she doesn't drown, she's a witch!) Fat is less dense than lean body mass. The obvious inconvenience of this method led to the development of a method using skin fold calipers. The person taking the measurements pinches a fold of flesh at critical body locations and uses the calipers to measure the thickness of the fold. The critical points on the male body are just above the knee and on the back just below the arm socket. The critical points on the female body are folds on a triceps and the midriff. The conversion table that comes with the device translates the skin fold measurements into percent body fat.

Physical Abilities Simulation

Several of the large steel manufacturers conducted studies to determine which job applicants would be best suited to the physically demanding work. It was not possible to determine just by looking at individuals whether they would be capable of doing the work. Medical examinations can determine if a person had a prior back injury, thus making the person unsuited for certain types of assignments, but they cannot determine suitable strength without the use of dynamometers for persons who have never been injured.

The physical abilities studies in occupational settings are formatted like standard personnel-selection studies: The independent variables consisted of measurements that could be taken during a preemployment interview, and the dependent measures were work performance measurements taken under controlled conditions. Campion (1983) reviewed these studies from industrial settings as well as for police and fire personnel. The following is a synopsis of a later study reported by Guastello and McGee (1987).

At the time the study was conducted, several of the major steel producers were working under a legal judgment from a suit brought by the Equal Employment Opportunity Commission requiring the employers to increase the proportion of women in traditionally male jobs to 20%. They needed a fair system for selecting applicants.

The independent variables were height, weight, percentage of body fat, the rail walk test, and dynamometer measurements of arm, back, and leg strength. Other indicators were the number of hours of physical exercise the individual engaged in during the course of a week (self-reported) and whether the individual was currently a member of the labor pool.

The sample consisted of 129 people, both male and female. The full compliment of two production mills were included plus others who volunteered from the management staff. The objective of the sample composition was to produce the widest range of measurements of both the independent and dependent measures. Wide score ranges will allow for maximum-sized correlations between independent and dependent measures and allow for the generalization of the final results to the widest range of job applicants.

The dependent measure was a composite of measurements taken during a structured 2-hr period of simulated work that started immediately after the dynamometer readings were taken. The first task was to shovel enough furnace slag into a wheelbarrow at level height, then empty the wheelbarrow; the measurement was the time required to perform the task. The shoveling station is shown in the photograph in Figure 8.8.

The second task was to push a wheelbarrow through an obstacle course three times with loads of sandbags weighing 100, 200, and 300 lb (Figure 8.9). The obstacles were designed as segments of railroad track that naturally occurred in the real work environment. The dependent measure was time elapsed. The other tasks were sledgehammering, lifting, and carrying a drum of material used in the plant that weigh a little over 100 lb, pushing and pulling a heavy bar of steel across a greased track, hauling sandbags, and dumping gravel.

The results of the study indicated that the composite of measures of work performance could be predicted by six variables (multiple $R = .87$). The two most effective variables were arm and leg strength. Arm strength was not surprising given the amount of upper body strength generally required for this type of work. Leg strength often acts to compensate for arm and back strength. Back strength was also correlated with work performance, but it accounted for no unique variance beyond that accounted for by arm and leg strength; thus, it did not enter the multiple regression model. This was a fortunate result because it was desirable to avoid the risk associated with measuring back strength among job applicants.

The remaining four predictor variables that showed unique associations with work performance were height, balance, age, and whether the person was presently a member of the labor pool. Height was important for two reasons: (a) Taller people have more skeleton on which more muscle is attached, and (b) the added height gave people greater leverage over some of the objects that they had to manipulate.

Older participants performed more work than younger ones; this was an unexpected finding. In normal circumstances, muscle strength declines about 5% starting at age 45. The explanation in this situation was that the older workers, who were either currently in the labor pool or who once were part of the labor pool but had moved on to other jobs, used their body motions and tools more efficiently than the younger workers who had been employed in the industry for less than 6 months.

FIGURE 8.8
Shoveling station that was used in the physical abilities study reported by S. J. Guastello and D. W. McGee (*Journal of Mathematical Psychology, 31,* 248–269, 1987).

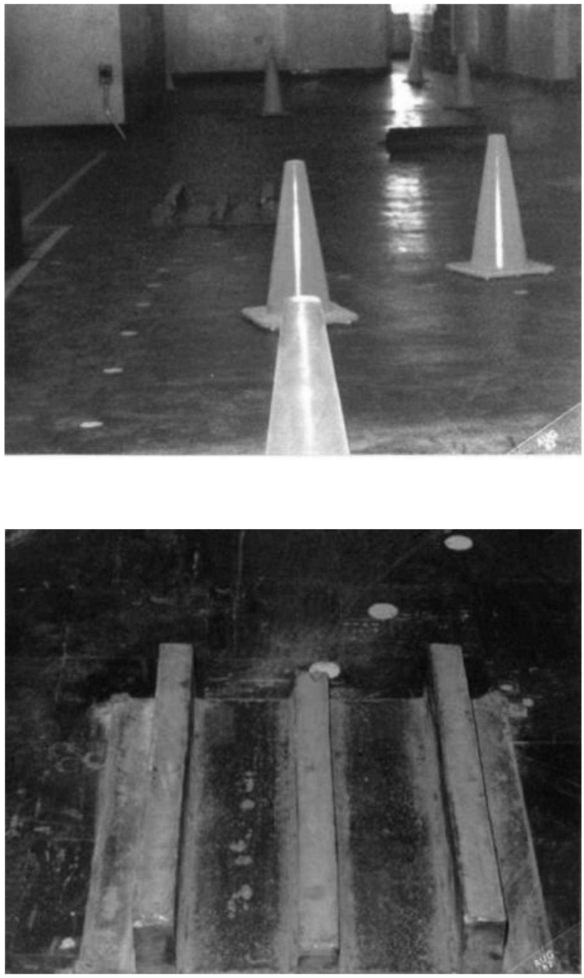

FIGURE 8.9
Wheelbarrow obstacle course used in the physical abilities study reported by S. J. Guastello and D. W. McGee (*Journal of Mathematical Psychology, 31,* 248–269, 1987).

Body fat did not have a direct impact on the work performance measure. It was related to greater fatigue, which was the difference in arm strength between the time of the initial measurement and when the measurements were taken again after the 2-hr work sequence.

Some Common Biomechanical Issues

This section considers some topics in biomechanics that are frequently encountered during workspace design. They are the mechanics of lifting, design of seating, handtools, and work conditions that are often associated with carpal tunnel syndrome.

The biomechanical issues might be redefined as sources of musculoskeletal disorders. NIOSH (1997) identified potential sources of musculoskeletal disorders in several industries. In meatpacking, for instance, the potential is associated with cleaning metal tubs, shank trimming, and removing lard and internal organs. In warehousing, the potential risks are associated with lifting and carrying containers of assorted weights. The point here is that each industry has its own array of potentially hazardous situations.

Lifting

Despite all previous remarks about the importance of back strength, there is a widespread advisement in labor environments to lift with one's legs, not with one's back. As it turns out, the weight per cubic centimeter on the vertebrae when the person lifting the object stands upright from a bent-over position is twice as great as the load on the back when the person stands upright with the load while unbending the knees.

Figure 8.10 illustrates an instance of work design that lessened the load on the worker's back. In the "before" picture, the operator lifted and moved heavy containers from an initial location on the floor. In the "after" picture, the containers were set on a table surface further from the floor, thus lessening the vertical lift. Work that involves turning with the heavy object also produce an elevated risk of spinal injury (Kumar, 2002).

NIOSH (1994) developed an equation for computing recommended weight limits. The equation requires several measurements that are multiplied together to produce the result: the weight of the object, horizontal distance of the object's initial position to the operator's body, vertical distance (which is less if the object is positioned about 1 m off the floor), how high the object needs to be moved, turning angle, frequency with which the lift occurs, and the grip strength required to grasp the object. A companion document (NIOSH, 2007) provides numerous and probably familiar illustrations of best lifting practices for minimizing musculoskeletal injuries.

Liberty Mutual (2005) developed a similar calculation routine where the calculation result is framed around percentile norms for lifting by males and females under a variety of lifting conditions. Their website provides a set of interactive pages for entering measurements similar to the NIOSH measurements. The result that is returned is the percentage of the population who can perform the task. They recommend structuring the tasks to match the capabilities of at least 75% of the occupational population. The website also has facilities for calculating the efficacy of pushing and pulling tasks.

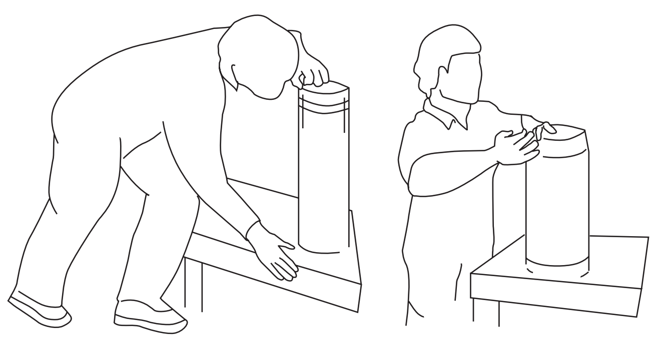

FIGURE 8.10
Workstation requiring repeated heavy lifting before and after reengineering.

The repetitiveness of a task could place a profound limit on the maximum acceptable effort associated with a lifting task. All other things being equal, the maximum acceptable effort value drops to 50% of peak effort if the operator is under load for 10% of the work cycle (Potvin, 2012). If the operator is under load for 30% of the work cycle, the maximum acceptable effort drops a bit to 45%, but fatigue sets in after 1 hr.

Walking Surfaces

Pushing and pulling tasks often involve carts, so in addition to the strength requirement for setting the cart in motion there are two other sources of demand on the operator. One is the controllability of the cart, such as steering and control of tipping, which are inherent in the design of the cart and its wheels. There are two types of forces involved, *initial force* and *sustained force*. A cart designed for high inertia will require less sustained force but might be more difficult to stop in an emergency. The image of a "freight train running downhill" comes to mind as an extreme example of this principle.

Optimal flooring for carts involves a trade-off between two physiological demands. The concern is not only strength requirements, but disproportional strain on leg and back muscles and the respiratory system that could be induced by different configurations of materials. Maikala, Cirello, Dempsey, and O'Brien (2009) compared the demands for pushing an equipment cart across a plywood floor and a Teflon floor; the coefficient of friction for the plywood was 2.6 times that of the Teflon. Muscle demand was 26% higher for plywood, but oxygen demand for the initial impulse was 19% greater for the Teflon.

Another type of trade-off to consider when choosing factory flooring is the extent of ordinary walking that one should expect on the floor, and whether the pedestrians would be carrying a load of any size. Slippery surfaces encourage falls.

Walking on overly rough surfaces can have a substantial impact on gait. Railroad workers often have to walk on ballast, which are three sizes of rocks used for drainage in and around the track area. The largest rocks create the roughest surface. Wade, Redfern, Andres, and Breloff (2010) found that ballast sizes affect just about every aspect of gait measurement (see previous chapter). The overall recommendation is to use the smallest ballast on walking surfaces or on surfaces where the operator stands while working to minimize the demands on the ankle and the knee as they coact to maintain bodily stability and to minimize muscle fatigue.

Seating

Kantowitz and Sorkin (1983) noted that there would be no particular seat that would suit everyone, but a few truisms about seating have emerged nonetheless: (a) The majority of the body weight should by carried by the bones in the buttocks rather than by other body parts, (b) the pressure of the thighs against the seat should be minimal, (c) feet should be placed squarely on the floor or a footrest should be used, (d) it should be possible to change posture easily in the seat, and (e) the lower back area should be supported (p. 478). Proper support of the lower back alleviates the weight on the vertebrate by a substantial amount; the precise amount depends on the person-seat fit (Pheasant, 1996) and thus lowers the potential for back pain or pinched nerves after many hours of sitting. The ability to change posture comfortably facilitates the flow of nutrients to the vertebrae, flow of waste out of the vertebrae, and lessens fatigue (Adams & Dolan, 2005).

Chair designs can be evaluated giving users a modest task to perform, such as maintaining focus on a computer monitor and keeping contact with the keyboard, while being asked to change positions in the chair periodically (Bush & Hubbard, 2008). The chair would be equipped with sensors along the arm rest and seat. The user would wear sensors on the neck, elbow, wrist, upper, and lower legs. There would be a sensor pad under the chair to capture movements of the chair. The criteria of chair performance would be sustained head, hand, and eye location on the part of the operator, sustained contact with the back of the chair while working or changing position, and the ability to change position comfortably.

There is now reason to think that too much time spent sitting can present health challenges even in the most comfortable seats (Husemann, Von Mach, Borsotto, Zepf, & Scharnbacher, 2009). Employees' work routines should be organized with a mix of standing and sitting and some opportunity or necessity to walk around.

Handtools

Handtools convey physical forces from the human hand to the working end of the device. Optimal handtool designs attend to the forces involved, the shape of the handgrip, and the size of the hand. Consider any handy assortment of handles on kitchen utensils. Note how the shapes of the handles vary from a straight piece of wood to a handle with a greater contour for the handgrip. Some utensils have a soft surface that is compatible with a wide range of handgrips so long as the diameter of the device is not uncomfortable.

Different tools require different handgrip shapes. The fist that grabs the hammer is using the *power grip* (Pheasant, 1996). The curved handle on the pliers facilitates the *torque grip*, which involves a motion of grabbing hold of the device and twisting it.

We use the *precision grip* when holding a pencil, pen, exacto knife, or similar device. Pencils have evolved to a standard diameter, undoubtedly with the help of the manufacturers of pencil sharpeners. Slight differences in the contours of the barrels of common pens may add a little comfort for some users. No one knows for sure whether the shapes make the handwriting any more legible, however.

Carpal Tunnel Syndrome

Many tools require or support repeated hand–wrist motions. When these tasks are performed all day every day over many years, carpal tunnel syndrome could result. The condition is an inflammation and irritation of the nerves going through the carpal bones of the wrist (Figure 8.11). The condition is painful, and the remedy often requires surgery.

It is possible to redesign workstations to lessen the risk of carpal tunnel syndrome. In some cases, it might be possible to design machines to perform the repetitive hand–wrist tasks. Another group of options is to redesign the handtools or simply adjust the seating so that the hand extends straight from the arm rather than bending upward from the wrist.

Computer keyboards offer some opportunities for preventative designs. Some users are comfortable with a palm reset that is placed in front of the keyboard; it elevates the palms so that the hand is working at a downward slope relative to the keyboard. Another group of innovations involves keyboards that are split down the middle so that the portions of the keyboard that are operated by each hand are angled at 15°. This orientation minimizes horizontal bending of the wrist, especially for augmented keyboards that contain many more function buttons than the standard QWERTY keyboard.

Vibration is also responsible for some incidences of carpal tunnel syndrome. A jackhammer hitting the pavement in the street has some impact on the street, but not so visible is the wave of vibration through the hammer back through the hand of the operator. If the jackhammer (or similar vibrating tool) does not produce carpal tunnel syndrome, it can

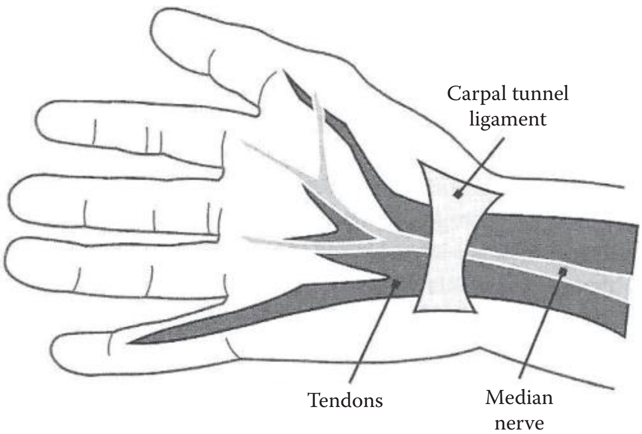

FIGURE 8.11
The carpal tunnel.

produce a nexus of disorders involving damaged nerves, bones, other connective tissue in the forearm, muscles, and joints (International Standards Organization, 1986). *Dead man's hand* is a signal that significant damage has occurred; the fingertips turn white in response to the cold.

Bus drivers have an occupational exposure to carpal tunnel syndrome. The large steering wheel requires larger turn motions than does an automobile, and conveys a substantial amount of vibration from the vehicle. Incidence levels of carpal tunnel syndrome among bus drivers is greater for drivers who also have other types of occupational stress exposures (Guastello, 1991, 1992). The connection that has been suggested is that the drivers who were experiencing greater levels of stress were holding the steering wheels more tightly than other drivers, thus transferring the vibrations to a greater extent. Other stress issues are considered in Chapter 9 in a broader context.

Computer Workstations

The visual display terminals that are common on desktop computers today were first introduced on a widespread level in the late 1970s, particularly in office environments. By the early 1980s, many office workers were reporting eyestrain, body pains, nausea, and even miscarriages that were allegedly due to computer terminals. The devices themselves emitted electromagnetic radiation at levels far below the acceptable levels for common television sets or other appliances. Wattenberger (1980) reported a series of studies that examined what the evils could possibly be. A series of concise studies were conducted with telephone company personnel.

Directory Assistance Operators

The sample consisted of directory assistance operators, half of whom did their work on a computer while the other half used the older paper-and-pencil method. (Computer-based directory assistance functions were just being phased in at the time the study was conducted, so it was not difficult to find samples of operators who were already accustomed to one or the other method.) Each group completed a report of body pains and the Job Descriptive Index (JDI), which is a standardized measure of job satisfaction (Smith, Kendall, & Hulin, 1969) that is still available today in an updated form. The JDI contained five scales for satisfaction with pay, promotion, supervision, coworkers, and the work itself.

Wattenberger (1980) reported significant differences in job satisfaction on all scales except supervision. Greater satisfaction was reported for operators doing the computer-based task. Note that pay, promotion, and coworkers have nothing to do with the experimental comparison, which was the difference in the work itself.

Of the body parts affected by the differences in work systems, only neck pain was significantly different for the two types of workers. Neck pain was reported by 68% of the computer terminal users, which was a higher rate than that reported by the paper-and-pencil users.

Technical Service Representatives

Wattenberger (1980) used the same quasi-experimental plan with telephone technical service representatives, half of whom performed their tasks on computer terminals whereas the other half used paper-and-pencil materials. The participants in the study completed the same inventory of aches and pains, and the JDI.

Job satisfaction results once again favored the computer terminal users, but this time the only effect was satisfaction with the work itself. There was no spillover to other aspects of job satisfaction.

Service representatives who used the computer terminal experienced greater physical discomfort in four body areas: headache, nausea, blurred vision, and sore buttocks. The former three symptoms go together as symptoms of eyestrain. The results of the studies on directory assistance operators and service representatives led to an experimental analysis of computer workstation experiences.

Workstation Experiment

Participants were computer users from a variety of telephone company job groups. There were two independent measures. One was whether the computer used a dark screen background and lighted lettering (common in those days) or a light screen background and dark letters (similar to contemporary screens for word processors or printed paper). The other independent variable was time of day; the researchers were looking for a circadian rhythm effect. Circadian rhythm is considered further in Chapter 9.

The two dependent measures were a subjective rating of discomfort and the *near point*. The near point is how close to the computer the person must sit in order to be comfortable. Results showed no effect for discomfort ratings, but both experimental factors and the interaction between them affect the near point. The results showed that the near point varied over time of day, and different trends over the day were recorded for light screen and dark screen.

Based on the studies, Wattenberger (1980) arrived at several conclusions: (a) Despite the folklore that was growing at the time, people liked computerized versions of the jobs better than they liked paper-and-pencil versions of the same jobs, (b) There were some true physical discomforts associated with computer workstations that could be isolated from imaginary discomforts, (c) Comfort levels did change in the course of the day, and the extent of change depended on the computer screens being viewed, (d) Reorganizing the workers' tasks so that the workers' entire day was not spent seated in front of a computer screen could reduce discomfort, and (e) Discomfort could be lessened by designing and using adjustable computer desks and computer chairs.

Other related findings (Eastman & Kamon, 1976) indicated, furthermore, that the optimal orientation of a computer screen relative to the operator was a 15° upward slope. This angle would match the natural angle of vision from the opposite direction. The reading target should be located within the range of 15° to 20° downward slope of the eye to minimize discomfort (Rempel, Willms, Anshel, Janchinski, & Sheedy, 2007). Although many studies on computer workstations have been published since 1976, it does not appear that the basic points that were made early on were the least bit shaken. A browse through the World Wide Web would turn up numerous sites promoting higher-priced office furniture that is "economically designed."

The Near Point

When the early studies on computer workstations were conducted, the typical screen was a dark background with illuminated characters. The characters were the standard ASCII typeface equivalent to Courier 12-point font on today's word processors. Today the white background is the norm in word processing. The type is usually black, and typefaces vary considerably in style and size. Colored type and images are far more common now compared with a generation ago especially on web pages and when working with computer graphics.

According to Gunnar Ergo (2012), "90% of people who use a computer for more than 3 hr a day suffer some symptoms of *digital eye fatigue*" (emphasis added). Computer in this case would also include hand-held devices, video games, and similar devices with back-lit screens. Symptoms include: "Dry or itchy eyes, eye strain, watery eyes, headaches [and] blurred vision." Their response was a design for eyewear that wraps around the head to some extent to keep in moisture around the eyes, a lens that filters spikes in fluorescent light sources. The lenses are coated and made (optionally) with progressive magnification to support one level of magnification for looking at a screen and a slightly stronger one for reading from source material that is positioned close to the keyboard.

In light of the growing number of people afflicted with digital eye fatigue and sore necks in addition, research was resumed on the best location for the near point. The near point itself represents a balance between sitting close enough to see clearly and minimizing the extra demand on the ocular system that goes with near-field vision. The participants in Rempel et al. (2007) performed 2.5 hr of visually intensive work on a computer while sitting at three distances from the screen; the distances were 46, 66, and 86 cm). The furniture was controlled so that the visual targets would appear within the 15°–25° window, and the back of the chair was set of a 110° slope. Once the set heights were adjusted for the individual partipants, they could move their body position within the chair but could not move the chair itself or the computer screen. Dependent measures were several indicators of posture within the chair, head angle, and effective visual distance, with several significant effects associated with viewing distance. The net effect was that the longer the viewing distances induced greater fatigue because the participants made more postural adjustments to compensate the distance.

Rempel et al. (2007) used standard text sizes found in most business applications (10- to 12-point type), and noted that longer viewing distances require larger size type in order to resolve visual accuracy than what is commonly used in business environments.

Most programs, however, are equipped with a visual zoom feature instead. So, although the stress on the oculomotor system is greater at shorter distances, visual acuity is resolved at shorter distances. Postural and seating adjustments, zoom, and visual correction all offer some degrees of freedom for resolving visual acuity and minimizing fatigue.

Workstations in Health Care Settings

Just when one is inclined to think that the ergonomics of computer workstations was figured out 30 years ago, and the implementation of the take-home messages should no longer be difficult, the classic issues of body aches and pains arose in the health care industry (Hedge & James, 2012). The impending new national health care system in the United States, which was scheduled to go live in 2014, requires all hospital facilities to convert all their paper and other nondigital records to digital form. As a result, male physicians in one hospital are reporting up to 5 hr per day working on digital data entry in addition to 2 hr per day working on a computer in their home environment. Their female counterparts are reporting up to an hour more per day engaged in the same tasks. Up to 30% are reporting pains in their (usually right) hands, wrists, forearms, and upper arms, and 50%–80% are reporting pains in their upper backs, lower backs, necks, and shoulders.

Lessons from the past strongly suggest that their workstations are not ergonomically suitable. Furthermore, the proliferation of laptop computer designs precludes the 15° rise in the keyboard, which affects the angle of the hand on the keyboard; laptops are typically designed with flat-to-surface keyboards, which either instigates or exaggerates the physical discomfort. If the problem turns out to be as widespread as it looks, several additional questions immediately arise. For instance, if the physician is spending so much of the work day doing data entry, what is happening to patient care? Why are highly skilled individuals spending so much time on data entry tasks?

The report by Hedge and James (2012) only considered the ramifications of so much typing. The friendliness of the software has not yet been unpacked. There are at least four different software vendors whose products are targeted to the formatting of a database that would be compliant with national health care requirements, and over 200 software vendors producing reports that need to be converted in some way or fashion. The point, for the time being, is that any sources of user unfriendliness in the design of the software product can be expected to add to the perceptions of physical symptoms.

DISCUSSION QUESTIONS

1. How would anthropometric data be valuable in virtual reality computer programs? Describe some different types of programs and the data requirements that might be associated with them. (This question will be revisited in a later chapter.)
2. Are there any potential musculoskeletal hazards in the electronics assembly industry associated with jobs that involve coil winding or trimming, circuit board wiring, fastening parts, and packing products? What types of injury potential are evident, and why?

3. Is an adjustable desk really necessary in a computer workstation? If this suggestion were implemented, would there be any side effects or deterrents?
4. Describe some consumer products that could be improved with better attention to anthropomorphic data.
5. Are some consumer or industrial products better suited to right-handed people than they are to left-handed people? How or why would that be the case?
6. Some designs for writing utensils might have placed a value on style and shape over functionality. Consider a variety of designs, and discuss their relative merits. Does the pen slip out of the user's hand or require the user to hold the pen unusually close to the point or unusually further away from the point?
7. The workstation studies were conducted with desktop computers which typically involve a desk and a chair. What about laptop users? Where do they place their laptops when using them? Is it really on their laps? What is the visual angle to the screen? What distancing do the users regard as comfortable?

9

Stress, Fatigue, and Human Performance

This chapter describes the nature of stress, its sources, health consequences, and how stressors combine to affect human performance. The contributing research emanated from numerous sources including industrial production, athletics, and laboratory experimentation. The experience of fatigue is often a result from stress exposure, but it is also a dynamic in its own right. A substantial portion of the relevant thinking for some of these areas dates back to the early 20th century. Cognitive load and fatigue effects are often difficult to separate, but recent developments with nonlinear dynamics and enhanced experimental designs have made the separation possible. Thus, the chapter elaborates four cusp catastrophe models for four types of stress–behavior mechanisms: the ability–stress model, the buckling model, the diathesis–stress model, and fatigue. The first contributions from a series of studies are described in the final section of the chapter.

Nature and Types of Stress

Stress, according to the classical definition, is the nonspecific reaction of an organism to any environmental demand (Quick & Quick, 1984; Selye, 1976). This definition is extremely broad and suggests that any stimulation at all is a form of stress and that the negative, undesirable consequences of stress are really a matter of degree and interpretation. Both desirable and undesirable events can produce stress (Holmes & Rahe, 1967).

The physiological dynamics of stress are rooted in the two subdivisions of the autonomic nervous system. The sympathetic nervous system is responsible for bodily arousal when it is activated. Once sufficient energy has been expended, the parasympathetic nervous system is activated and the body relaxes; relaxation is a physiologically active process, not a passive process. Stress reactions occur when arousal is frequent, but there is no opportunity to expend energy at the moment when it is necessary to do so, thus prolonged arousal results. Exercise programs are thus frequently recommended to people who need to alleviate the effects of stress in their lives (Zwerling et al., 1997).

There are many types of stress that a person can experience. Although they have common results on the person, it is convenient to classify them according to their sources because the goal is, ultimately, to reduce unnecessary stress both in the workplace and in one's personal life. The four broad categories are, therefore, physical stressors, social stressors, speed and load stress, and dysregulation due to irregular work schedules. The latter has some entanglement with fatigue in the sense of prolonged time on task.

Physical Stressors

Toxins

Physical sources of stress would include toxins, excessive heat, excessive cold, and noise. Toxins comprise a large and varied category. As a general rule a person is likely to experience the common stress symptoms, but with specific health impairments peculiar to the toxin (Travis, McLean, & Ribar, 1989), many of them are severe to the point that we think of them as immediate and present dangers more often that sources of stress. OSHA standards for the safe handling and ventilation of chemicals should be consulted; requirements are specific to the type of toxin involved. Toxins present in confined spaces present additional dangers and challenges as described in the previous chapter.

Extreme Temperatures

Excessively cold temperatures in the work environment can impair work performance and cause health problems, notably hypothermia. Nonetheless, appropriate clothing and training to perform the work under cold conditions can counteract the performance effects of cold (Enander, 1989). For less extreme environments, the standard of comfort is a normally ventilated room that is heated to 72°F, although people do express some individual differences in preference for a little more or less heat (Kantowitz & Sorkin, 1983). Turbulent airflow produces the sensation of colder temperatures, or windchill. Windchill is the psychologically equivalent temperature of outdoor temperature that results from actual temperature plus turbulence.

Clothing compensates for cold temperatures, of course. Each article of clothing has a number of clo units associated with it. A person should wear enough clo units to produce a comfort level equivalent to the standard. Ultimately people pick clothing based on what is in their closets and the known comfort levels associated with those items. Chicago railroad workers advise each other as to how many layers of thermal underwear and sweaters are needed in addition to the outer garments to get through a night of –35°F with the windchill factor.

Gloves are part of cold-weather apparel as well as standard personal protection gear in better weather. Gloves can be problematic sometimes because they interfere with fine psychomotor skill by exerting sheer forces between the operator's hand and the outer environment. The choice of glove is a trade-off between psychomotor flexibility, warmth, and protection from cuts and scrapes, and it is possible to purchase gloves that suit a range of possible requirements. For instance, when dress gloves were part of women's standard evening apparel, one manufacturer advertised, "Gloves so thin you can pick up a dime."

Excessive heat is a different matter, as there is a limit to the amount of clothing a person can remove. Performance decrements begin to appear at 86°F after 8 hr of exposure, but begin after 45 min of exposure to 100°F heat. The critical temperature of 86°F is the point where the body starts to build up heat internally. Ventilation can ameliorate some but not all of the effects of severe heat (Kantowitz & Sorkin, 1983). Current OSHA standards for heat exposure are predicated on the type of industry and the chemical or other hazards that are also present. Two standard responses to extreme heat exposure are to stay hydrated by keeping water available and using salt tablets. The salt helps the body retain water and slows the loss of electrolytes through perspiration.

Hancock, Ross, and Szalma (2007) conducted a meta-analysis of 48 studies published over the previous 60 years on the effects of thermal stressors on task performance. A meta-analysis is a set of statistical techniques that isolates the effect sizes for independent variables reported in a variety of studies conducted in different settings and converts them into a common metric that permits conclusions about overall effect size and the search for explanations for the variability in effect sizes that are reported. They found that the overall effect size for thermal stress on performance was –0.34 standard deviations of the performance metric, which in turn was comparable to an 11% loss in performance overall. Effect sizes varied by whether the individual studies focused on heat or cold and the extremity and durations of the exposures that were studied.

In the case of heat exposure, performance loss increased relative to thermo-normal conditions as the ambient heat became more intense. The drop in performance was not as severe as the increase in internal temperature build-up. Heat exposure affected perceptual performance the most, followed by motor performance; cognitive performance was the least affected of three types of performance.

In the case of cold exposure, performance loss began at 52°F and dropped proportionately with colder temperatures, relative to thermo-normal conditions. Perceptual performance was again the type most affected, but the effect was stronger for the cold (–1.13 SD) than for heat (–0.78 SD). The effect of cold on motor performance (–0.41) was less than the effect of heat (–0.78). The effect of cold on cognitive performance actually produced an *increase* in performance (+0.41) instead of a negative effect (–0.23) for heat.

The effect of duration of exposure was not so simple. An acclimatization effect occurs whereby people adapt to the temperatures in much the same way as they adapt to other forms of repetitive or constant stimuli. Hancock et al. (2007) found that a cross-product of exposure time and intensity was a good predictor of the effect size, however.

Noise

Noise is another physical source of stress, in addition to its ability to interfere with signal detection and produce hearing damage with prolonged exposure. Noise sources produce arousal, which can be either good or bad, but if the task is already challenging enough it produces discomfort. High noise levels require the operator to scan the noise for signals, some of which could denote danger, which means a greater cognitive processing demand to separate true signals from unwanted sound.

The research on the effect of noise on performance produced some generalizable trends (Cohen, 1980; Kantowitz & Sorkin, 1983). First, a good deal of the stress-related impact of noise is simply the annoyance level associated with the sound. People are less annoyed by a noise source if they believe it serves a useful purpose. Second, the noise is more stressful if it is intermittent rather than continuous. Third, noise is more stressful if it is uncontrollable, rather than controllable. Thus, one means for controlling the effect of stress is to give those who are exposed to it some modicum of control over it, teach them to look for degrees of control, or remind them of degrees of control that are already available that they might not have recognized.

Social Stressors

Stress can be induced by social sources, and it is convenient to subdivide those into source categories: work related and nonwork related. Work-related social stressors would include role ambiguity ("What am I really supposed to be doing here? No one is making it clear what they want"), role conflict (trying to meet conflicting demands), obnoxious supervisors and coworkers, job insecurity, new job assignments, and insufficient authority to perform tasks necessary for work assignment and planning for retirement, to mention a few.

A personal success can be stressful too. Not all stressors are unwanted events. Stressful events are arousing, which is what makes them stressful. Nonwork-related stressors include illnesses or death in the family, divorce, changing homes or hobbies, change in work patterns of family members, changes in eating or sleeping habits, financial difficulties, the holidays, and Christmas in particular (Holmes & Rahe, 1967).

On the other hand, it is also known that a good social support system can counteract the other forms of stress to produce an adaptive result. The relationship between negative life crises among submariners and illnesses was lower for people reporting stronger social support systems than for people with weaker social support (Sarason, Sarason, Potter, & Antoni, 1985). The same point has been underscored for promoting the well-being of military personnel returning home (Cornum, Matthews, & Seligman, 2011; Reivich, Seligman, & McBride, 2011) and people experiencing substantial stress from any life sources (Pincus & Metten, 2010). Engaging social support is one aspect of a broader construct of *resilience*, which is considered further later in this chapter.

Personal social networks, that is, friends and family, are the primary sources of social support. In contained environments, comraderies within a workgroup commonly develop around shared objectives, shared challenges, and shared interests that are unrelated to the work environment, such as athletics, music, or other entertainment preferences. In severely stressful and isolated environments, the social cliques sometimes ostracize people who are showing signs of poor adaptation, which can be expected to make matters worse for those people (Palinkas, 1987).

Crowding and Isolation

Crowding and isolation are forms of stress that derive from both social and physical sources. In the case of crowding, workgroups in offices show better performance and less absenteeism if there are fewer walls around their personal workspaces, greater distance between one person and another in the work area, and fewer people in the office container overall (Oldham & Fried, 1987).

Note that there is a qualitative difference in the experience of working alone, versus working with one other person, versus working with a larger N of people, versus $N + 1$. There is a bit of folklore that traces back to the British Navy wherein the boss assigns two men to a job no matter how small the job. This writer has seen firsthand that the technique worked profoundly well for industrial maintenance and house renovation crews. The idea did appear silly to people who encountered it for the first time, but before the work day was over, they reported that they completed their tasks faster, and they did eventually need to have a coworker assist with some aspect of the assignment. In the case of the house renovation crew, the work for 15 people that would have required 8 hr of work with individual assignments was done in 5 hr with the two-person system.

Although these cooperative behaviors are usually expected to produce positive effects on performance, sometimes the opposite occurs (Barnes et al., 2008; Stachowski Kaplan, & Waller, 2009). The presence of others, the depth of interactions among participants on the task, the complexity of the task, and evaluation apprehension can have positive, negative, or interactive effects on group performance (Bond & Titus, 1983, Hertel, Kerr, & Messé, 2000).

The case of working alone rather than with a coworker is only one form of isolation. Much more stressful forms of isolation are experienced in Antarctic winter-overs and outer space travel. Submariners are also isolated from communication with the topside world for weeks at a time, but their daily experience appears to be dominated by the close crowding rather than the isolation. The valuable support system mentioned earlier does require interpersonal sensitivity and responsiveness on the part of each submariner in order for the crew to remain effective (Sandal, Endresen, Vaernes, & Ursin, 1999).

It is also noteworthy that the predictors of psychological well-being among workers in extreme environments do not generalize across all extreme environments. Differences in situations with respect to physical, individual, and social demands affect the profiles of the stress-resistant person (Sandal, 2000). Observations such as these have led to a broader proposal that individual differences that affect performance in person–machine systems should be given a lot more attention than they have been given traditionally (Szalma, 2009).

Electronic Monitoring

Electronic monitoring of employees' work performance became a concern in the early 1990s as a result of widespread conversion of work to computer-based media. Monitoring capability facilitated feedback from supervisors or their avatars. Although, according to principles of basic learning theory, feedback reinforces effective performance, correct ineffective actions, and raise performance overall, employees were seeing the presence of electronic monitoring as stressful regardless of whether the substance of the feedback was positive or negative (Aiello & Kolb, 1995). Restricted mobility from being tied to a computer terminal is a source of stress by itself, but it became reportedly worse when monitoring systems were implemented.

Aiello and Kolb (1995) recommended that building a climate of positive social support was essential for minimizing stress and for turning feedback systems into positive influences on performance. Organizational climates can exacerbate stress by encouraging aggressive and oppositional management practices and similar relationships among coworkers (Van der Velde & Class, 1995). Organizational members respond to the negative features of the social environment in turn with passive–defensive strategies such as angling for approval; acting in conventional or conforming ways when individual thought, action, and expression would be appropriate and valuable; dependence and avoidance.

The foregoing climate factors were part of the constructs of role stress and role ambiguity that were first introduced by Kahn, Wolfe, Quinn, Snoek, and Rosenthal (1964). A voluminous literature on these and other constructs of organizational climate has developed since that time. A particular subvariety of organizational climate, safety climate, is considered in the next chapter.

Speed and Load

In the previous chapter on psychomotor skill and control, the relationship between errors and RT was observed for situations where work speed was allowed to increase and workload was held constant. Chapter 6 addressed the issues connected to human cognitive channel capacity and the effect of its limitations on speed and accuracy. A classic study characterized what happens when workload and speed are both allowed to vary, which follows next.

Conrad (1951) made a distinction between speed stress and load stress. *Speed stress* is a reaction on the part of persons working on a task that has the effect of worsening their performance beyond what might be expected from the physical characteristics of the work or equipment involved. *Load stress,* on the other hand, changes the character of the task. As the number of signal sources (e.g., visual displays in the experiments) is increased, more time is needed to make judgments simply because of the greater amount of information that is being processed. Conrad's experimental participants were engaged in a clock-watching task in which they pressed a key as a pointer approached the 12:00 or 6:00 o'clock position on any of the clock dials used. In the various experimental conditions, two, three, or four dials were used, and speed was varied. Errors increased as the product of speed and load increased.

Working Too Slowly

Errors in the low-load condition are less readily explained. Boredom or stress resulting from boredom is one possible answer. There could be other viable explanations as well, but the literature that addresses slow-speed conditions is sparse. Consider the following anecdotes and the available data to support them:

A physical rehabilitation specialist has a regime for muscle training using relatively standard progressive weight resistance equipment, but with the special regimen of pushing against the weights very slowly (Ware, 2011). The trainee's legs visibly shake while doing the slow push movements in the early stages of training; the initial deficit could have contributed to this effect. The training regime has reportedly a high rate of success. The explanation for the effect is that the neural mechanisms are being forced to find new pathways or firing combinations that are more viable than the initial pathways. The newer and more viable pathways consolidate through further training.

Kremen (1966) studied a tracking task in which the target moved at different speeds and in different directions. Accuracy and smoothness of the tracking motion were the two dependent measures. Increasing the target speed resulted in smoother motions; slow-speed performance was more erratic.

Hausdorf et al. (2001) studied the variation in gait for people walking at their normal pace or at the same, faster, or slower pace that was regulated by a metronome. They were looking for differences in the fractal dimension of the time series of stride intervals under those conditions; fractal dimensions would correspond to levels of complexity and turbulence in a time series that arose from a deterministic process rather than a random process (see Chapter 2). Hausdorf et al. (2001) found that the coherence of the time series dropped under the three metronome conditions and that randomness increased. The reduction in gait coherence was greatest in the slow-speed condition. West and Scafetta (2003) developed a mathematical model that emulated Hausdorf et al.'s results; it required two internal control centers, one located in the motor control mechanism and another in the central nervous system. The external driver appears to disrupt the correspondence between the two internal systems (West, 2006).

The picture taking shape from the foregoing vignettes is that deliberate too-slow pacing disrupts preexisting psychomotor coordination or cognitive automaticity. The further implication is that more than one control center is involved in the psychomotor process, and there would have to be a feedback loop of some sort between them or else they could not coordinate. This explanation would be consistent with what is known about synchronized systems. The minimum elements required to synchronize two subsystems are two oscillators, a feedback loop between them, and a control variable that speeds up the oscillators

(Strogatz, 2003). Once the oscillators reach a critical speed, they suddenly undergo a phase shift whereupon their actions synchronize. The theory of synchronization does not carry a direct implication for human error rates. Rather, if one viewed the synchronized state as intrinsically undesirable, such as the hypersynchronization of neurons firing that occurs in an epileptic seizure, the synchronized state itself would be characterized as one large system error. Similarly if synchronization represents a desirable state for any reason, the desynchronized and uncoordinated state would be regarded as one large system error.

Most of the empirical work to date on synchronization in individual psychological processes has been done with psychomotor tasks (Jirsa & Kelso, 2004a; Sleimen-Malkoun, Temprado, Jirsa, & Berton, 2010). The most common application involves bimanual finger-tapping, which exhibits the classic synchronization structure very clearly: Subjects tap alternating index fingers to a metronome. When the metronome speeds up enough, a phase shift occurs whereby the two fingers tap simultaneously. Although possible models for the synchronization of cognitive process are still in the formative stages, one suggestion is to look for combinations of cortical and thalamic activation centers, or the behavioral manifestations of those processes (Jirsa & Kelso, 2004b; Ward, 2004). Another suggestion is to look for patterns of engagement and non-engagement of executive control functions while performing a task (Hollis et al., 2009).

Recent studies on thought–action patterns also suggest that decoupling could be involved. Although one explanation for errors in vigilance tasks attributes errors to the depletion of cognitive resources associated with cognitive workload and fatigue, an alternative explanation is *mindlessness* (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997), which is the tendency for the mind to wander while performing the task. It now appears, however, that the experimental task structure that was thought to induce mindlessness was actually producing workload and fatigue effects of the more conventional variety that result from relatively *high* mental loads (Helton, Weil, Middlemiss, & Sawers, 2010). Although the propensity for mindlessness to occur under conditions of *low* cognitive workload has not been considered explicitly, Balagué, Hristovski, Agagonés, and Tenenbaum (2012) recorded periods of task-relevant and task-irrelevant thoughts while participants performed physically demanding exercises. Their results showed that the patterns of task-relevant and task-irrelevant thought oscillated aperiodically through the work session. The switching between the two types of thought was interpreted as a coping mechanism for the boredom and ensuing physical fatigue. Task-related thought appeared to be a part of efforts to reshape the performance of the task and consistently became dominant shortly before the exhaustion point.

Signal Detection Tasks

The next question is whether there are any similar effects of working too slowly apparent in a signal detection or vigilance task. Evidence favoring the existence of multiple performance processes could be particularly enlightening. Signal detection tasks offer the opportunity for two types of errors, misses and false alarms. It is also well known that the errors can be readily biased in favor of one error or another by the rewards or penalties associated with them. Balci et al. (2011) designed an experiment that induced participants to make a trade-off between accuracy and rewards. They found that bias toward accuracy dominated the early stages of learning and performance, and a bias toward higher overall rewards dominated later on. In this case, the two subprocesses gained coordination with practice. Guastello, Boeh, Schimmels et al. (2012) reported a similar shift in performance strategy in an episodic memory task before and after an hour of work on another task.

In a visual search example, Zenger and Fahle (1997) manipulated load by increasing the search target area to produce a pressure on response time. Missed targets were more common than false alarms, and they were traceable to a speed–accuracy trade-off.

Szalma and Teo (2012) constructed a vigilance task that varied the number of screens containing possible targets and the speed of screen presentation. For that set of conditions, there was a linear decline in accuracy as a function of speed for low-demand content, but the classic breakpoint was observed as the number of screens increased from 1 to 8. False alarms declined with increasing speed for the 8-screen condition only.

According to Koelega, Brinkman, Hendriks, and Verbaten (1989), research prior to their study had shown that the density of targets and the memory load inherent in the task affected whether error rates set in during slow- or fast-working conditions. They assessed performance on four vigilance tasks for speed, accuracy, electrodermal response, and subjective ratings. The tasks varied in memory demand and by type of stimuli. A principal components analysis of the measurements produced two components that indicated that speed and accuracy could result from two different underlying cognitive processes.

Work Schedules

The standard workweek in the United States is 8 hr per day, 40 hr per week. Employers who need staffing for 16 or 24 hr per day add second and third work shifts. More businesses in recent decades have found the need for staffing 24 hr a day, 7 days a week, and this norm is not limited to the United States. The puzzle is how to define work shifts that produce the best results with the least stress and strain on the workforce. There are some people who prefer evening or night shifts by virtue of activity patterns in their family lives or natural tendencies to be night owls. The need for employees during nonstandard times, however, greatly exceeds the number of people for whom such work arrangements are actually convenient and stress-free.

According to Costa (2003), there are thousands of different work schedules in operation throughout the world. They can be distinguished, however, by several characteristics: (a) the number of work shifts in the course of the day, (b) the length of the work day, (c) the extent to which night work is involved, (d) whether weekend work is part of all work crews' schedule or whether weekend-only crews are involved, (e) whether rotation of work hours is involved and the direction of rotation, (f) the amount of time between changes in work schedules, and (g) whether work starts at different times of day for a particular crew. One common type of rotation is across times of day, such as 2 weeks on the 11:00 p.m. to 7:00 a.m. shift, followed by 2 weeks on the 3:00 p.m. to 11:00 p.m. shift, followed by 2 weeks on the 7:00 a.m. to 3:00 p.m. shift. Other employers might rotate in the opposite direction.

The stress impact of work schedules arises mostly from interference with circadian rhythm, which is a daily biological cycle. Another large component of stress is the interference with the patterns of home life. For instance, workers on the 3:00 p.m. to 11:00 p.m. shift might experience greater stress than those on the 11:00 p.m. to 7:00 a.m. shift because they leave for work when nobody is home and arrive home when everyone is asleep. Those who work through the night may be out of synchrony with their natural biological processes, but at least they cross paths in the morning and evening with the rest of the household.

Circadian Rhythm

The human body exhibits a natural cycle of waking and sleeping known as *circadian rhythm*. Within the waking and sleeping parts of the day, there are patterns of wakefulness and sleep. The full cycle takes 25 hr if the person is in a contained environment, such as a cave, without contact with the outside world. The cycle is closer to 24 hr long for most people who are exposed to the normal planetary light–dark cycle (outside the polar regions).

The purpose of sleep is to restore the physical and mental well-being of the individual. Prolonged sleep deprivation eventually results in mental imbalances that resemble psychoses, and multiorgan failures. At the same time, sleep disturbances are symptomatic of many forms of mental illness, so it is better to say that sleep and mental well-being have reciprocal effects on each other. Sleep occurs in 90-min subcycles, each of which consists of four stages. The four stages range from light and almost awake (Stage 1) to deep (Stage 4), and are characterized by distinct electroencephalogram patterns associated with each stage (Andreassi, 1980).

At the end of Stage 4 sleep, the sleeper returns through the other stages quickly to a different form of Stage 1, known as REM sleep, where the acronym stands for rapid eye movements. REM is also known as paradoxical sleep. REM sleepers look like they are going to wake up because the eyes are darting back and forth as if looking at something, but these sleepers are very difficult to wake if someone should try to do so. REM is the stage in which dreams are remembered most often, although dreams occur in other stages.

Repeated 90-min cycles exhibit progressively less time in the deepest sleep. The subjective feeling of restfulness is related to the amount of time that one actually spends in Stage 4 sleep. Thus, the first sleep cycle is perhaps the most important of all.

Waking periods display different patterns of alertness, body temperature, and adrenaline levels in the bloodstream (Akerstedt, 1977; Colquhoun, 1971; Wojtezak-Jaroszawa, 1978). If the person awakes at 6:00 a.m., alertness is mediocre at first, but increases to its peak by midmorning. The peak may be sustained until after lunch, when the parasympathetic nervous system is doing its job to process lunch. Arousal level reaches a nadir by the later afternoon, but a natural shot of adrenaline around that time improves the alertness levels up in time to drive home from work. Alertness peaks again in the late evening, and tapers off around a normal bedtime of 10:00 p.m.

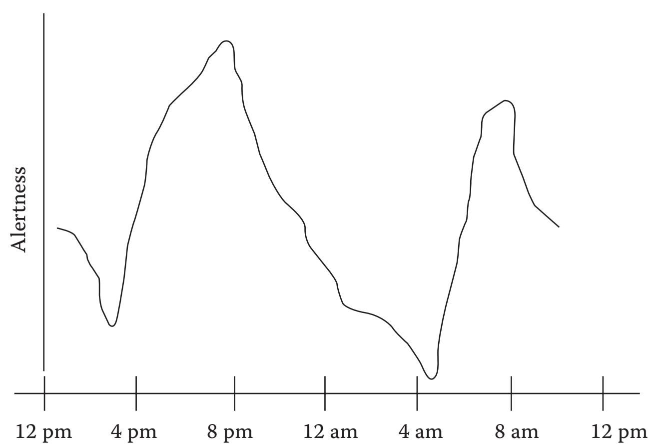

FIGURE 9.1
Alertness levels during a 24-hr period.

If the person is staying up all night instead of going to sleep, the alertness level continues to decline until it reaches its lowest point around 3:00 a.m. People who are compelled to stay up all night will report that if they can force themselves through the 3:00 a.m. period, the adrenaline shot that comes in the early morning will carry them through into the next day. A chart of the circadian alertness level appears in Figure 9.1.

Sleep is triggered by natural cycles of melatonin levels in the brain. Wakefulness is triggered by adrenaline and also by light. People who need to force themselves awake during the night are thus advised to turn lights on. People who are trying to sleep in the day should darken their rooms. In the Arctic and Antarctic Circles, the pattern of outdoor lighting where it is night for 6 months and daytime for 6 months can pose a serious challenge to sleeping for those who have not acclimated to the environment. In the less extreme subpolar regions, and to a lesser extent in the temperate zones, the short daylight hours can induce *seasonal affective disorder* toward the end of the dark season for some of the residents. Seasonal affective disorder is a mild form of depression traceable to living in darkness for too long; it is a temporary affliction but it is probably not helping people who might be prone to mild clinical symptoms for other reasons.

Dysregulation

Dysregulation is a particular form of stress that is characterized by a combination of changes in eating and sleeping habits, and difficulties with them. It is often caused by shift work or irregular work schedules (Depue & Monroe, 1986; Kundi, 2003). Shift work often has deleterious effects on individual work performance because it interrupts or conflicts with the daily biological (circadian) rhythm cycle, which includes the sleep–wakefulness cycle, concomitant brain wave patterns, and variations in body temperature and adrenaline secretion. Permanent night workers have been found to show out-of-phase rhythms, which, in turn, had an effect on their productivity and effectiveness (Agervold, 1976; Malaviya & Ganesh, 1976, 1977). Rotating shift workers express greater amounts of sleep inadequacy, gastrointestinal ulcer, eating problems, irritability, and job dissatisfaction compared with nonrotating shift workers (Tasto, Colligan, Skjel, & Polly, 1978).

Military operations often require prolonged work periods at irregular intervals in order to execute missions or maintain vigilance and readiness under hostile conditions. Depending on the demands of the exercise, the military personnel could experience shortened sleep periods but not sleep debt (need to catch up on sleep), or they might indeed experience shorter and more fragmented sleep periods (Naitoh, Kelly, Hunt, & Goforth, 1994; Thompson, 2010).

In addition to the disruption of circadian rhythm and collision with family demands, job demands, and social conditions, individual characteristics play a role in the amount of strain experienced by a worker (Costa, 2003). Job satisfaction issues and related labor conditions can make a stressful situation worse. For instance, in this writer's experience, some rotating crews reported that they preferred their night shifts to the day shifts because there were fewer interruptions from management and other business activities that only took place in the daytime. Night-shift workers in public transportation reported the opposite experience; all the rowdy and dangerous characters come out at night even though there might be fewer passengers overall.

The social conditions that were mentioned in Costa's (2003) review article included the tradition of shift work in a particular industry, the supportive nature of the working atmosphere, and transportation ease. The worker's social support outside the workplace is also important. Work–family conflicts may vary in size and scope. Workers who do not have families have social networks that may support particular work schedules or run contrary to them.

Individual characteristics include age, gender, general health, personality characteristics, and the extent to which the workers are able to manage sleep at nonstandard times. One important personality characteristic is internal locus of control, which is the belief that good and bad outcomes in life are under the individual's control even though the role of the environment may appear large to other people (Smith & Iskra-Golec, 2003). Generally, people with internal locus of control are less susceptible to the negative effects of stress than those with external locus of control. Externals believe that good and bad outcomes in life are under the control of the environment, and there is not much that an individual can do about them.

It is not a simple matter to optimize a work schedule for a given situation given the number of variables involved—schedules that might be relatively good in one situation may be relatively poor in another (Costa, 2003). Kundi (2003) offered some general recommendations concerning night work. Whenever possible, the schedule should: (a) minimize the number of hours of night work, (b) minimize the variability of night work hours, (c) spread the night assignments evenly across work crews, and (d) support longer time periods on a given work schedule rather than shorter periods.

Consequences of Stress

The consequences of stress can be grouped in performance-related outcomes and health-related outcomes. The health-related outcomes are both psychological and physiological.

Performance

The effects of stress on work behavior can take the form of errors, accidents, absenteeism, and turnover. Figure 9.2 depicts the effect of arousal on performance (Yerkes & Dodson, 1908). Too much or too little arousal, besides being subjectively stressful, results in

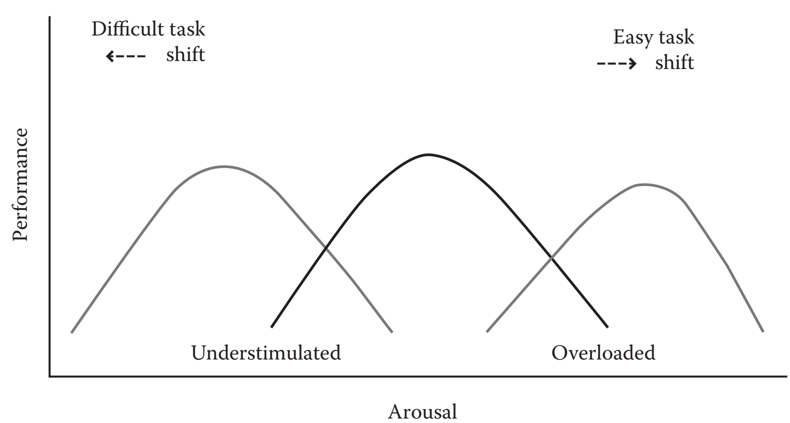

FIGURE 9.2
The Yerkes–Dodson law for performance as a function of arousal.

suboptimal work performance. Of further interest is that difficult tasks require less arousal from other sources for optimal arousal, whereas easy tasks are best performed with alternative forms of arousal added, such as playing a radio in some situations (McGrath, 1976). Some of the more complex effects of stress and work performance are elaborated further on in this chapter.

People can cope very effectively with jobs that do not engage the optimal level of arousal. According to Hancock and Warm (1989), people maintain a steady and effective level of performance in the neighborhood of the optimal point. When workload gets too far outside the normal comfort zone, they engage in coping strategies to stretch their zone, as depicted in Figure 9.3. It is not clear what all the possible coping strategies could be, but some have come to the foreground in the discussion already: off-loading complicated or time-sink tasks to other people or another time, ignoring social interactions that are irrelevant to the task, using automatic thinking processes and less executive control, and working for greater speed and less accuracy. In the case of work underload, the individual might engage in conversation with coworkers, play the radio, or do something else while the jobs in the low-volume task pile up to a critical mass. The choices of strategies would be limited by what is afforded by or inherent in the work itself. It follows, nonetheless, that strategies for counteracting overload are different from strategies one might use to counteract underload.

Inevitably, demand can exceed the coping zones in either direction. Hancock and Warm (1989) characterized the sharp drop in performance that occurs when the coping zone has been breeched as possibly resembling a catastrophe or chaotic function. Augmented cognition or variations thereof are attempts to stretch the effective coping zone. Such help is often paired with decisions about whether to trust the system to perform tasks and make decisions that would be correct for the situation. There are also situations where setting up the relevant software to do the task is more troublesome and time-consuming than doing the job in the usual fashion.

People who work under conditions of stress typically do not work as effectively as those who can work under more hospitable conditions do. Stress has a greater negative impact on performance when the stressors are uncontrollable and unpredictable (Cohen, 1980). One of the more theoretically challenging observations in the stress experiments is the

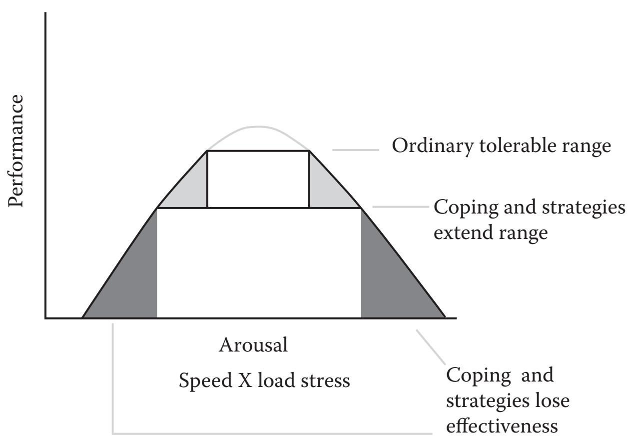

FIGURE 9.3
The impact of coping strategies on the arousal–performance relationship as described by P. A. Hancock and J. S. Warm (*Human Factors*, *31*, 519–537, 1989).

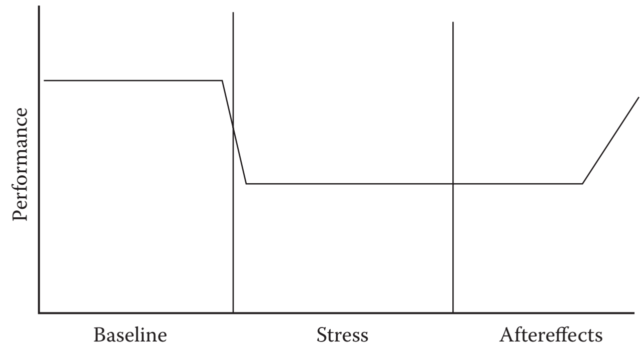

FIGURE 9.4
Aftereffects of stress on performance.

aftereffect of stress phenomenon. After a stressor has been introduced, work performance decreases. When the stressor is removed, however, depressed performance levels continue; the typical effect is shown in Figure 9.4. A viable theory about stress and performance needs to account for the aftereffect phenomenon. The three best-supported explanations for the phenomenon are, according to Cohen, persistent coping, adaptive cost, and learned helplessness.

According to the persistent coping thesis, people under stress reorganize their mental efforts such that they are actively fending off the stressful stimulus in the form of a coping mechanism. The key point is that, once the coping mechanism has been initiated, it is difficult to turn it off immediately. The act of coping with the stressor persists after the stressor is removed (Deutsch, 1964; Epstein & Karlin, 1975).

The adaptive cost thesis (Glass & Singer, 1972) holds that the process of adaptation depletes cognitive resources and, as a result, performance remains depressed until the individual recovers or is sure that the "coast is clear" before regrouping resources and ignoring the possibility of stressful stimuli. The depletion of resources, or cognitive shrinkage, explanation is consistent with theories of human information processing: People scan the stressful stimuli for information that could be meaningful to their health, safety, or success. By doing so, mental channel capacity is consumed, and the available capacity with which to perform the primary task becomes smaller. Once the channel capacity has been reconfigured to include the stress channel, it takes a while to reallocate that channel capacity to the main task; perhaps the person is typically waiting to be sure that the threatening situation has truly ended.

According to the learned helplessness thesis, people under conditions of uncontrollable stress either learn, or have already learned, that there is nothing they can do about the stressor or its effects on their performance, and therefore they do not try to do anything (Seligman, 1975). Learned helplessness reactions have been demonstrated empirically in experiments with both people and animals. Learned helplessness was a premier theory of clinical depression when it was first introduced. Depression is symptomatically similar to severe stress reactions.

Health

Anxiety is an irrational fear, according to classic clinical psychology. Although it could be based on real-world events, anxiety is essentially apprehension gone out of control.

Anxiety can be experimentally conditioned in animals and humans (Wilson & Davison, 1971), and the conditioning aspect of anxiety offers a strong partial explanation for numerous psychological disorders. Anxiety is a nonspecific symptom of many psychological disorders, and could be regarded as the mind's alarm system that goes off when something is wrong. Anxiety is frequently cited as a symptom of prolonged or chronic stress. Indeed, the clinical profile of traits observed for people who have experienced chronic stress exposure are not distinguishable from the personality profiles of people suffering from clinical depression. Posttraumatic stress syndrome, on the other hand, is a severe psychological reaction to acute and intense stress, and the symptoms may be schizotypal as well as depressive (Glover, 1982; Natani, 1980; Walker & Cavenar, 1982).

Strain is the amount of deformity caused by stress. In other words, strain is a measure of damage in the sense of a material science metaphor that is explored in this chapter. The material science concept of elasticity translates into notions of psychological resistance to stress. Some personality variables, such as locus of control and Type A personality syndrome, are relevant to resistance to stress, but the broader concept of buckling and elasticity is less well developed in conventional psychology. The buckling concept is elaborated later on in this chapter. For the most part, health professionals will use the word *strain* to indicate any of the health consequences or deformities produced by stress.

Fatigue is the loss of work capacity over time. It is not synonymous with stress as defined earlier, although it can be thought of as a specific result of a specific type of stress, for example, a heavy and fast-paced workload. Fatigue can be physical or cognitive in nature, or both types could be operating simultaneously. Fatigue is a natural consequence of doing a large quantity of work and is not symptomatic of a problem in the work environment by itself, except in cases where people are compelled to work for too long without a sufficient rest period. Fatigue is reported sometimes from people who have not been doing extensive quantities of work, but they have the subjective feeling of tiredness nonetheless; in these cases, the report of fatigue signifies an emotional difficulty, which could have origins in the workplace.

The physical health consequences of long-term stress include hypertension, heart disease, kidney failure, alcohol and drug abuse, and medical consequences of alcohol and drug abuse. Psychological health consequences of stress include anxiety, impaired interpersonal relationships, impaired decision making, and clinical depression. The onset of many of these health disorders can be traced to background variables and trigger variables (Guastello, 1992, 1995). Hazards in the workplace and physical types of stressors act as background variables. Anxiety and social forms of stress act as trigger variables, with the medical disorder having a discrete (some might say "sudden") onset. In order to obtain an understanding of the impact of stress in a particular workplace, it is necessary to separate the effect of prolonged exposure to stress from the effect of advancing age, which is another background variable.

Stress and Performance Dynamics

This section of the chapter describes four catastrophe models connecting stress precursors to performance outcomes. The essential points leading to the development of these models (Guastello, 1995, 2003) are considered next and followed by synopses of recent studies that examined the stress and fatigue dynamics in the same contexts. There are four types of models: anxiety and performance, buckling stress, diathesis stress, and fatigue.

Arousal, Anxiety, and Performance

Catastrophe models describe sudden changes of events over time. In the case of stress–performance relationships, they describe sharp declines and improvements in performance as a function of stress and other variables. The size of the performance changes are governed by a bifurcation parameter. The proximity of the system to a point of change is governed by the asymmetry parameter. There may be more than one research variable associated with either parameter.

The anxiety and performance models were framed around sports performance (Hardy, 1996a,b; Hardy & Parfitt, 1991). Although Hardy (1996a,b) did explore catastrophe models that were more complex, the cusp model that is shown in Figure 9.5 appears to be the theoretical resting point. Physiological arousal, as measured by heart rate, contributed to the asymmetry parameter. Cognitive anxiety contributed to the bifurcation variable; note that arousal by itself does not necessarily transform into anxiety. Self-confidence also functioned as a second asymmetry variable. (It is possible to have two or more research variables contribute to the same underlying control parameter.)

As physiological arousal increased, the golf player was more likely to move into the manifold region of the behavior surface. Anxiety would have the effect of bringing out the best or the worst in a player, as observed in the basketball study. Self-confidence would bias the player toward higher levels of performance; self-confident people can experience the same sources and levels of arousal as other people, but their interpretation of the arousal does not allow them to be overwhelmed.

Note that the foregoing relationships are consistent with some standard findings in social psychology (also summarized in Aiello & Kolb, 1995) regarding the effect of an audience on performance in sports, public speaking, or other activities. If the performers are confident of their abilities to perform, the added arousal generated by the audience produces better performance. If the performers are not so confident, the presence of the audience has a detrimental effect. The negative effects of anxiety of performance are stronger when the task requires a heavy load on working memory and has its most pronounced negative impact by disrupting automatic processes in the executive cognitive functions (Baumann, Sniezek, & Buerkle, 2001; Sartory, Heine, Mueller, & Elvermann-Hallner, 2002; Williams, Vickers, & Rodrigues, 2002). Anxious performers tend to refocus their attention on lower level tasks such as tracking a ball (Williams et al., 2002).

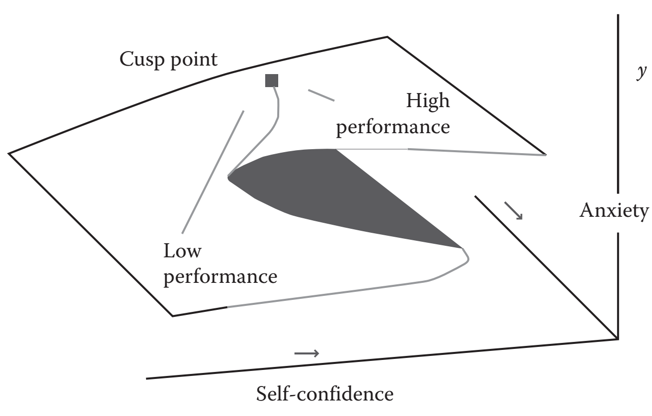

FIGURE 9.5
Cusp catastrophe model for the effect of anxiety and self-confidence on sports performance.

Note that "arousal" in this context might be a better word than "anxiety." The level of excitement associated with arousal often only becomes anxiety when it is interpreted as a negative experience or a sign of an impending negative experience.

Anxiety can also arise from a long-term personality trait that could have genetic origins; in that case it would not be just a transient condition (Eysenck, 1995; Learny & Kowalkski, 1995). It can have both positive and negative effects on performance under stressful conditions (Ein-Dor, Mikulincer, Doron, & Shaver, 2010; Guastello & Lynn, 2010; Matthews & Campbell, 2009). The negative impact of stress is that it leads to broken concentration, response delays, distortions of judgment, and errors; in this context, it might be regarded as a resilience deficit. The positive impact is that it is sometimes associated with greater vigilance and attentiveness to small cues that could signify a large impact. By having both positive and negative roles, anxiety should function as a bifurcation variable. Additional control variables were considered in the experiments with the cusp model for buckling stress reported later in this chapter.

Levels of Performance

Some researchers have reported three stable states of work performance, rather than two, however (Guastello, 1981, 1987, 1995; Kanfer, 2011). The three states are (a) not good enough to meet the performance criterion, (b) good enough, and (c) a higher level of excellence that represents one's personal best in the matter. Three-state systems can result from substantial individual differences in ability, motivation, and differences in the surrounding organizational context that affect motivation within the organization. The motivation component requires two control parameters, one for extrinsic forms of motivation and one for intrinsic sources. Taken together the three states and four control parameters comprise a butterfly catastrophe model (Guastello, 1981, 1987, 1995). Motivation in this context is both positively and negatively valenced; an oversupply of negatively valenced experiences would be interpreted as "stressful" in the conventional sense by most people.

Three levels of functioning are often (but not always) associated with entire jobs rather than specific tasks, and usually require longer time horizons to unfold rather than what is usually the case in contemporary experimental analyses of fatigue. Vestiges of three states can be seen in Figure 9.6 although the lowest level of performance is not stable. In the experiment, that was (Guastello, Boeh, Gorin et al., 2013), 105 undergraduates completed the seven computer-based tasks seven times. They worked under one of four experimental conditions: tasks fully alternated, tasks aggregated with the multitask module performed

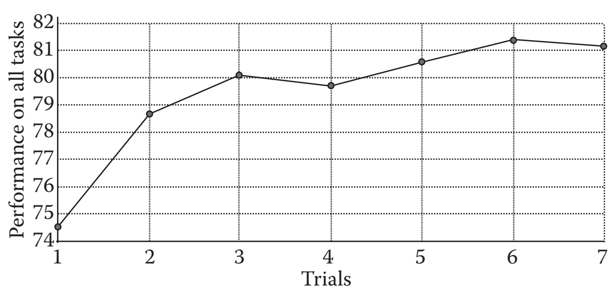

FIGURE 9.6
Performance on seven tasks across four experimental conditions. (Reprinted from Guastello, Boeh, Gorin et al., in press, with permission of the Society for Chaos Theory in Psychology & Life Sciences.)

first, tasks aggregated with the multitask module performed last, and where the participants chose the task order themselves. Figure 9.6 shows the performance trends over time for all seven tasks performed under all experimental conditions. The trend analysis showed that all of the visible bends in the curve were statistically significant. Performance improved up to Trial 3 (learning effect), was followed by a downturn (fatigue effect), followed by a further increase in performance through Trial 6, and a second downturn on Trial 7. The performance improvement from Trials 4 to 6 strongly suggested a strategy change of some sort. At present the range of possible strategies that a person could have deployed in this context is not known, but an analysis of the degrees of freedom associated with the tasks might uncover the answer, however. It was apparent, nonetheless, that some type of strategy shift was taking place fairly consistently.

Buckling Stress

Physical Demands

The development of the buckling stress model (Guastello, 1985, 1995) began with definitions of stress and strain as material scientists know them, which are a bit different from the psychological definitions, although similarities can be seen. *Stress* in material science is defined as the amount of load per unit area of a test piece of material, usually measured in pounds per square inch. There are a number of stress types, such as bending, pulling, and torque. Given a pulling stress, *strain* is the ratio of the change in length of the test piece to its original length, measured in inches per inch. The *modulus of elasticity* is the ratio of stress to strain. A plot of stress versus strain results in a curve that has both linear and nonlinear segments, with a peak known as *ultimate tensile strength.* In a materials fatigue test, stress is plotted against the number of stress cycles to failure, *N,* for a given level of stress, where a failure is a break in the test piece. The stress-*N* curve has an asymptotic lower limit of stress, called the *endurance limit*, which is linearly related to ultimate tensile strength and indicates the range of stress the test material can withstand before fracture. Some materials increase in strength whereas others decrease with successive stress cycles.

In Zeeman's (1977) application of catastrophe theory to beam buckling (Euler buckling), an elastic horizontal strut was given a vertical load and a horizontal compression. The amount of buckling is the deflection of the beam under the vertical load. The system was modeled as a cusp with deflection as the dependent measure (a difference score). Compression, which is the ratio of two times the modulus of elasticity times to length, functions as the bifurcation factor. Vertical load is the asymmetry factor. Buckling occurs at the cusp point, which is the most unstable point in the system.

Euler buckling is actually a group of related phenomena. The situation just described was one in which both ends of the horizontal beam were pinjointed. If both ends were fixed, buckling load is four times that of the pinjointed example. If one end is fixed and the other pinjointed and free to move sideways, the critical buckling load is one quarter of the value for the pinjointed example. In one of the original problems, the objective was to determine how long one could make a thin pole before it buckled under its own weight. On the other hand, beams that are too short fail by crushing rather than buckling (Gordon, 1978).

The same tests of strength, elasticity, and so forth that are made on building materials can be made on living samples of bone, kangaroo tendon, and butterfly wing. As the systemic level shifts from nonliving to human, the task is redefined from one of measuring the strength of various materials to measuring the strength of composites of materials. The composites can also think, which, in turn, necessitates subdividing the problem into primarily physical and primarily psychological forms of stress. The stress and performance problem considered next pertains primarily to the physical form.

The first application of Euler buckling to work performance required a close-up study (Guastello, 1985) of the wheelbarrow obstacle course data from the physical ability project that was mentioned in Chapter 8 (Guastello & McGee, 1987). In that simulation, subjects pushed a wheelbarrow over a 50-ft course, where the obstacles were traffic cones and barriers on the floor that were designed to simulate railroad tracks. The course was completed three times, first with a load of 100 lb (sandbags), then 200 and 300 lb. The criterion was timed to complete the course in seconds.

Subjects completed the 100-lb course with a model response time of 50 s. The response–time distributions were bimodal for both the 200- and 300-lb courses, with the lower modes at 40 s. The upper modes were comprised of all those who could not complete the course; their score was set equal to the longest completion time for the sample, which was 664 s.

The act of pushing a loaded wheelbarrow (strut) over a floor obstacle is more similar to the strut that is fixed at one end and free to rock sideways at the other. The worker exerted the compression force; the wheelbarrow tipped when buckling occurred, dumping sandbags on the floor. The worker then had to set the wheelbarrow and its contents upright and continue if possible, thus increasing the performance–time score.

Performance–time differences in the human model replaced strut deflection in the mechanical, which shows the cubic potential denoting critical change. The asymmetry term was vertical load as in the Zeeman (1977) model. The modulus of elasticity/height ratio (bifurcation factor) was represented by several experimental variables: body fat (as measured by calipers), exercise habits (number of hours of exercise per week outside of work), height, body balance (rail walk test), and gender (the sample of participants contained 79 men and 50 women). The cusp catastrophe model for buckling in the application to the wheelbarrow obstacle course appears in Figure 9.7.

Cognitive Demands

Of course, the specific physical qualities that signified elasticity in a wheelbarrow obstacle course might be different for other physical tasks. The control parameters might be very different again, at least by outward appearances, when mental workload problems are

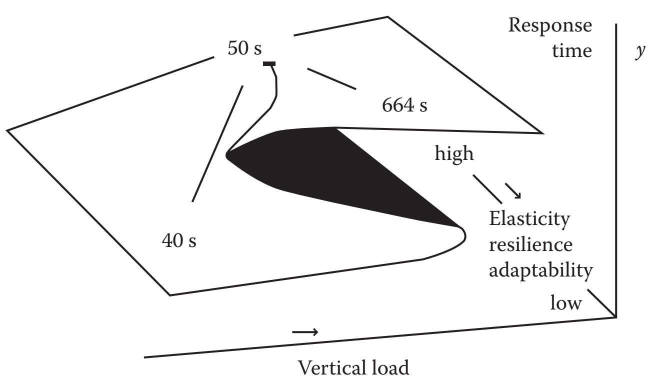

FIGURE 9.7
Cusp catastrophe model for buckling stress showing the effects of load and elasticity on a physical labor task.

considered. According to Thompson (2010), resilience consists of both cognitive and emotional elements. Cognitive resilience includes preplanning, rehearsing, focus, dividing tasks, prioritizing, and practice. Emotional resilience would include the ability to recognize one's own emotions and the emotions of others and act appropriately. Furthermore, demands from an onerous physical environment can add to the demands from the cognitive task.

The model shown in Figure 2.9 depicts the broader range of cognitive workload reactions where resilience is comparable in meaning to elasticity in physical applications. Indifference to cognitive workload is actually a response of a stable system that has a sufficient capacity for withstanding peaks in demand. Beyond a critical point, however, performance breaks down suddenly, in which case we have a different stable state that is undesirable. The person or system could make an adaptive response at this point, however, to preserve the level of performance (low response time or error rate). Too much flexibility, however, can be costly to the system and can actually destabilize it; this point is elaborated on further in this chapter. As Figure 2.9 shows, the resilience portion of the system is actually located in the part of the response surface that does not contain the stable states. It contains the cusp point instead, which is the most unstable point in the system.

Diathesis Stress

Diathesis stress models attempt to explain why the effects of stress take the form of one somatic symptom in one person but a different symptom in another. The explanation is that there is an underlying vulnerability in a system, or a hierarchy of system strengths and vulnerabilities, and the consequence of stress affects the weakest element in the system. Thus, a weakness in the stomach area could predispose a person to ulcers, but a cardiac weakness in someone else would translate into a different symptom.

Shiftwork

Figure 9.8 depicts a diathesis stress model for the effect of shift work as a stressor on the effectiveness of a manufacturing organization (Guastello, 1982). The organization was a precision printing organization in which the workforce worked rotating shifts; this is the

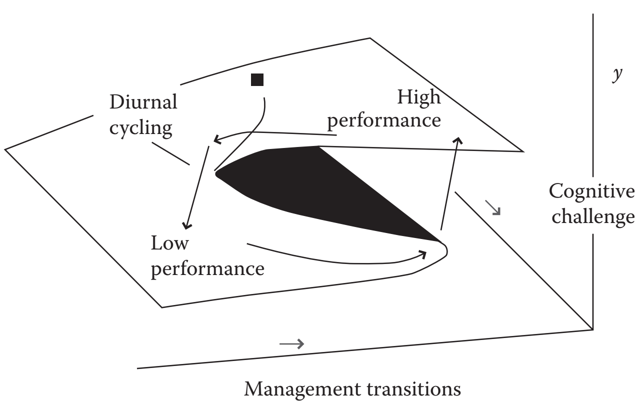

FIGURE 9.8
Cusp catastrophe model for diathesis stress showing the exacerbating effects of cognitive demand and management-induced stress.

underlying source of vulnerability. The stressors took two forms. One was a cognitive demand as the bifurcation parameter. Some of the printing jobs (woodgrains utilizing four to eight colors in a rotogravure process) required relatively short print runs, and some required very long press runs. The printing cylinders wore down during the longer print runs and the ink flow changed by subliminal amounts over short periods. Changes in ink colors that resulted from the cylinder wear produced color-matching difficulties and considerable waste of materials and labor in the process. The other source of stress came from management influences. The asymmetry parameter in Figure 9.8 represented some management personnel and relationships over a 13-week period, which were ultimately for the better.

The dependent measures were changes in material waste and machine–labor time for two versions of the same print run done on two different shifts, for example, morning versus afternoon or afternoon versus night. For short runs, which were less demanding, there were small fluctuations across shifts, but the more demanding runs produced larger fluctuations. In this situation, the morning and night shifts were equivalent in performance, but the drop in productive efficiency occurred in the afternoon (second shift) when it did occur.

During times of negative management conditions, the flux in productivity was anchored at the low efficiency mode of the cusp response surface. During the last weeks of the data series that involve the positive management conditions, the flux in productivity was anchored at the high efficiency mode of the cusp response surface. The largest fluctuations in efficiency occurred during the weeks between the two endpoints. The management solution to the production efficiency problem was to retain the new operations management procedures and to place a limit on the length of a print run so that they could re-engrave the cylinders before the color-matching crisis was likely to occur. Their customers were given price incentives to place their job orders in batches of workable size.

The diathesis stress principle generalizes a bit further. It basically consists of an ongoing source of variability in performance, which would be the fluctuations induced by circadian asynchrony. The workforce was sufficiently practiced at keeping the impact of those fluctuations to a minimum, but the effect loomed in the background of the system's activities waiting to be activated by new stimuli. The new stimuli consisted of a background variable and a trigger variable. The background variable was the management situation, which went from not so good to something much better; this was the asymmetry variable in the model. The trigger variable was the cognitive demand of the task, which could be traced to a quantifiable feature of the task, and for which the machine system was partially responsible; this was the bifurcation variable in the model.

Occupational Health

The occupational health effects of stress mentioned earlier are another example of diathesis stress dynamics (Guastello, 2013). The concept of diathesis stress dynamics actually originated in behavioral medicine, which by itself falls outside the scope of human factors. Nonetheless, the analysis of medical disorders over a 3-year period of occupational exposure showed that the cusp catastrophe model was a statistically superior explanation for the medical outcomes compared with linear interpretations.

Fatigue

Early theory (Ash, 1914; Starch & Ash, 1917) suggested that physical and mental fatigue processes are similar; as a result, the study of one becomes a contribution to the study of another. This section details the central concepts pertaining to fatigue curves, which are mostly related to physical work, the results of the earliest mental fatigue studies, and contemporary views on the topic.

Fatigue is operationally defined as a decrement in work capacity over time; it is often inferred by decrements in performance on physical tasks (Kroll, 1981; Starch & Ash, 1917; Weinland, 1927). The *work curve* is a graphic plot of performance over time; it was first introduced by Mosso (1894, 1915). The work curve can be decomposed into two parts: the curve itself and deviations from the curve. The curve denotes a loss of work capacity and decrement of performance caused by fatigue (see Figure 9.9). The variability from the curve increases dramatically toward the end of the work period and denotes loss of control over one's neuromuscular system (Ash, 1914; Starch & Ash, 1917; Weinland, 1927). Seven types of work curve were identified, those that (a) dropped gradually, (b) dropped suddenly, (c) maintained stable but high performance, (d) maintained stable but low levels of performance, (e) maintained stable levels of performance prior to, and sometimes after, the output drop, (f) were highly unstable, or (g) increased suddenly (Crawley, 1926; Ioteyko, 1920; Marks, 1935).

Physical Fatigue

Some participants in the physical fatigue experiments showed an increase in strength or work capacity after having been exposed to a physical work set that would have fatigued other participants. Two explanations emerge from the literature: exercise and scalloping. Crawley (1926) found that subjects increased work output in a subsequent test session if the same muscle groups were utilized in a previous session. More generally, exercise is known to increase work capacity by an additional 50% compared with preexisting levels (McCormick, 1976). Crawley also found that individuals who knew their work period would end expended their greatest effort just before the close of the session. Subjects who were not informed exhibited a more typical work curve.

Ioteyko (1920) determined a general formula for the work curve that satisfied all varieties of work curve that were known at the time. After a few other interesting studies that intervened over the ensuing 67 years, Guastello and McGee (1987) showed that the effects of fatigue on physically demanding work performance could be modeled as a cusp

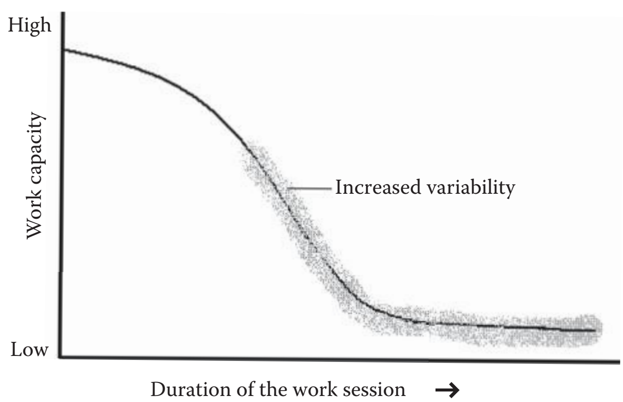

FIGURE 9.9
The work curve for a physical task showing a decline in work capacity and greater variability in performance as fatigue sets in.

catastrophe, and that Ioteyko's formula was very close to the cusp catastrophe model, even though it predated catastrophe theory by a half century.

The cusp catastrophe model for change in arm strength in Figure 9.10 accurately accounted for the actual data and more so than linear comparison models that were based on the same qualitative variables. Several bifurcation variables were found in the study, some of which had negative regression weights and some had positive weights. Bifurcation variables with negative weights in the model, indicating that the variables promoted fatigue, were the amount of work done, greater amounts of exercise outside of work, labor experience, and higher body fat percentage. Bifurcation variables with positive weights in the model were weight and motivation (subjectively assessed by the experimenters based on the participants' demeanor). Leg strength emerged as a strength compensation variable.

Cognitive Fatigue

The early fatigue researchers reported that cognitive fatigue follows the same essential temporal dynamics as physical fatigue. Work capacity declines after successful intervals of demanding mental work, and recovers after a rest period (Ash, 1914; Starch & Ash, 1917). Another important feature of the early studies, however, was that two human performance phenomena were taking place instead of one. Although the experiments were designed to reduce work capacity over time, which would induce an increase in RT, repeated testing would induce a practice effect, which would have the impact of lowering RT.

The studies of cognitive fatigue were sparse after the Ash (1914) experiments. Two reviews of research surfaced in the early 1980s (Holding, 1983; Welford, 1980). Both reviewers reported that the effects of fatigue on performance and stress were often obscured in the way researchers defined problems and experiments. Some laboratory research, nonetheless, distinguished some interesting phenomena. Bills (1931) reported a *blocking effect* in cognitive fatigue experiments (as in "mental block") where excessive reaction times would occur after prolonged work on an aperiodic basis. Blocking effects became typically defined as an instance of RT greater than two times the mean RT (Welford, 1980). The blocking effect was thought to occur as a means of dissipating the effect of fatigue. There are two critical aspects of the blocking effect that signify fatigue; one is the onset of the block, and the other is the increased number of blocks that appear as the experiment continues.

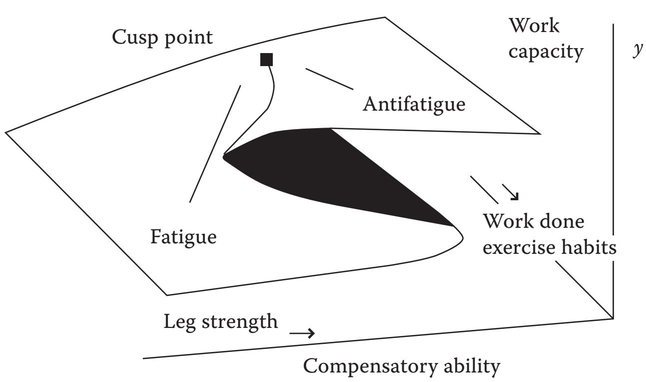

FIGURE 9.10
Cusp catastrophe model for arm strength fatigue after 2 hr of physically demanding work.

Although some studies showed that the time needed to recover from cognitive fatigue is less than the time needed to recover from physical fatigue, recovery time could be dependent on the specific examples of mental and physical tasks used in those experiments (Welford, 1980). In some situations, recovery requires a prolonged recovery time, and is not qualitatively different from the aftereffects of stress phenomena discussed earlier, according to Holding (1983). Aftereffects of fatigue often do not transfer to subsequent tasks, and may be related to the repetitiveness of the task and level of arousal or boredom. Perceptual fatigue, according to Holding's interpretation, is largely related to arousal, and can be manipulated by reward structures. Underarousal, on the other hand, is typically regarded as a stress phenomenon, as previously discussed in the context of the Yerkes–Dodson law Hancock and Warm (1989).

Experiments on cognitive task performance after prolonged work periods did not consistently result in a decline in performance for all people who were subjected to the same or similar experimental conditions (Dureman & Boden, 1972; Holding, 1983; Poffenberger, 1928). For instance, Poffenberger reported that people who performed prolonged mental work reported feelings of tiredness in much the same way as they do after having performed extensive physical work. Performance on cognitive tasks, however, may remain stable, or improve after prolonged work, despite reports of tiredness. Again, practice or automatization effects could counteract other reasons for declines in performance. The clear differences in cognitive work curves produced by the same stimuli and experimental conditions for different people may be explained by a more inclusive equation and theory in much the same way as Ioteyko (1920), Crawley (1926), and Guastello and McGee (1987) developed.

Two other trends in cognitive fatigue research suggest that nonlinear dynamic processes could be involved. First, the physical work curves show greater variability from the curve with prolonged work, and similarly, RT tends to become more erratic after prolonged mental work (Holding, 1983). Second, there is a provocative theory suggesting that at least some forms of fatigue are the result of increased noise in the neural system, rather than the failure of the neural system to transmit signals. The variability from the work curve over time, particularly blocking phenomena, and the neural noise concept, strongly suggest that chaotic processes are operating (Guastello, 1995), although this particular proposition has only received sproatic attention (Clayton & Frey, 1997; Hong, 2010).

Figure 9.11 defines a cusp catastrophe model for cognitive fatigue (Guastello, 1995) showing two stable states of behavior, which were RT under conditions of fatigue, and RT under conditions of practice. The perception task involved the reversal of ambiguous figures from Ash's (1914) data. In principle, RT in both instances gravitates toward asymptotic values, or attractors, both of which are known to exist on the basis of the learning curve and some varieties of fatigue curve that were explained earlier. The bifurcation parameter is the number of units of work performed by the human subject. A greater change in RT is expected as more opportunities for measuring the RT to the figure-reversal task accrue.

The cognitive fatigue data from Ash's (1914) experiments involved participants who were exposed to more than one version of the experiment. In other words, they were not naive subjects when a lot of their data were generated, and they accumulated practice from their previous participation. Confounding effects of this type may have rendered those early experiments only marginally interpretable from the vantage point of conventional experimental methods. The data lent themselves to a cusp catastrophe analysis of the experimental data, however, which could separate the fatigue and practice effects. In

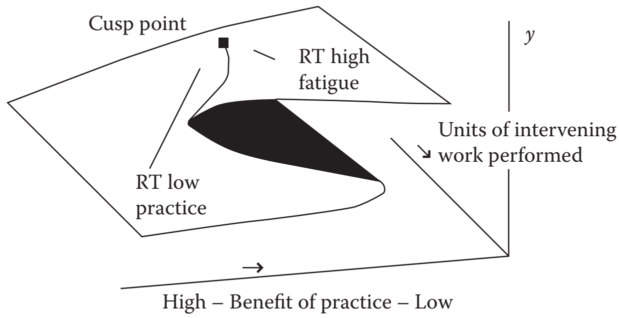

FIGURE 9.11
Cusp catastrophe model for mental fatigue showing the effects of work volume and practice.

so doing, it was possible to show that the dynamics of mental fatigue are closely analogous to the dynamics of physical fatigue.

The model for cognitive fatigue is a generalization of the control parameters found in the model for physical fatigue. First, the total quantity of work done would be the main contributor to the bifurcation parameter: If the individual did not accomplish much in a fixed amount of time, there would be comparably little drain on work capacity. Those who did could exhibit either positive or negative changes in work capacity.

The asymmetry parameter would be one or more compensatory strength measures. For instance, in Guastello and McGee (1987), laborers displayed differences in arm strength as a result of about two hours worth of standard mill labor tasks, which primarily demanded arm strength. Workers were measured on isometric arm strength and leg strength before and after the work session. Leg strength showed little change after the work session, which was not surprising, but it did act as a compensation factor for arm strength; those with greater leg strength experienced less fatigue in their arms.

A similar type of compensatory function is thought to be operating in cognitive fatigue dynamics. Here it is necessary to assume a model of intelligence. Cattell's (1971) hierarchical model of intelligence, which was described in Chapter 6 is perhaps the most viable candidate because of its consistency with recent experimental findings concerning working memory. Working memory is nested under fluid intelligence. It too is hierarchical with several domain-specific modules (e.g., numerical, verbal, and spatial domains) that are hierarchically organized with a modicum of executive control over attention (Conway et al., 2005; Logie, 2011). Less executive control is involved to the extent that the processes are automatically executed. More control is involved in processes that involve applying the cognitive functions a little differently for each time an example of the task is encountered (Ackerman, 1987). Executive control functions also focus attention, more so under conditions of high demand or other forms of stress. It is thought that cognitive fatigue occurs when the executive function is being overly taxed (van der Linden, 2011). It is possible that memory modules are recruited as needed when demands increase or change (Drag & Bieliauskas, 2010; Logie, 2011; Reuter-Lorenz & Cappell, 2008; Schneider-Garces et al., 2009). Functions that might be central to one task could be compensatory to another. A compensation factor, therefore, would derive from the recruitment of working memory capacity from operational areas not directly associated with the task. In other words, compensation factors are somehow adjacent to the primary area performing the task.

Cusp Models for Cognitive Workload and Fatigue

Current research on cognitive workload and fatigue still has several unresolved problems that were addressed in the four studies that are described next. (a) It is very difficult to separate the effects of workload and fatigue using conventional experimental paradigms (Ackerman, 2011; Guastello, Boeh, Shumaker, & Schimmels, 2012; Hancock & Desmond, 2001). (b) The upper boundaries of cognitive channel capacity are sometimes fixed and sometimes variable, and the bottleneck phenomenon often appears when multitasking is involved. (c) It has not been forthcoming as to whether fatigue is really produced by the total amount of time working or by time on a specific task. (d) Cognitive underload can also produce fatigue. (e) Prolonged time on task can produce the performance decrements resulting from fatigue, but it can also produce improvements in task performance or work capacity that would be associated with practice and automaticity. (f) It is sometimes possible to alleviate fatigue by switching from a low-demand task to a higher-demand task or vice versa (Alves & Kelsey, 2010; Lorist & Faber, 2011), but switching tasks could be mentally costly, as described in Chapter 6. Cognitive strategies of an uncertain nature are involved and might change during a work period to keep performance up to the desired level.

The four studies applied the cusp catastrophe models to the analysis of tasks performance in laboratory tasks. By constructing an experimental design to assess both phenomena, it was possible to separate the two effects. In addition to addressing the six dilemmas mentioned above, the research program also engaged a search for situational and individual differences or variables that correspond to the resilience constructs in the workload model and the compensatory ability factors in the fatigue model. Another part of the research agenda was directed toward capturing as many fatigue and workload phenomena within the framework of either cusp model or other NDS processes that were thought to give rise to the cusp effects when they were observed.

The evaluation of a cusp model's presence in the real data can be done with some adaptations of some well-recognized statistical procedures. The cusps involve a change over time as the dependent measure, and the elements of the model: the cubic polynomial term, a cross-product term for bifurcation, and a simpler linear term for asymmetry. The cusp is then compared with an alternative model that could be used to assess the problem, which is most often a linear model containing all the variables that are expected to contribute to bifurcation and asymmetry. The researcher would be looking for a reasonably high R^2 for the cusp, which should be at least as large at the R^2 for the alternative model. All the parts of the model should be present, as determined by statistical significance on the regression weights. Multiple variables can be tested as bifurcation and asymmetry control variables (Guastello, 2011). The foregoing criteria may require somewhat of a trade-off in the event that researchers do not find everything they could be looking for in a given instance.

Episodic Memory

The first experiment of the series focused on episodic memory (Guastello, Boeh, Shumaker, & Shimmels, 2012). The participants were 181 undergraduates. Prior to performing the main task, the participants completed timed 5-min tests of arithmetic and spelling abilities and a brief questionnaire measuring trait anxiety. The main task, which was a computer-based game, required the participant to view a sequence of visual stimuli that had an auditory accompaniment and repeat the sequence by clicking a mouse on a graphic display. The sequence started with one stimulus, and the sequence got progressively longer and faster with each successive trial. This task was performed for 20 min. The change in behavior that was used as a depending measure was the score (total number of correct stimuli in an unbroken repetition of the displayed sequence) collected during the first 3 min of work and the last 3 min.

The longest sequence of correct responses that a participant was able to achieve was tested as the asymmetry variable in the buckling model for workload (Figure 9.12). Anxiety was tested as a bifurcation variable in the workload model. The results supported the presence of the cusp structure overall and the longest sequence of correct responses as the asymmetry variable for vertical load. Anxiety did not work as a bifurcation variable here. The key results of this analysis appear in Table 9.1 along with comparative results of the other cusp models that are described next.

The fatigue model also assessed differences in performance between the first and last 3-min period. The amount of work done was measured as the total number of points achieved during the middle 14 min; this measurement was tested as the bifurcation variable. Arithmetic and spelling scores were tested as compensatory abilities that comprised the asymmetry parameter. The longest sequence of correct responses, which represented working memory capacity, was tested as another compensatory ability. The results supported the presence of the cusp structure overall and the hypothesis concerning the bifurcation variable. Arithmetic and working memory capacity both functioned as compensatory abilities.

Although the maximum score on an unbroken run of correct responses worked as an asymmetry variable in both the workload and fatigue models, the meaning of the variable was different in each context. In the workload model it represented the largest and fastest load that the individual could endure or apply. In the fatigue model, it represented working memory capacity as a compensatory ability. Granted it was a more direct ability than arithmetic ability; the latter was closer to the intended meaning of "compensatory." This particular element of ambiguity between the two models was fortunately not occurring in the following studies where workload and fatigue dynamics were also considered simultaneously.

Pictorial Memory

The second experiment of the series focused on pictorial memory (Guastello, Boeh, Schimmels et al., 2012). It also provided an opportunity to determine whether cognitive fatigue was more closely related to total time spent working or time spent on a particular task. The participants were 130 undergraduates. The tasks were organized in two experimental sessions, the first of which went as follows: Prior to performing the main task,

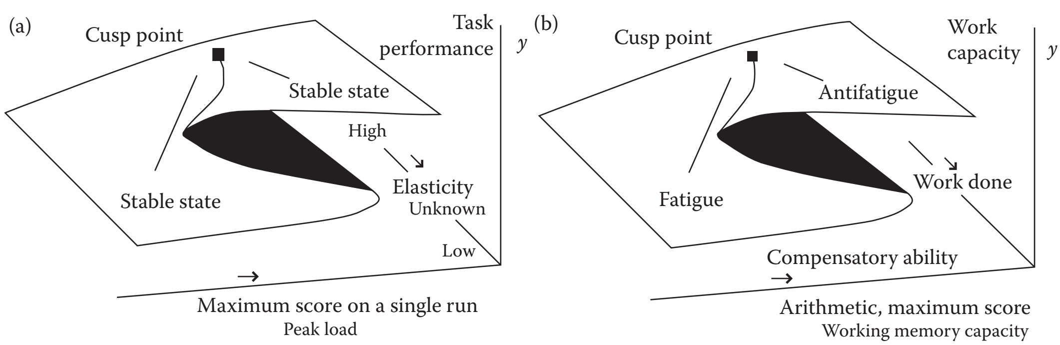

FIGURE 9.12
Buckling model for (a) cognitive workload and (b) fatigue in the episodic memory task.

TABLE 9.1

Summary of Results for Cusp Models for Cognitive Workload and Fatigue

		R^2			
		Cusp	**Linear**	**Bifurcation**	**Asymmetry**
Workload	Episodic memory	.44	.32	Unknown	Peak load
Fatigue	Episodic memory	.53	.50	Intervening work	Arithmetic peak load
Workload	Pictorial memory	.53	.13	Anxiety	Incentive condition
Fatigue	Episodic memory	.30	.16	Unknown	Unknown
	Episodic peak	.39	.07	Unknown	Unknown
	Pictorial memory	.52	.59	Intervening work	Episodic peak
Workload	Multitask at Time 1	.49	.35	Unknown	Task difficulty
	Multitask at Time 2	.75	.18	Self-determined task order	Unknown
Fatigue	Accuracy	.47	.56	Fully alternating order Self-determined task order	Spelling
	Memory	.53	.33	Self-determined task order Intervening work	Unknown
	Perception	.44	.20	Intervening work	Unknown
	Coordination*	.43	.53	Intervening work	Unknown
	Multitask*	.35	.52	Unknown	Unknown
	Reaction*	.35	.61	Unknown	Unknown
	Speed*	.39	.46	Unknown	Spelling
Workload	Security cam vigilance	.33	.29	Frustration	Load increase
Fatigue	Security cam vigilance	.43	.17	Puzzle completed	Slow speed first
Summary	Average all models above	.45	.35		
	Average all except *	.47	.30		

the participants completed the same measures of arithmetic, spelling, and anxiety, and a 5-min version of the episodic memory task used in the previous study. The main task was a memory game consisting of two decks of cards. One deck was spread out on the table and consisted of complex pictures. The other deck contained riddles that the participant read with the objective of matching four picture cards with the riddle. The score was the number of riddles solved during a period. The total work time on this task was 50 min, and the task was followed by another 5-min version of episodic memory task.

For fatigue in the main task, the dependent measure was the difference in scores between the first and last 10-min interval. The amount of work done was counted as the number of riddles solved during the central 30-min period. The results supported the presence of the fatigue model, work done as the bifurcation variable, and the peak correct sequence on the first episodic memory task as the compensatory ability to fatigue.

Fatigue in the episodic memory task was assessed for both changes in the total scores and longest correct sequences before and after the main task. The bifurcation variable was the total score from 50 min on the intervening main task; the same compensatory abilities were tested. For both these criteria, the results supported the presence of the cusp structure, but not as strongly (R^2 criterion) as it was found in the main task. None of the hypotheses for bifurcation or asymmetry variables were supported by significant regression weights. The conclusion was the fatigue resulted from time on a particular task and not necessarily from total time working.

There were significant differences on the two episodic memory measures between the two measurement times that were also of interest. The number of points scored increased from the first to the second observation time, but the peak sequence length decreased. This pair of observations suggested that the participants were making a strategy shift whereby they collected points from shorter sequences and thus placed less demand on their working memories.

Cognitive workload effects were assessed in the second session of the experiment when the participants performed the pictorial memory task competitively with another participant who had also been in the first session. The load manipulation was the difference in performance between working alone in the first session and performance when working with the competitor. Two variables were manipulated: Participants were either informed they would receive bonus credits for actually winning the game against the competitor or they were not so informed. Participants were either given some modest time pressure by being informed that they needed to complete their turn within 90 s, while others were not so informed. Anxiety was tested as the bifurcation variable, with the expectation that anxiety would be more influential in competitive situations than when working alone as in the study on episodic memory. The incentive condition and time pressure condition were dichotomous variables that were tested as potential vertical load variables.

The results supported the presence of the cusp in the workload data. Anxiety did have the hypothesized effect as a bifurcation variable, and the incentive condition worked as the asymmetry variable for vertical load. Time pressure did not have a unique effect here, but an interaction effect was uncovered such that time pressure had a negative effect on task performance for people working in the incentive condition, but a positive effect for people working in the no-incentive condition. This relationship is shown in Figure 9.13.

Multitasking and Ordering of Tasks

One goal of the third study (Guastello, Boeh, Gorin et al., 2013) was to investigate cognitive workload and fatigue outcomes for multitasking compared with single tasks. The second goal was to assess the impact of task ordering, particularly on the multitask. Participants were 105 undergraduates who once again completed measurements of arithmetic, spelling, spatial visualization, and anxiety prior to the main task. The main task was actually a set of seven gaming tasks that tapped different combinations of perceptual and motor ability:

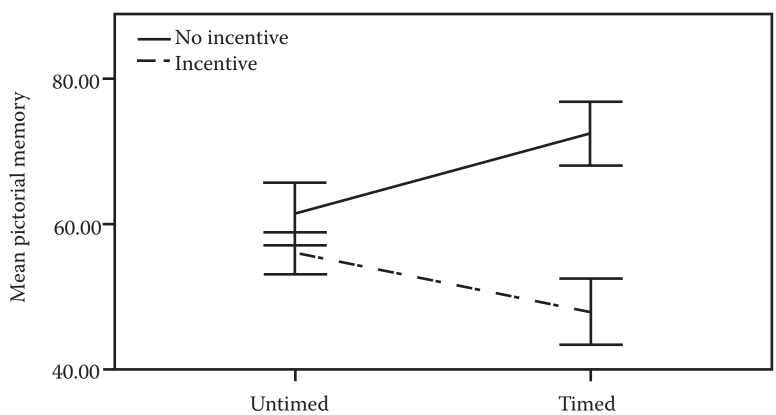

FIGURE 9.13
Interaction between time pressure and incentive condition on performance in the pictorial memory task. (Reprinted from Guastello, Boeh, Schimmels et al., in press, with permission of the Human Factors and Ergonomics Society.)

accuracy, coordination, memory, the multitask, perception, reaction time, and speed of response. The multitask required two responses: (a) selecting the one of four spatial stimuli that matched a target shown on the screen and (b) making a single-digit addition or subtraction calculation. Both answers needed to be entered by computer mouse within 2 s to produce a correct response. Each task was performed seven times, totaling 50 min. There were four experimental conditions: (a) the seven tasks were completed in completely alternating sequence; (b) all examples of one task were completed before moving on to the next task, and the multitask was done first; (c) same as previous but the multitask was done last; (d) participants could do the tasks in any order they chose so long as they did seven examples of each task.

Results supported the effect of task ordering on the multitask, such that allowing participants to choose their own order produced better performance than any of the other three orderings. There was also an interesting result for performance over time. All seven tasks showed the same temporal trend as shown in Figure 9.5. All bends in the curve were statistically significant. There appeared to be a learning effect at first, followed by a fatigue effect, followed by an apparent shift in cognitive strategy at Time 4, further improvement, then another performance decline.

A cusp model for fatigue was tested for each of the seven tasks. The performance difference was between the first and seventh rendition of the task. The amount of intervening work accomplished was the score on the middle five renditions of the task. The four experimental conditions were treated as three dichotomous variables and tested as bifurcation variables also with the reasoning that some conditions might produce more fatigue or faster learning, or both, than others. Spelling, arithmetic, and spatial visualization were again tested as compensatory abilities. Results supported the cusp model for two out of seven tasks, memory and perception, which in retrospect were the two tasks that seemed to place the greatest demand on working memory. This finding lent credence to the speculation (Logie, 2011) that cognitive fatigue resulted from the demands on working memory; this point still needs to be explored further. The intervening amount of work accomplished was supported as the bifurcation factor for both the memory and perception tasks. The self-determined task ordering condition was a second bifurcation variable in the fatigue model for memory. Significant effects were not found for the compensatory ability variables, however.

The cusp model for fatigue in the accuracy task was relatively strong, but not as accurate a fit to the data as the linear alternative. There were two bifurcation variables, the fully alternating condition and the self-determined condition, and spelling ability was the compensatory ability variable.

The cusp model for workload was centered around the multitask and was tested in two forms. In one form, the total score on the multitask was treated as the Time 1 measurement of performance, and one of the other single tasks was treated as the Time 2 measurement. For the Time 2 measurement, participants were randomly assigned to one of the six other tasks. The six tasks were rank-ordered according to the average percent accuracy attained by the sample, thereby producing a difficulty score ranging from 1 to 6. The difficulty score was hypothesized as the vertical load variable, and the four task ordering conditions were tested as bifurcation variables. Results supported the cusp structure and the role of the difficulty variable, but a bifurcation effect was not isolated.

For the second form of the cusp model for workload, performance on the randomly assigned task was treated as the Time 1 measurement of performance and the multitask as the Time 2 measurement. The cusp structure was noticeably stronger for this version of the model. The self-determined task order was supported as the bifurcation variable, but the difficulty variable was not supported as the asymmetry variable.

Vigilance Dual Task

One objective of the fourth study (Guastello, Malon et al., 2012) was to approach the multitask aspect of cognitive load from the dual task perspective. The experimental situation was designed to be analogous to the work of a hospital nurse who monitors security cameras in several patients' rooms while performing other tasks. The vigilance task, which was the task of primary concern, was composed of a sequence of virtual reality images from a security camera that was monitoring a building when no one was supposed to be in any of seven different rooms. An intruder, who was designed to look threatening, appeared on a random basis in 10% of the frames. When the participants spotted an intruder, they rang a desk bell. The experimenter had a script of the timings for the appearance of intruders and marked whether the participant produced a correct hit, a miss, or a false alarm. Because false alarms were relatively rare compared with missed errors, only the missed errors were analyzed.

For the secondary task, participants completed a 300-piece jigsaw puzzle. The puzzle task contributed one score, which was the total number of pieces assembled at the end of the 90-min session.

There were three experimental manipulations. One was the speed of presentation of the security camera frames; there were three speeds comprising a repeated measure. The second independent variable was whether the participant experienced changes in speed from slow to fast or from fast to slow. The third independent variable was whether the participant worked on both tasks alone or in pairs. The study on pictorial memory showed what could happen when two people were working competitively; in this study, the second goal was to observe what would happen when two people worked cooperatively, or at least had a chance to do so.

The third goal of the study was to expand the repertoire of variables that could function as resilience variables in the workload model. The new candidates were conscientiousness, work ethic, emotional intelligence, and the frustration scale from the NASA TLX. Anxiety was included again as a potential elasticity variable in the buckling process. Arithmetic and spelling tests were included again as potential compensatory abilities in the fatigue process.

Conscientiousness is a personality trait whereby someone with a high score would be attentive to details in their work and daily life, adherent to rules and regulations, and would do their best to complete a job properly (Cattell, Eber, & Tatsuoka, 1970). Work ethic is a work value, or set of beliefs about work, that emphasizes independent action and an obligation to work (Buchholz, 1977). A person who endorses the work ethic would agree with statements that emphasize independent action and obligation to work such as, "Work is good regardless of how hard or boring it is" and "A person should avoid dependence on others."

Salovey and Mayer (1990) introduced the concept of *emotional intelligence* (EI) as "a type of emotional information processing that includes accurate appraisal of emotions in oneself and others, appropriate expression of emotion, and adaptive regulation in such a way as to enhance living" (Mayer, 2001, p. 9). Mayer and Salovey (1997) refined the definition to reflect more strongly that EI was the ability "to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth" (Mayer, 2001, p. 10). The two definitions promoted different measurement models, some of which emphasized the cognitive aspect of EI, while others emphasized personality aspects of EI. The measurement model adopted in Guastello, Malon et al. (2012) was the version by Schutte et al. (1998) because it captured the theme of *alexithymia*, the inability to interpret one's emotions and having no words to express one's emotions; this construct was influential in the development of the construct of EI and was conveniently brief for a study that involved measuring a number of variables and a core task that was meant to be time consuming by itself.

The cusp model for cognitive workload was supported with frustration as the bifurcation variable. Frustration obviated any additional effect from EI, work ethic, or conscientiousness. People who reported being frustrated by the two-task assignment had either a high or low error rate compared with less frustrated people. Sometimes, a task produced frustration for people who do turn out to be successful at the task nonetheless and for those who are not so successful. The asymmetry variable was *load direction*, which was the change in miss rates on the vigilance task between slow to medium speed, slow to fast, or the reverse conditions. In essence, the results were consistent with the theory, if we consider frustration as simply having the opposite effect of EI.

The fatigue model was supported with the amount of work done on the puzzle (number of pieces assembled) as the bifurcation variable. The asymmetry variable was whether the participant started with the fast condition first or the slow condition first. The increase in misses was greater if they started with the slow condition. Instead of an ability variable working here, the operating variable denoted a situation where the participants developed two strategies for managing the two tasks with performing the slow condition before the fast condition being less advantageous for fatigue.

Conscientiousness, work ethic, and the EI measure did not contribute to the cusp model as expected, but they did play modest roles in the linear models that were tested. Participants scoring high on conscientiousness made fewer miss errors on the vigilance task overall and showed smaller changes in their error rates when the rate of target stimuli increased. Participants scoring high on work ethic actually made *more* miss errors on the vigilance task, completed *less* of the puzzle, and showed more performance loss under fatigue; the reasons for this counterintuitive effect are not clear. Participants scoring high on the EI measure showed a slight tendency toward more errors on the vigilance task; again, the results were counterintuitive and warrant further investigation.

Summary of Cusp Models

Table 9.1 contains a summary of the results from the four studies where the cusp models for cognitive workload and fatigue were used jointly and lists the effect sizes and variables that made significant contributions to the asymmetry and bifurcation parameters. If all the 17 cusp models listed in the table are included, the cusp accounts for an unweighted average of 5% more variance than the best alternative linear models. If the four models marked with the asterisk are eliminated because the researchers concluded the cusp effect was not apparent to any meaningful extent, those 13 models that did work accounted for 17% more variance than the alternative model, or a ratio of 1.56:1. This is a reasonable increment of improvement in prediction utility given the small number of variables that are involved in the model. Nonetheless, research continues the search for better variables for each task situation and stronger manipulations of workload and fatigue in the laboratory and elsewhere.

Degrees of Freedom

The principle of degrees of freedom was introduced in Chapter 6 along with its plausible role in cognitive workload. The degrees of freedom principle underlies fatigue as well. According to Hong (2010), the increase in performance variability that is observed during the low-production phase of the work curve suggests an internal search for a possible reconfiguration of degrees of freedom. There are two plausible directions for such a search. According to the *redistribution principle*, the individual is searching for a lower-entropy means of accomplishing the same task or goal. If a *loss of total entropy* was occurring, however, the individual would not only be trying to regroup internal resources, but also reducing the need to respond to the total complexity of task situation and gravitating instead toward easier task options, simplification of the tasks, other ways of "cutting corners" or just stopping the task. It is a question for further investigation to determine the circumstances in which each principle might be engaged.

DISCUSSION QUESTIONS

1. How would you design a stress management program based on the occupational stress effects described in this chapter?
2. Describe a strategy that you would use to evaluate some choices in nonstandard work schedules. What criteria would you use in an empirical study? (Hint: How is this question similar to, or different from, evaluating a human–machine system as discussed in previous chapters?)
3. What new information about the control of stress effects comes uniquely from the nonlinear cusp models?
4. What does the set of cusp models for stress and work performance tell us about the nature of stress?
5. How would you evaluate mental fatigue in modern technological work?
6. What are some other cognitive strategies a person might use to buffer the effects of a challenging mental workload?
7. For all the public health messages that continue to circulate telling people what they should eat or not eat, do or not do, or the help they can get to minimize health disorders, have any of those messages tried to cure people of "workaholism"?

10

Occupational Accidents and Prevention

The study of accident analysis and prevention did not cross into the area of human factors engineering or ergonomics until the mid-1980s when it became clear that there were substantial psychological contributions involved that explained many connections among system design, social processes, and human error leading to the unwanted results. In earlier times, the role of human error was prominent in the human factors framework, and it was generally assumed that error rates would lead to accidents and injuries eventually. Today, the role of other variables in the system are better understood. This is also an area of human factors and ergonomics where management plays perhaps its greatest role.

Although there is a great deal of generalizability from occupational situations to traffic, home, and public situations, there are differences as well. Occupational environments are a bit easier to study for some types of questions because, unlike traffic situations, the principal actors stay in a contained environment. Things in traffic tend to move out of the range of immediate direct observation!

This chapter begins with an overview of the leading causes of death and the placement of accidental death in the hierarchy of leading causes. It continues with risk analysis models, the cusp catastrophe models for accidents, the nexus of psychosocial variables that are involved; group dynamics, safety management and culture, and resilience; an evaluation of accident-prevention techniques, and emergency response. The specific case of traffic accidents is deferred to Chapter 14.

Causes of Death and Injury

The word *cause* is a treacherous word because it requires a human evaluation of the outcomes produced by a complex system. Much of the information on causes of death, especially from past decades, is only available from death certificates, and the designation *of cause* on certificates is inevitably predicated on the status of medical knowledge available at the time of death. The task of defining a cause of death requires a gross simplification of the process leading up to many *proximal* causes of death. Thus, a bit of skepticism is warranted when interpreting alleged causes that were recorded at any one time in history.

Social psychologists have specifically studied the attribution of cause, which is the process by which ordinary people view situations and assign cause to either the primary actor or the immediate situation. To make a long story short, one typically chooses between a dispositional attribution and an environmental attribution. The former assigns cause to some characteristic of the person who is the accident victim or something the person did to promote the resulting event. The latter examines the forces at play in the environment and how those forces may have shaped a particular outcome and the victim's apparent actions while the event was unfolding. Humans tend to have biases in their causal attributions that place greater weight on dispositional explanations rather than on environmental explanations, especially when actual evidence is insufficient for determining which explanation is more appropriate. In other words, when in doubt, there is a tendency to blame the victim. This bias should be avoided or prevented as a general rule, but it happens often enough. An accident investigation is usually required, however, to uncover a deeper explanation for how the primary actors ended up in the critical situation.

Death Statistics

With the foregoing caveats in mind, we can still see some broad trends in causes of death. In 1900, accidents were the seventh most frequent cause of death after influenza, tuberculosis, gastroenteritis, heart disease, stroke, and kidney disease (Sternberg, 1995). By 1986, the accidental death rate was cut roughly in half from 72.3 deaths per 100,000 population to 39.5, but it was ranked fourth of all causes of death (National Safety Council, 1989) after heart disease, cancer, and stroke. As infectious diseases became curable, other causes of death moved up the ranks. The same rankings persist in more recent data based on 34 countries (National Safety Council, 2002). As of 2010, the top four causes of death in the United States are still the same. Medical technology has made some progress in preventing, or at least slowing down the death rates from the top three origins, but Alzheimer disease has risen to the sixth most prevalent cause of death after respiratory diseases (Alzheimer's Association, 2010; Centers for Disease Control and Prevention, 2007).

Heart disease, cancer, and stroke are stress-related to some extent; age and genetic dispositions play major roles also. A good deal of the stress in life is occupational in origin, and links to major medical outcomes have been documented (Guastello, 1992, 1995; Landsbergis, Schnall, Schwartz, Warren, & Pickering, 1995). Greater incidences of cardiovascular disease have been associated with a three-way interaction among male gender, job strain, and age older than 40 years. Job strain in this context was defined as a combination of high job demands with little decision control (Landsbergis et al., 1995). It is noteworthy, however, that the occupational conditions that could lead to cardiovascular disease could also lead to other medical disorders or psychological health problems, as discussed in Chapter 9.

Occupational Accident Trends

Of all the accidental deaths recorded in the United States in 1988, motor vehicle (nonoccupational) accidents accounted for 47% of incidents. Occupational accidents (including motor vehicles for transportation workers), home, and public accidents accounted for 11%, 23.4%, and 18.8% of accidental deaths, respectively (National Safety Council, 1989). The occupational death rate of 4.3 cases per 100,000 working people is only one third the rate reported in 1928 when specific statistics first became available. It is not clear at first blush, however, how much of the progress at the societal level has been due to safer work practices and how much of the progress can be attributed to shifts in technologies introduced for reasons other than safety, and industry composition in the economy overall, where a substantial portion of the hazardous work has been exported to other countries.

The median occupational death rate over 40 countries was 4.2 per 100,000 in 1998 (National Safety Council, 2002). The rate for the fourth quartile (top 10 countries), however, ranged from 6.5 to 31.0 deaths per 100,000. Interestingly, the later statistics for the United States showed a drop in the rate of motor vehicle accident death to 43% of all accidental deaths, and an increase in the occupational death rate to 16% of all accidental deaths. The occupational death rate for the United States had also increased to 5.0 per 100,000.

Although deaths represent a high watermark that is helpful for understanding accident trends, nonfatal injuries are both a major concern and the criterion used most often for evaluating hypotheses or the efficacy of prevention strategies. In 1988 there were 1.8 million disabling injuries, which was equal to the number of disabling injuries in motor vehicle accidents that year (National Safety Council, 1989).

If a particular type of accident cannot be eliminated altogether, it could be reduced in severity. OSHA defined a reportable injury (on its OSHA 200 form, which is completed by all employers with more than 10 employees) as a "cut, fracture, sprain, amputation, etc. which results from a work accident or from an exposure involving a single incident in the work environment," and requires medical attention. Such an incident is reportable whether or not there is a lost workday. An incident is not reportable if it only requires one treatment of first aid, whether or not medical personnel are available to give the first aid treatment. There is obviously a judgment call involved in the reporting of the least severe cases.

It is important to call attention to the nearly 2000-fold scale factor between deaths and disabling injuries just reported, and realize that large multipliers also exist between disabling injuries and injuries with no lost work time. For instance, Heinrich (1931) also reported that for every nonfatal accident, there are, in turn, approximately 10 near-miss encounters; this was perhaps the first report of a scaling relationship. The multipliers have been recently updated to show that for every work-related fatality, there are 30 lost workday cases, 300 reportable injuries, 3000 estimated near-misses, and 300,000 estimated at-risk behaviors (Gnoni, Andriulo, Maggio, & Nardone, 2013). The specific scaling factors can be expected to vary across industries and over time, but the general principle suggests at the very least that there is an exponential distribution of severities, and a number of factors are probably involved between the near-miss and the actual accident. Some analysts have also found merit in studying near-miss accidents in hopes of preventing real ones. Near-miss accident investigation is considered later in the chapter as a prevention technique.

Figure 10.1 illustrates the most disabling occupational injuries, based on cost information available from insurance company sources (Braun, 2008). The MSK group contains

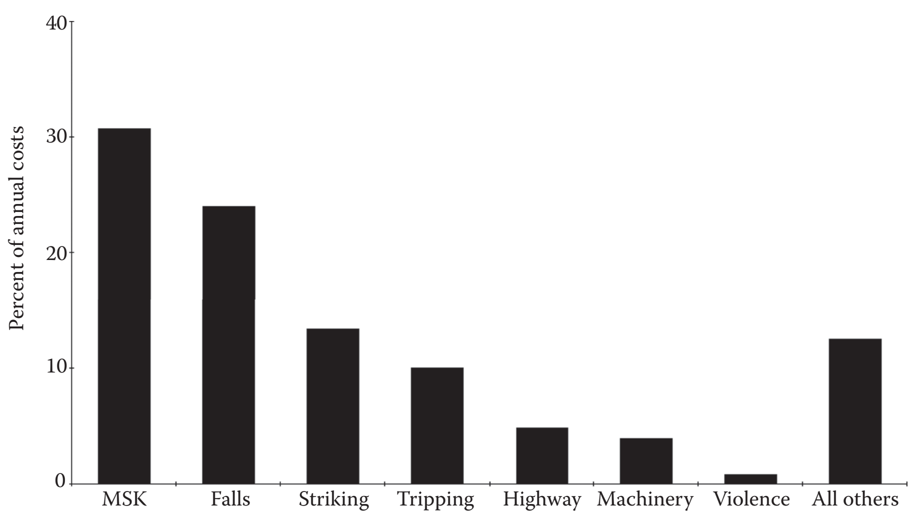

FIGURE 10.1
Most disabling injury groups by type, based on U.S. insurance data from 2005.

musculoskeletal injuries incurred through overexertion or repetitive motions such as carpal tunnel syndrome. The falls include falling to the same level or a lower level of a structure. Striking includes incidents where the individual was either struck by an object or struck against an object. Machinery accidents involve being caught in machinery somehow. Other groupings should be self-explanatory.

Structural Risk Models

Structural risk models have been widely offered as heuristics for understanding accident occurrence. One or another of the models is implicit in virtually every form of accident control. Structural risk models vary in complexity and can be ordered as: single-cause mechanisms, chains of events, factorial approaches, process-event sequences, fault trees and Petri nets, multiple linear models, deviational models, and catastrophe models (Benner, 1975; Guastello, 1989, 1991; Rowe, 1977).

Individual Accident Proneness

The concept of accident proneness first appeared in the 1920s when insurance statisticians discovered that approximately 90% of the industrial accidents involved only 10% of the people in the workforce. That finding led to the premature conclusion that those 10% were chronically doing something wrong, and they were thus labeled *accident prone*. The label provided an illusion of explanation for the mysterious probability structure. Later, however, it was shown that the 90%–10% finding could occur by chance if one assumes that a Poisson statistical distribution, rather than a Gaussian (or normal) distribution, generates accident incidence rates (Mintz & Blum, 1949). With the change in the assumed statistical distribution, several data sets no longer showed abnormal accident frequency rates. Those that continued to show deviations from Poisson expectations could not be interpreted as evidence of individual accident proneness; environmental causes could be responsible just as readily.

Poisson distributions have become very useful for problems involving unusual rates. For instance, if the average computer hard drive crashes after 84 months of service, how likely would it be for a drive to crash after 48 months? Even if we are crashing drives faster than average, how unusual is it for us to be crashing drives this fast? For an accident problem, a similar question might be: If our industry incurs 7 injuries per 100 workers per year, how likely is it for 1 person to incur 2 of the injuries?

Exponential distributions are also useful for studying accident data. In an exponential distribution, a zero frequency of an event will be the most common observation (statistical mode), one event will be the next most common, and so on; its shape is similar to the power law function with the negative shape parameter in Figure 2.8, except that its mode is at 0 and not slightly above 0. For distributions where the overall incidence rate of an event is low, the shape of the Poisson distribution can be approximated by the exponential distribution (Evans, Hastings, & Peacock, 1993).

The exponential distribution has an interesting property whereby the mean is equal to the standard deviation in theory. Thus, when some engineers want to apply a “safety factor” to a system, they might be prepared for incidence rates of the mean times 6. (“Six

Sigma" is a popular criterion). A Six Sigma criterion would indicate that the system would respond correctly to all but 0.25% of situations.

Despite Mintz and Blum's (1949) conclusions, the subsequent 40 years hosted a parade of individual variables studied for purposes of advising employers how to avoid selecting accident-prone job applicants. The prevailing concepts in the past 20 years have centered on impulsivity, personal or social maladjustment, and alcohol or drug use. These personal attributes of people will be considered later in this chapter under the section on accident-prevention programs.

Single-Cause Models

Single-cause models take the form, "If condition *X* is present, then *Y* will occur." Nothing especially complicated is taking place in the modeling sense. Some examples might be: "The spinal cord injury was caused by diving into shallow water," "The train wreck was caused by a faulty connection in the signaling system," "The deaths in the underground mine fire were caused by a failure to develop an evacuation plan," and "The worker was knocked unconscious by a wrench that fell from an overhead platform."

Although these explanations for events look convincing and concise, the pictures of the events become a bit more complex when we ask additional questions. In the case of the diving accident we might ask, "Was it a pool or outdoors?," "If it was a pool was there a deep end?," "If there was no deep end, was there a sign saying 'No Diving'?," "Could the diver see the sign, the deep end of the pool, and so on?," "If not why did he dive?," "Was the diver familiar with the body of water?," and "Was alcohol involved?" As soon as we address these questions we have the potential for multiple single causes, chains of events, factorial models, and other structural varieties.

The assessment of cause is statistically problematic. A significant correlation coefficient, or one of its popular substitutes, does not determine causation; it only determines an association between two variables. Although it is usually the case that a proximal cause will be correlated with accidents, correlation is not sufficient to support causation. To determine causality, the researcher must manipulate the independent variable experimentally or else use a quasi-experimental research design. Archives of accident statistics typically do not lend themselves to that type of analysis. Such information may be very valuable nonetheless, and it holds its value if it is not improperly interpreted.

Single-cause models are often predicated on linear relationships, and other times the relationship is categorical. As another alternative, there could be a breakpoint relationship between the hypothetical cause and effect. In a breakpoint relationship, the amount of *X* must increase up to a critical threshold before any *Y* can occur (Rowe, 1977). This relationship is essentially nonlinear, and implies a transition from a stable safe state to an unstable state.

A common statistical method for assessing potential single causes is the *odds ratio*. All we need to get started is a dichotomous predictor, a dichotomous outcome, and some frequency counts. Consider the following problem: Flying metal debris often causes lacerations to people in a radius of a few feet from the metal-cutting work. Suppose we observed that work sites with one type of metal-cutting tool and machine equipment produced more reportable lacerations than work sites with a different set of tools and machines. We would set up the data as shown in Table 10.1. The independent variable is Machine System 1 versus System 2. The dependent measure is the number of people who suffered no lacerations versus those that did.

TABLE 10.1

Example Data for Machine Systems and Accident Outcomes

	Accident	No Accident	Row Total	Odds
System 1	2 (5.48)	40 (36.52)	42	0.048
System 2	13 (9.52)	60 (63.48)	73	0.178
Column total	15	100	115	

The first step in the analysis is to determine whether there is an association between the two variables. For this purpose we could use the χ^2 test with the Yates correction for continuity. However, because the expected frequency for one of the cells is very low—and in many situations the expected frequencies could be even lower—the likelihood χ^2 would be the favored choice of statistic. (The likelihood χ^2 can be used for problems with larger subsamples also.) The expected frequencies of accident events are shown in Table 10.1 in parentheses. The expected frequency is the row total times the column total, divided by N = number of observations for this problem. The χ^2 statistic is computed as:

$$\chi^2 = 2\sum Ob\ln(Ob/Ex), \tag{10.1}$$

where Ob is the observed frequency for a cell, Ex is the expected frequency, and the summation is over all four cells (Howell, 1992). The χ^2 statistic for this problem is 4.58, which is significant at $p < .05$. (A p value of <.05 is sufficient when testing only a small number of potential single causes.)

Now that an association between the two variables has been established, the next step is to produce the odds ratio. The odds of an accident with System 1 would be 2/42, or 0.048. The odds of an accident with System 2 would be 13/73, or 0.178. Take the ratio of the two odds, and we can state that the odds of an accident in System 2 are 3.74 times as great as the odds of an accident with System 1.

It is important to remember, however, that odds ratios can be misleading when taken out of context. It is one thing to say that the odds of an event are "twice as likely" under a circumstance when the base-rate chance of occurrence is 0.1, but quite another when compared with a base rate of 0.000001. The orders of magnitude matter.

Multiple-Cause Models

Single causes can expand into factorial models as soon as multiple Xs are found to be associated with the unfortunate Y. The logic then becomes, "if $X1$, $X2$, or $X3$, are present, either alone or in some combination, then event Y will occur."

The diving accident scenario that was introduced earlier is based on some real events. A drought in Wisconsin in 1988 resulted in severely low water levels in many public waterways. There was in fact a spate of 11 spinal cord injuries resulting from diving into shallow water. Branche, Sniezek, Sattin, and Mirkin (1991) assembled a control group of 22 men who matched the 11 injured in several ways: They lived in the same counties of Wisconsin, used public water facilities that summer, and were between the ages of 18 and 45 years.

Ten independent variables were tested for comparison between the injured group and the control group. The injured were 12 times more likely to have entered the water from a dock or pier, 10.8 times more likely to have dived rather than jumped, 7.5 times more likely to have dived into water less than 5 ft deep, 6.0 times more likely to have used medication on their last water recreation, 4.3 times more likely to have been unfamiliar with the body of water, 4.0 times more likely to have used alcohol on their last water recreation, and 2.9 times more likely to have a high school diploma or less education.

The statistical analysis does not offer any clues as to which variable or condition preceded another, but the risk factors themselves might suggest a logical pathway. At the end of the chain, however, all the so-called risk factors can be channeled into two important proximal ones: the water was too shallow and the person dove in from a dock or pier.

Domino Models or Event Chains

Domino models are chains of events. The Chicago fire is a classic example. The cow knocked over the lantern, which ignited the hay, which ignited the barn. The fire from the barn spread to the next house, then to the next house, then to the next house, until it burned the last available building. Note the critical importance of each step in the process on each subsequent step.

Some multiple-car collisions are examples of event chains. Imagine a line of cars stopped at a red light. A new car joins the pack from the rear, but its driver fails to apply the brakes soon enough. The new car hits the one in front of it, forcing that car into the one in front of it and so on. The chain of collisions might be broken if there was enough space between the cars, the drivers who were about to be hit had their feet firmly on the brakes, or one of the vehicles was large enough to absorb enough shock without being forced forward.

Another type of traffic event chain might start with the lead car of a moving pack. The lead car hits a surprise patch of ice and spins out of control, whacking the next car in line. The third car takes evasive action but slides into a ditch while doing so.

A common but short chain of events is often detected or assumed in occupational accidents. In behavior modification approaches to workplace safety, management introduces a plan to recognize and reward specific safe behaviors, such as wearing protective eyewear. Higher utilization of protective gear should produce fewer accidents involving injuries to the eyes.

Reniers and Dullaert (2007) developed a computer program that would uncover the potential for domino scenarios in a process industry. In chemical processing, for instance, a chemical accident of one type leads to additional accidents, making the end result a lot worse compared with what might have been the case if the chain were interrupted after the first event. The execution of the analysis requires collecting a database of probabilities of events and collateral events. It calculates how much danger is exported to adjacent events given that a source event occurred. From that point, it is possible to calculate the odds of particular domino scenarios by cross-multiplying the odds of events within a hypothetical chain. The underlying logic is not very different from that used to construct fault trees, which are considered here a bit later.

Factorial Models

Single causes can expand into factorial models as soon as multiple Xs are found to be associated with the unfortunate Y. The logic then becomes, "if X_1, X_2, or X_3 are present in some combination, then a known level of Y will occur." Here we are saying that an accident is explained by a combination of risk factors. More commonly, however, we are saying that a work unit's accident rate is explained by a weighted combination of factors, such that:

$$AR = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n. \tag{10.2}$$

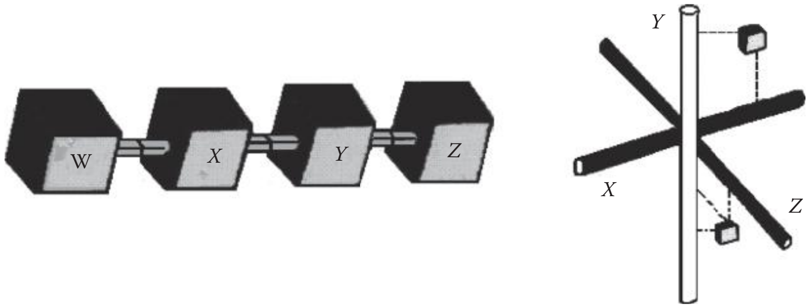

FIGURE 10.2
Visual comparison of event chains and factorial models of accident causation.

Equation 10.2 is a multiple linear regression equation where AR is the accident rate, b_i are regression weights, and X_i are the independent variables. Note: In many cases, better results might be obtained by converting AR to $\log_{10}(AR + 0.05)$. A visual comparison of the chain of events and factorial models appears in Figure 10.2.

Consider an example of an accident analysis that was conducted in this fashion. The situation involved 658 employees at a secondary metal fabrication facility (Guastello & Guastello, 1988). The employees completed a paper-and-pencil survey that covered a number of variables related to safety management, different kinds of stress, and different kinds of hazards. There was one question added to the end of the survey, which asked how many accidents they were involved in during the past year. Note that "involved" denotes not only actual injury, but includes workers who may have played some role in a collective task in which someone in the work area was injured. The involvement rate in this situation was about double the OSHA-reportable accident rate for the same period.

There were many results from the survey that were useful to the safety management objective at the plant. Three variables were isolated with a stepwise multiple regression procedure, where individual accident involvement was the criterion. The factorial model for the location, however, was composed of three variables: anxiety levels, the number of environmental hazards (from checklists), and danger levels. Questions regarding danger levels indicated the level of severity of potential injuries in a work location.

The coefficient of multiple regression was .35 ($p < .001$), indicating that the three variables accounted for 12.3% of the accident involvement variance (R^2). The regression weights indicated that anxiety accounted for approximately 10% of the accident involvement variance, and the other two variables accounted for an additional 1% each. Although these may seem like low numbers, they are within the common range of what to expect in factorial investigations of this type. There is indeed a lot of unpredicted behavior variance, however, and statistical analyses that are stronger than multiple linear regression are probably needed to account for it; this particular need was part of the motivation for the cusp catastrophe model and other accident process models that are considered next.

Process-Event Sequences

The simplest form of event sequence model accords less attention to causes and more attention to the outcomes leading up to an accident. The nuance here is that an accident is a process, rather than a single discrete event. Surry (1969) conceptualized the accident process as a hazard build-up cycle. At first, the workplace is safe with no uncontrolled hazards. As people start to work, however, tools are left out in workspaces, and different people enter the workspace to do different things with different tools and equipment. People and objects move around and make opportunities to bump into each other. Eventually, hazards accumulate to a critical level when an accident occurs. Notice that there is an entropy concept implicit in the hazard build-up view of an accident process.

An intervention based on the hazard build-up cycle would emphasize training for good factory housekeeping. Other possible forms of training would center on the best use of tools, and procedures that would minimize the acceleration of the hazard build-up. Workers should learn to recognize the build-up cycle and to spontaneously intervene by reorganizing their workspaces for safer operation.

Fault Trees

A fault tree is a diagram that represents a sequence of events. The sequence could represent a process or a chain of events. The distinguishing characteristic of a fault tree is that it contains one or more branching functions that denote choices, options, or opportunities for multiple outcomes. A fully defined fault tree would contain estimates of probabilities associated with each branch at a node.

The fault tree in Figure 10.3 is a reinterpretation of the Sunshine Silver Mine disaster (Denny, Gilbert, Erdmann, & Rumble, 1978). The disaster started several floors underground in a trash heap in an obscure part of the working cavern. The ventilation system was inadequate, and caverns filled up with smoke. Not everyone was able to detect the smoke before it blocked their exit to the nearest escape route. Analysis showed that the mine did not have a working evacuation method, so the miners had to figure the way out extemporaneously. Those that survived to the next set of options faced insufficient elevator space and respirators that often did not work. If the devices did function properly, many of the miners did not know how to use them properly. The net result was 91 deaths out of 173 miners within an hour.

It is also possible to construct a fault tree prior to an unfortunate event. In those cases, it would resemble a decision tree. If the possible nodes where an error can occur can be mapped out in advance, even if the error rates are not known, then it would be possible to insert diagnostics into the system that will alert operators or system controllers to impending errors.

Figure 10.4 depicts the core elements of a fault tree whereby events and subevents have probabilities (p) associated with them. There is an initial condition that triggers possible combinations of outcomes. At each juncture, there is an opportunity for a risky situation to continue or terminate. Event 3 in Figure 10.4 is where the accident or injury actually occurs or a recovery is possible. The odds of an accident would be $(p_1) \times (p_{2a} \mid p_1) \times (p_{3a} \mid p_{2a})$. Breaking the sequence down into steps illustrates points of possible recovery as well as a picture of how it happened.

A closer look at Figure 10.4 reveals that the hypothetical calculation just made here constitutes the odds of a particular domino sequence occurring. The essential difference between a fault tree and a domino structure, however, is that multiple overlapping domino structures could exist, and "danger units" could spread in multiple directions from the source. Fault trees, specify fixed event structures that could be prevented from reaching criticality if certain yes/no gates were open or closed. In some simpler contexts, the distinction between the two types of processes is not blatant.

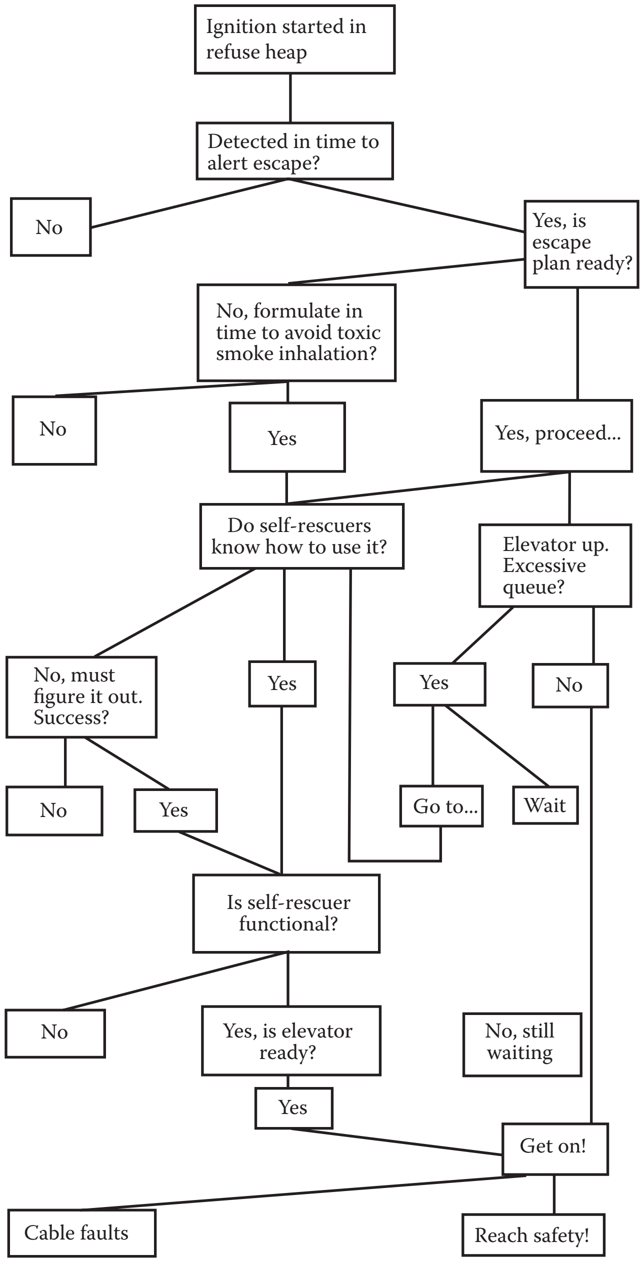

FIGURE 10.3
Fault tree analysis of an underground mine fire. (From *Chaos, Catastrophe, and Human Affairs: Applications of Nonlinear Dynamics to Work, Organizations, and Social Evolution*, p. 210, by S. J. Guastello, 1995, Hillsdale, NJ: Lawrence Erlbaum Associates. Copyright 1995 by Lawrence Erlbaum Associates. Reprinted with permission.)

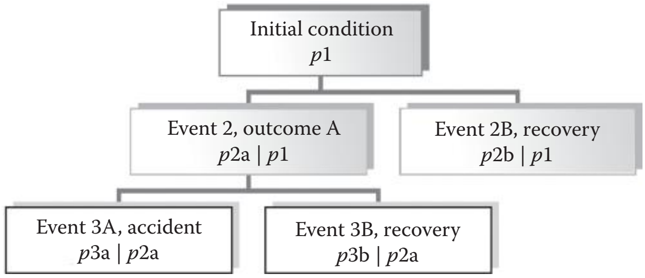

FIGURE 10.4
Segment of a generic fault tree annotated with conditional probabilities associated with events.

Flow Charts and Petri Nets

The social science literature is speckled with models of phenomena that are characterized by the prominent use of boxes and arrows. The contents of the boxes could indicate objects or events. The arrows convey a vague notion of causation. If the flow of arrows heads in one direction only, the model might be verifiable through use of linear structural relational analysis. That type of analysis produces patterns of linear correlations such that higher correlations are interpreted as stronger causal links, as if the correlations resulted by a causative pattern. The analysis would also ascertain whether the patterns that were predicted within a hypothetical model were stronger than links that one would obtain from alternative patterns of correlation.

Petri nets look like boxes and arrows also and often resemble fault trees. With Petri nets, however, the definition of system elements depicts more literally the flow of information among subsystems of person–machine networks. Marked Petri nets are embellished with big dots in the boxes that indicate that a parcel of information has reached a particular box or subsystem (Peterson, 1981). Marked nets can be found in industrial engineering where some form of person–machine communication is taking place.

Love and Johnson (2002) used the Petri net method to display the conditions leading to the Claptham Junction train wreck in the United Kingdom. An improper repair on a signaling system gave several trains the wrong cues leading to a collision between two of them. A third train crashed into the wreckage of the first two. Four other trains passed through the same junction and could have easily been involved in the multitrain collision had the timing of their arrival been different. Further investigation into the event uncovered that the technician responsible for the faulty repair had been working excessive hours because there were too many problems to fix in too little time and without enough qualified help. The event was actually a casualty of fatigue, something that was not well defined in the Petri net model. One might drop backward in time a step further, however, and ask why management allowed the situation to develop as it did.

Advances in computer technology have made it possible to develop simulations of accidents in animation form (Johnson, 2002). One way to use the technique would be to vary system attributes, for example, the speed of the vehicles and the timing of possible evasive maneuvers, and determine visually if a particular event would have taken place under specific conditions. Such simulations may be worth the proverbial 1,000 words for conveying to nontechnical audiences, such as courtroom juries, what actually happened in a given situation. On the other hand, the asset of this method is also its limitation: Erroneous assumptions that are built into the scenarios in the simulation would also be easy to understand and remember and could be more convincing than a more accurate explanation that is conveyed in a less dramatic form.

Complex and Circular Causation Network

The following substantive model of industrial accident causation is an attempt to organize numerous qualitative research findings on the linkage between human and environmental characteristics and occupational accidents. It is intended as a transitional concept from which we can further derive a working nonlinear dynamic model of the accident process.

Figure 10.5 shows the apparent connections among stress, anxiety, errors, hazards, perceived danger, beliefs about accident control, and errors and accidents. To begin, it is only necessary to say that the definitions of stress in its many forms and anxiety that were presented in Chapter 9 are operational here. Correlations between anxiety and accidents are known (Guastello, 1993), but the relationship appears to be bidirectional. Leary (1990) identified two paths by which anxiety could lead to a behavioral disruption. In the first case, a person's cognitive evaluation of a situation, here a dangerous one, leads to anxiety, or the excessive fear that something bad will happen. As a result actions are not taken that could have been taken because the person could not decide what to do. A second path is one where a person engages in a behavior that would ordinarily be appropriate, but anxiety intrudes and interrupts the behavior sequence. The interruption may take the form of an inopportune hesitation, or of dropping everything and making a quick exit to go do something else. Thus, people climbing a mountain or working on a tall building tell each other not to look down; the arousal level produced by the height and perception of the possibility of immediate danger could result in a behavioral interruption at a very bad time.

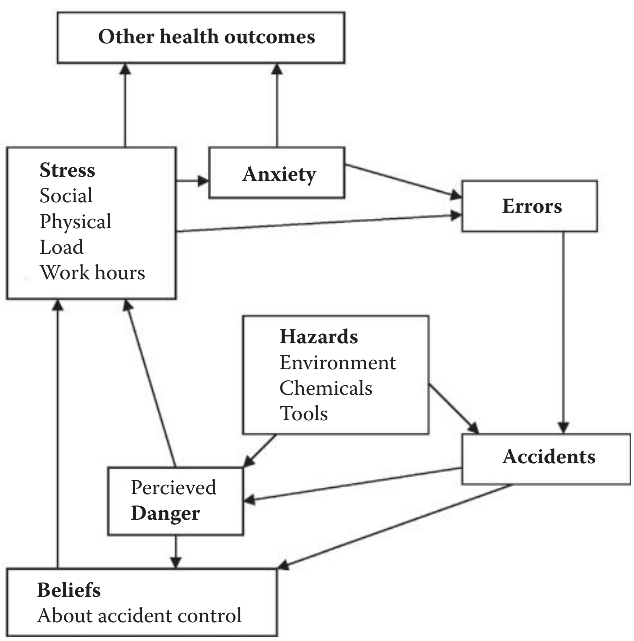

FIGURE 10.5
A network of psychosocial factors in the accident process.

Accidents, or even some near misses, could easily trigger elevated anxiety on the part of the worker (Guastello, 1995). Even if the person involved was not personally harmed or as badly harmed as others, the effect of seeing what happens to others can be greatly traumatic. What we have essentially established are two mechanisms by which stress can be regarded as contagious. In one sense, there is within-person build-up caused by repeated exposures to stress and personal accident involvement. Each accident or serious near-miss serves the purpose of elevating stress beyond the threshold at which anxiety sets in. The environmental cues that trigger anxiety would have a greater effect if the individual's level of state anxiety was also elevated. Alternatively, there need not be an experience of anxiety for stress to affect behavior.

In the second contagion mechanism, an accident experienced by one person leads to the experience of stress by another. Thus, there is transpersonal stress build-up that is the more conventional notion of contagion.

The next two loops in Figure 10.5 express how danger and beliefs about accident control are connected to the circular stress–anxiety–error–accident cycle. Hazards in the environment are perhaps the first prerequisite for any type of accident process. If we all worked in rubber rooms, there would be little point in discussing these concepts. Although they make direct contributions to the accident process, hazards also lead to perceptions of the level of danger, and that level of danger serves to modulate the person's stress level. People also habituate to stress and danger levels, which could be a benefit in one sense, but a limitation in another if it means that the individual become oblivious to the hazards that need to be negotiated. Hazards can be further separated into their sources of origin, such as environments where, for instance, fires and explosion occur, tools, or chemicals.

Some people believe that the good and bad things that happen to them in life are the results of their own behaviors, whereas others believe that most events in their lives a matter of luck. This polarity is well known as the *internal* versus *external locus of control* concept (Rotter, 1966). Several attempts have been made to link a variation of the locus of control construct to accident involvement. Some people believe that accidents are controllable whereas others believe that accidents are a matter of bad luck. The concept of beliefs about accident control was first advanced as an *outcome* of accident involvement (Guastello & Guastello, 1986), although other researchers have attempted to use it predictively (reviewed by Arthur & Doverspike, 1992) in what amounts to a concept of individual accident proneness.

Each connection in Figure 10.5 presents an opportunity for safety management. The concept of safety climate was first expressed by Zohar (1980) who was investigating safety practices, and workers' views of those safety practices, that distinguished factories with good safety performance from those with poor performance. Attitudes toward the organization's safety program and its effectiveness, worker training, availability of needed tools and personal protection equipment, and the foreman's attentiveness to rule violations, all served to distinguish high- and low-performing groups.

Cusp Catastrophe Model of the Accident Process

The cusp catastrophe model is a risk analysis structure that can be used to describe a single accident event or a system of individual or group accident experiences. The strength of its predictive value resides primarily on empirical analyses done with occupational samples. Extensions to the interpretation of a single accident can be made based on the collective results. An accident investigation could be framed around two questions: What were the background conditions or variables? What were the trigger variables?

The cusp catastrophe response surface that appears in Figure 10.6 is labeled for the general accident model (Guastello, 1989, 1991, 1992, 1995, 2003). See Chapters 2 and 9 for further explanation and other examples of the cusp model. There are two groups of variables in the accident model, which correspond to the two control parameters. The asymmetry parameter consists of environmental hazards. The bifurcation variable consists of several types of stress or stress-reducing variables that are collectively known as *cognitive load*. The concept of cognitive load includes various forms of stress such as work pace and other contributions of safety management; physical stressors such as heat, cold, noise, and crowdedness; and social stressors such as poor job security, changing job assignments, and difficulties with superiors or coworkers. Workgroup size also contributes to the bifurcation parameter in workgroups, wherein the accident rates of small groups are more variable than those of larger groups (Guastello, 1988, 1989). On the other hand, small groups also have access to locally stable low accident rates; I return to this point later on.

The process for a single accident works as follows. The system begins at a stable state of low risk. Risk remains low while ambient hazards and load variables slowly change. Once the control variables reach a critical point, however, there is a sudden change to a high risk where an accident occurs. Subcritical risk levels are captured by the response surface also. The process for group-level accident rates works in a similar fashion. At the group level, however, we can specify returns to lower accident rates when the values of the control parameters have receded sufficiently far.

Literal tests of the cusp model have been made with industrial work units (Guastello, 1989; Guastello & Lynn, in press), bus drivers (Guastello, 1991, 1992), and hospital

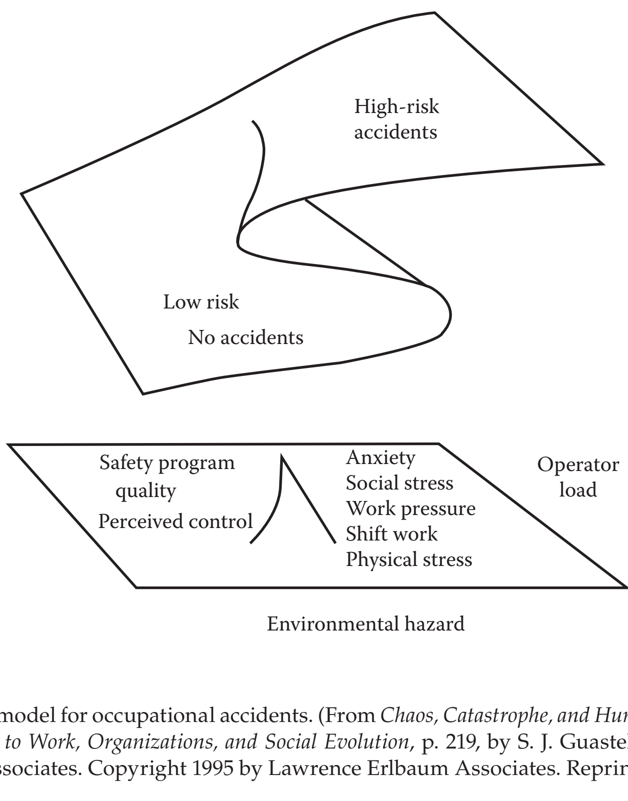

FIGURE 10.6
The cusp catastrophe model for occupational accidents. (From *Chaos, Catastrophe, and Human Affairs: Applications of Nonlinear Dynamics to Work, Organizations, and Social Evolution*, p. 219, by S. J. Guastello, 1995, Hillsdale, NJ: Lawrence Erlbaum Associates. Copyright 1995 by Lawrence Erlbaum Associates. Reprinted with permission.)

personnel (Guastello, Gershon, & Murphy, 1999). The critical variables were measured by a survey that was completed by employees. The relevant variables are going to be somewhat different in different situations, but they do represent the general themes specified previously.

Table 10.2 provides an index of accident situations with their formal constraints. The strongest group dynamics appear in industrial settings. Employees tend to work together in an intact unit. The output of one person affects the others. Outputs include contributions to the relative safety of the work environment, and atmosphere of stress, anxiety, and confidence in safety management. Bus drivers, however, work mostly alone, although there should be some radio contact with a central controller who can authorize help where

TABLE 10.2

Comparison of Accident Models and Situations

Model Property	Manufacturing Facilities	Public Transportation	Health Care
Group dynamics	Employees tend to work together in an intact unit. The output of one person affects the others. Outputs include contributions to the relative safety of the work environment, and anxiety, land confidence in safety management.	Employees usually work alone, but there is some radio contact with a central controller. Hazard levels change as bus moves through different neighborhoods and as traffic situations change. Stress, anxiety, and confidence in management are shared by virtue of common experiences.	Employees usually work in cross-functional teams that are reconfigured many times per day according to the needs of patients. Stress, anxiety, and confidence in management are shared by virtue of employee cohesion and common experiences.
Type of system	Relatively closed	Relatively open	Relatively unstable
Type of observation pairs	OSHA-200 accident reports for workgroups measured at two points in time.	Self-reports by individuals for bus accidents in personal automobiles.	Self-report by individuals for occupational accidents generally and for blood-specific risks (needle sticks, sharps).
Environmental hazards (a)	Tools, toxins, poor housekeeping, crowded walkways, ratings of danger severity.	Unruly and potentially violent passengers, need to reprimand passengers for radios, smoking, etc. Hazard levels vary by shift.	Verbal abuse from coworkers and patients. Exposure varies by professional group.
Cognitive load (b)	Physical and social stressors, safety climate, beliefs in accident control, workgroup size, amount of time spent working on recommendations from the ergonomic report.	Social stressors (job insecurity, changing assignments, difficulties with coworkers or management).	Shift work, safety climate, depressive symptoms, work pace, physical stressors (noise, heat, cold).
R^2 cusp model	.42	.63	.75
R^2 best linear alternative	.05	.26	.07

Source: S. J. Guastello: Nonlinear Dynamics, Complex Systems, and Occupational Accidents. *Human Factors and Ergonomics in Manufacturing*. 2003. 13, 293–304. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

needed. In fact, the primary recommendation for this group was to provide a better radio and response system for drivers in emergency situations. Health-care employees usually work in cross-functional teams that are reconfigured many times per day according to the needs of patients. Industrial workgroups are relatively closed systems, whereas public transportation work systems are relatively open, and health-care groups are relatively unstable.

The accident rates in any business' experience sometimes vary sharply over time. The cusp model would explain the variation as change around the bifurcation manifold. Reason (1997) also observed the oscillations as well as the two stable states. His explanation was that the safety management factor was actually changing, such that it sometimes favored low accident rates and less productivity. At other times management favors faster production, which promotes more human error and accidents. Management is likely to gravitate toward higher risk and higher production after it has experienced an extended period of lower risk, lower production just because it loses the memory of what could happen if it speeds up the work processes.

The model requires that the dependent measure (accident count or rate) be measured at two points in time. In the industrial setting, the dependent measure was the workgroups' accident rate based on their OSHA-200 report at two points in time. For the bus drivers, accidents were measured as self-reported counts at work and self-reported counts of accidents in personal automobiles; this difference would convey the occupational risk exposure controlled for a base rate of driving accidents for that driver. For the health-care setting, the two dependent measures were self-reported individual accidents of all forms and self-reported incidents of blood-specific events such as needle sticks and sharps. In the health-care setting, it was possible to predict the general accident rate from blood-specific incidents rather than the other way around (Guastello et al., 1999). The qualitative variables for asymmetry and bifurcation parameters were measured using a version of the Occupational Hazards Survey, which is completed by the participating employees.

Complex Dynamics, Events, and Deviations

A multilinear event series (Benner, 1975) is a compilation of multiple simple-event series and fault trees. The new premise is that a complex manufacturing system is composed of activity patterns that occur simultaneously. In principle, the workflow patterns have been designed by management to maximize work output and to minimize accidents. When patterns of activity fall out of sequence, or outside the toleration limits for those sequences, risks escalate.

Kjellen and Larsson (1981; Kjellen, 1984a, 1984b) extended the multilinear event concept by introducing the concept of deviations. The conceptual model assumes that there is a normative and functional work pattern with which to begin. When deviations from norms are introduced, they are carried through the system from subprocess to subprocess and their impact magnifies. At some point in the magnification process, a burst of energy is released that suddenly transforms the system into the second stage of the accident process. During the second stage, personnel have the opportunity—maybe—to take evasive action either to prevent the accident itself, thereby transforming it into a near-miss event, or to minimize the damage. Large accumulated deviations may interfere with successful evasive maneuvers. The third stage of the process is the actual delivery of harm to the employee; by that time, the event is final.

Sources of deviation in a work setting may pertain to the flow of materials, changes in personnel assignment, flow of information, equipment anomalies, intersecting activity patterns, and environmental disturbances (Kjellen & Hovden, 1993, p. 421). Kjellen and Hovden presented an example of how a confluence of deviations described a construction accident: A regular worker was out sick and was replaced by an apprentice. The task of the day was to erect a beam. A crane that would have been very useful was needed somewhere else. There was ice on the beam. The worker erected the beam manually, but it was crooked, so he walked out onto the beam to realign it, slipped on the ice, and fell to the floor below, breaking a rib and puncturing a lung.

The deviation principle in accident investigation is similar to the deviation principle encountered in systems for ensuring quality in manufactured goods. Kjellen and Hovden (1993) observed through experience, however, that the standards against which one might compare deviations are more rigorously defined in product quality management and involve fewer sources of deviation. Operationalization of the model requires detailed information sources. Both issues, in their opinion, explained the limited uses of the deviation principle in actual practice. Larsson (1993) observed that the deviation models should include cognitive and decision-making aspects of the work process that could describe the onset of some forms of deviation and correct others.

A simple self-organized system appears in Figure 2.6. Its important feature is that the parts have loops backward and forward so that part 1 affects part 2, 2 affects 3, 3 affects 4, but the output of 4 also affects 3 or 1. The flow of information may be steady, but it may just as readily be periodic or chaotic over time. Feedback loops may serve to dampen the variance over time (negative feedback) or accelerate it (positive feedback). A system's boundaries will have an impact on the structure that forms spontaneously from a system that started as chaotic.

When a system makes a distinctive change from one form of organization to another, or from a state of chaos to self-organized structure, the change in system entropy over time can be represented mathematically as a cusp catastrophe model (Guastello, 2002; Haken, 1988). The cusp suggests conditions for reverse dynamic patterns, and indeed reverse patterns are often found. When a system is subjected to large levels of entropy, its behaviors change from self-organized to chaotic. Chaotic systems then reorganize according to a more effective, coherent, but different pattern of information flow.

Work organizations naturally form hierarchies (as do some nonliving systems) as a means of self-organizing in response to entropy. One would then expect the system to remain stable by virtue of strong, top-down, hierarchically driven management forces, and often that is the case. On the other hand, new information is generated within the experiences of the lowest level of the hierarchy, particularly if management directives contradict, or are indifferent to, the decision requirements at the lower levels. As a result, the workgroups, even efficient ones, can destabilize the behavior of management (Guastello, 2002; Guastello & Johnson, 1999).

Many examples from biology and organizational behavior indicate that healthy systems are complex adaptive systems. Complex adaptive systems are poised on the edge of chaos and self-organization. As such they are receptive to a wide range of stimuli from their environments and make a comparably wide range of responses to the incoming stimuli (Dooley, 1997; Guastello, 2002). This new thinking about organizations, together with the value of chaos as an explanation for seemingly random events through deterministic processes, has led to the conjecture that chaos might indeed explain some occupational accidents (Nielsen, 2000). Chaos and complexity could be hiding in work situations that involve at least three coupled oscillators or semi-independent subunits. Thus, group- and organizational-level dynamics are considered next.

Group Dynamics, Safety Climate, and Resilience

Although accidents happen to people who are often carrying out activities on their own initiative, occupational accidents are group phenomena to varying extents, as mentioned earlier in conjunction with Table 10.2. This section of the chapter expands on the group dynamics leading to accident occurrence or control, and the concepts of safety climate, safety culture, and resilience.

Group Dynamics and Complex Technologies

The contrasting group dynamics for industry, transportation, and health-care situations indicate one source of self-organization. The group work environment involves numerous local interactions among the individual employees and management representatives. In the course of an industrial group task, the employees coordinate their actions with each other, and wherever possible watch for hazardous acts of others that could affect themselves personally, or warn others when a hazardous situation is developing.

In transportation situations, the information flow among workers is much weaker, and occurs through a central controller at best. The majority of safety-related decisions are going to be made on the spot by the driver in response to momentary changes in conditions.

The health-care situations often involve cross-functional teams of nurses, physicians, and technicians. Emergency (stress) levels can vary at a moment's notice. In hospital emergency rooms, it appears that a substantial number of adverse outcomes can be traced to communication failures among personnel (Xiao, Patey, & Mackenzie, 1995).

Coordination training among team members will help any one particular person integrate well with others. This is an important point if a smooth assimilation of a new employee is desired, which is often the case. Experimental studies on group coordination indicate, however, that coordination will become chaos when more than 50% of the component employees are not familiar with each other (Reason, 1997). Group coordination can occur without any management help (Guastello, 2002), although placing "coordinator roles" within large workgroups (Comer, 1995) might facilitate coordination. The optimal ratio of workers to safety coordinators appears to be about 15:1 (Guastello & Guastello, 1987b).

Not all information flows contain useful information. Stress and anxiety are contagious within workgroups too. Positive feedback loops within workgroups can accelerate stress and anxiety effects over time. The unwanted impact of stress can spread throughout a workgroup also, particularly when an injury occurs.

Computer-driven automation can have a favorable impact on safety insofar as a greater distance is placed between the operator and mechanical forces that induce injury. On the other hand, the trend in automation is to place greater cognitive loads on the operators while reducing physical loads (Eberts & Salvendy, 1986). To complicate matters further, the operators work through computers, which are virtual representations of real systems. Unlike virtual reality software, however, the benefits of immersion and the fidelity of cause and effect in the software are not present. Instead, operators such as air traffic controllers must detect signals on a screen, maintain radio contact with many aircraft, and remember where they put the airplanes. Consider this scenario reported by NASA researchers for aviation safety:

> In 1991 a tower controller at Los Angeles airport cleared a Metroliner to position on a runway and hold for release while she instructed other aircraft to cross the other end of the runway, a common practice. Several unavoidable delays occurred, and the Metroliner was not conspicuous in the twilight with poorly designed airport lighting. The controller forgot that she had not released the Metroliner to take off, and she cleared another aircraft to land on the same runway, which it did, crashing into the Metroliner. (Stone, Remington, & Dismukes, 2001, p. 1)

At one level of analysis, the Metroliner accident can be regarded as another example of environmental hazard and operator load reaching critical points in a cusp catastrophe model. At a second level of analysis, the Metroliner accident is an example of mode error: Complex technical systems may perform many functions automatically and thus involve relatively few humans. The humans, however, must keep track of the state of the system and its recent history; some control actions by the humans are permissible during some states of operation, but not in others (Hudson, 2001). At a third level of analysis, we are concerned with how situations like this one develop. Other explorations into airline errors have examined the extent to which errors made by air traffic controllers, flight captains, or crew are detected and corrected by one another (Casner, 1994; Fischer et al., 2001). These complex systems offer many opportunities to apply concepts from nonlinear dynamics, especially where the specific interactions among flight personnel must be understood.

Safety Climate

The understanding of the group phenomenon known as *safety climate* is starting to undergo some development and with some added complexity. The original construct of safety climate (Zohar, 1980) captured a social or organizational climate that was characterized by the adoption of safe operating practices, workers' evaluations of those safety practices. Safe and unsafe work environments could be distinguished, all other things being equal, on the use of personal protection equipment, vigilance on the part of workers and management, the safe behavior of employees, and good housekeeping.

The concept of climate was not far removed from the broader concept of organizational climate (Schneider and Reichers, 1983) that was prominent at the time for distinguishing the social atmospheres of organizations. An organizational climate describes patterns of interactions between the individual and organization as a whole and can distinguish organizations in much the same way as personality distinguishes individuals. Some of the classic distinctions in climate include humanistic and participative versus authoritarian and exploitive (Likert, 1961; McGregor, 1960) or trait-like constructs such as level of achievement orientation, degree of risk taking, level of individual autonomy, and warmth and support (Saal and Knight, 1988). The more specific notion of safety climate can be expected to combine with other aspects of the organization's climate to produce a more unique experience regarding safety matters (Wallace, Popp, & Mondore, 2006).

Although the original concept of safety climate is closely tied to the experiences of workers on the shop floor, safety climate, like other notions of organizational climate, is strongly affected by policies of upper management. Management policies affect the relative priority of work speed and profits versus safety, the choice of accident-prevention programs, the efficacy of those interventions, equipment maintenance policies, and adequate maintenance of equipment (Reason, 1997). Reason suggested, furthermore, that some of the fluctuations in an organization's accident rate could be traced to waffling in priorities that are expressed by the various safety policies and concerns about production.

There is also some bottom-up influence in the production of a climate. Consistent with the broader construct of climate, climate levels within a work unit are usually thought to be similar to the perceptions of safety climate in the organization as a whole (Zohar and Luria, 2005). Greater variations in climate are found in workgroups, however, when the supervisors have greater autonomy for setting work pace or enforcing standards. Neal and Griffin (2006) saw level of participation in safety functions as being an important facet of climate and contributor injury levels. Other researchers have taken the idea step further to advocate developing a strategy for participatory ergonomics more broadly (Halpern and Dawson, 1997; Jensen, 1997; Moore and Garg, 1997). Recognizing that bottom-up influences can present some difficulties, Kjellén and Hovden (1993) advocated the development of safety information systems to assist with greater top-down control over deviations in work processes and to identify sources of unwanted novelties that build up into hazardous circumstances.

Despite the strong conceptual case for the role of management in safety climate, the correlations between measures of safety climate and actual individual accident involvement or group accident rates are not always consistent, however (Clarke, 2006; Fullarton & Stokes, 2007; Johnson, 2007). According to Clarke's (2006) meta-analysis, the correct mean correlations (ρ) between safety climate with safety compliance and safety participation were .43 and .50, respectively. The ρ between compliance and participation with accidents and injuries were .09 and .14, respectively, indicating a substantial attenuation effect between employee actions and actual outcomes. Because the lower bound of the 90% confidence intervals was still positive, those four values of ρ are considered generalizable, however. Safety climate actually showed a stronger relationship to injuries ($\rho = .22$, 28 studies, $N = 17{,}695$), but did not generalize because the lower bound on the 90% confidence interval was negative and only 22% of the variation in correlation coefficients could be explained by artifacts. Those results indicated that 10% of the relationships between climate and accidents were likely to run in the counterlogical direction, and a thus substantial amount of situational specificity was occurring overall.

Several writers suggested other sources of situational specificity for safety climate. Some of the candidate variables are the technology (Zohar, 2003), beliefs about accident control (Guastello, 1989, 2003; Neitzel, Seixas, Harris, & Camp, 2008), stress and anxiety levels (Dunbar, 1993; Gandit et al., 2009; Guastello, 1989, 2003; Hellesoy et al., 1998; Kerr et al., 2009; Nagashima et al., 2007; Parkes, 1998; Quick, Murphy, & Hurrell, 1992; Sauter & Murphy, 1994), organizational size (Fullarton & Stokes, 2007), and organizational subunit size (Guastello, 1988; Guastello & Guastello, 1987), and other variables that are part of the broader spectrum of organizational climate (Wallace et al., 2006). Anxiety as a safety climate variable is considered next.

Anxiety is often observed or experienced as a consequence of job stress. Job stressors variously include workload and time pressure, stressors with social origins such as job insecurity, or those with physical origins such as noise, heat, cold, shift work, and physical danger sources (Quick et al., 1992). It could be transient, arriving and disappearing as the stress sources come and go, or it could be a longer term effect. Longer term effects might reflect personality traits, but could also reflect long term job exposures, particularly if anxiety levels are widespread in the workgroup. The psychosomatic symptoms take a while to build up. Some plausible pathways for anxiety to disrupt behavior were discussed above. Although anxiety has origins that are not simply reducible to job stress, stress management programs have been effective for reducing accident rates by an average of 15% (Murphy, 1984; Murphy et al., 1996), and by implication, reducing anxiety levels as well.

More recent studies, however, have reported positive effects of stress on performance, with the interpretation that coping mechanisms appear to be playing a decisive role in channeling arousal into greater focus on the task instead of transforming it into disruptive anxiety (Gandit et al., 2009; Matthews & Campbell, 2009). Others have observed that anxiety can serve as a positive force in a group because it results in greater vigilance over risk factors (Ein-Dor et al., 2010).

Notwithstanding the cusp catastrophe model for safety climate and the accident process that is considered next, the extant literature on safety climate, anxiety, and accidents has not articulated a role for environmental hazards in conjunction with the other variables.

Cusp Model for Safety Climate

In light of the aspects of safety climate that have not been sufficiently addressed yet, the cusp catastrophe model mentioned earlier was considered once again, this time with specific attention to the safety climate concept (Guastello & Lynn, in press). The instability that is intuitively associated with an accident is actually located in the *change* between the stable states. The bifurcation manifold is actually a *pattern* of instability. Safety climate (in its initial definition) and anxiety are the two primary bifurcation variables of interest in this study. If the safety climate is negative (favoring the occurrence of accidents), a time series of events would actually show times of low accident rates and times of high rates, or workgroups would be split between those with lower and higher rates. Which outcome would prevail would depend on the hazard level. Positive scores on safety climate favor low-accident conditions, but although they are desirable they are not necessarily stable. Positive safety climate can be disrupted by a new hazard or other aspects of load. Reason (1997) noted that work pace by itself can have what amounts to a *hysteresis* effect on accident rates; hysteresis is the irregular oscillation between stable states that is often observed in catastrophe phenomena: An organization might make concerted efforts to reduce accidents including adjusting the work pace demands. Eventually, however, the organization starts to demand higher production output and as a result the accident rates start to zigzag up and down until the control point lands on "up" (i.e., higher risk). Figure 10.7 is a stylized rendition of Reason's illustration (p. 5). It is a view of the cusp surface from the top down. The point of the "<" is the cusp point, and the open side of the "<" is the high bifurcation side of the surface. Although Reason did not invoke catastrophe models as part of his exposition, the movement across the bifurcation manifold is essentially what he was describing.

Tenner (1996) identified a similar phenomenon as a *revenge effect*, which is an unexpected negative consequence that is produced when trying to fix another problem. Higher quality athletic protection equipment apparently lends itself to a revenge effect where the user feels protected and safe, but ends up taking greater levels of risk that exceed the limits of the production equipment. In another example, automobile drivers who report using their seat belts more often also drive 8–15 km/hr (5–9 mph) over the speed limit, as determined by speed monitors that were installed in their vehicles for experimental purposes (Berlin, Reagan, & Bliss, 2012).

Anxiety displays the attributes of a bifurcation variable: The high end of the scale predisposes the individual or group to performance disruptions that could be precursors to accidents or it could predispose the individual or group to closer monitoring of the

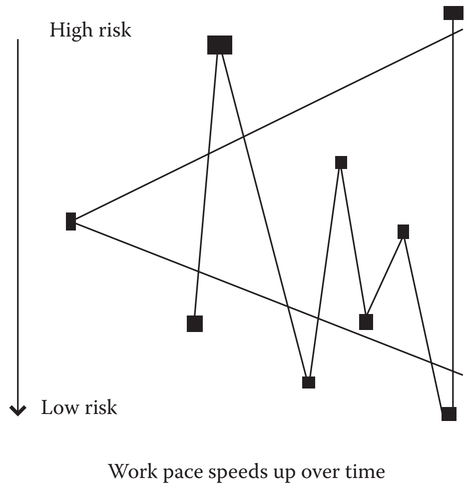

FIGURE 10.7
Hysteresis and risk levels.

environment for threatening signals. Low anxiety takes the individual or group away from either extreme, but the absence of anxiety does not generate a locally stable outcome.

The respective roles of asymmetry and bifurcation variables correspond to what early epidemiologists called background variables and trigger variables, respectively (Fine, 1979; Guastello, 1995). If everyone worked in rubber rooms, stress could produce errors and problems, but injuries would be unlikely. On the other hand, a reasonably skilled worker in a hazardous environment might avoid harm well enough until some unexpected event occurs, or work pace or other demands become too high. Here the demands induce human errors that could have serious consequences.

Catastrophes are often the result of phase shifts, which are critical moments when a situation emerges or self-organizes to produce a result that is not readily reduced to elementary contributing causes. Systems tend to self-organize when their internal entropy levels are high, whereupon events take on a structure without any input from outside the system, such as a manager. For instance, several accident theorists have focused on the build-up of energies in the system that begin with deviations in procedures or materials, carry through further parts of the work process, and have gone out of control by the time the energies transform into an injury to an operator (Kjellén & Hovlan, 1993). Shifts from one stable state to another, as when water turns to ice or energies are in control or out of control, are phase shifts that can be described analytically as cusp catastrophes (Gilmore, 1981; Guastello & Liebovitch, 2009).

Situational specificity plays a substantial role in self-organizing processes (Hazy et al., 2007). Although a process may take on a generalizable structure, the specific variables that are salient in the environment will shape the particular outcome. For instance, if one were to include safety climate as it is generally known as part of a broader array of constructs that are potentially operative—safety *management*, stress from different sources, anxiety, perceptions of danger levels, safety locus of control, and physical hazard inventories—different variables would be more pronounced in some situations than in others. Differences in the tight and loose nature of the workgroups' boundaries also affect the outcomes; contrasts among steel mills, transit operation, and health-care environments have been noted previously.

An investigation into the contributing dynamics of a workgroup and occupational injuries should thus be defined broadly enough to cover the probable range of possible contributing psychosocial and environmental elements, but also anticipate that some contributing elements could be more salient than others even though the underlying dynamic structure is essentially the same across situations. In an illustrative example, 1262 production employees of two steel manufacturing facilities who completed a survey that measured safety management, anxiety, subjective danger, dysregulation, stressors, and hazards (Guastello & Lynn, in press). Instead of using intact survey scales, however, the survey items were factor analyzed to produce four scales that reflected the ongoing dynamics of the facilities: safety management, which was close to the original definition of the construct, anxiety, a cluster of hazards associated with lifting, and a cluster of hazards associated with the use of construction tools. Safety climate and anxiety were tested as two bifurcation variables. The two hazard groups were added together to form a single asymmetry factor. Age and experience were combined into another asymmetry variable. Nonlinear regression analyses showed, for two steel manufacturing facilities, that the accident process was explained by the cusp catastrophe model; the accuracy of the cusp model ($R^2 = .72$) exceeded that of the next best log-linear model ($R^2 = .08$) that was composed from the same survey variables.

Another interesting finding that came to light was that the older and more experienced people were involved in more of the accidents over the same 3-year period. Ordinarily the opposite is true. The interpretation was that the more experienced workers got the more challenging assignments and were thus the ones who were more often exposed.

Of further interest, the R^2 coefficients for the cusp models appearing in Table 10.2 along with the R^2 values for the next best linear alternative measures can be combined with the results of Guastello and Lynn's study to assess the relative efficacy of the cusp model as a theoretical approach for risk modeling. The average value for the cusp models was .63, and the average value for the comparable linear alternatives was .12. The cusp showed an advantage of 5.25:1 in terms of accident variance accounted for by the model.

Safety Culture

The concept of safety climate was eventually transformed into *safety culture* in the semantics of safety officers. The transition took place in the wake of the Chernobyl nuclear reactor disaster in the Ukraine in 1987, when it was noticed that different societies had different views of what constituted a risk and what the risk was worth in terms of societal benefits (Douglas & Wildavsky, 1982; Pidgeon, 1991). Policymakers in the European Union, meanwhile, have been grappling with the task of developing uniform occupational health and safety standards (OHS) for EU employers. Perhaps the most fundamental problem facing transnational policymaking is the extent to which program policies should balance uniform compliance, national interpretation, employers' discretion, and worker participation. Next, there is a need to balance physical environment and process standards. Each of the constituent countries has its own history of labor and safety litigation. These differences will affect the form of the general solution and the norms of compliance (Frick, Jensen, Quinlan, & Wilthagen, 2000).

The agglomeration of organizations through buyouts produces situations where multiple approaches to OHS can exist within one organization. Confusion or diffusion of responsibility could result. For instance, for Swedish organizations, if one organization buys out another, the responsibility for OHS still belongs to the unit that was purchased where the dangerous work presumably is being carried out. Here the OHS objectives could be in conflict with other ill-conceived directives from the holding company's management. The exportation of dangerous work to countries outside the European Union, where norms and expectations for OHS support greater risks to the workers, presents more opportunities to exploit workers on the one hand or diffuse responsibility for OHS on the other. The growing size of the secondary labor market in the United States and European Union also reflect many opportunities for organizations to shirk OHS responsibility by outsourcing to other smaller operators.

OHS management within any one organization has generally stood outside the purview of most management functions. There is now a greater perceived need to integrate OHS objectives into the mainstream of an organization's business concerns. The implementation of OHS programs is thought to benefit from worker participation initiatives, and policymakers at the EU level are placing a strong emphasis on employee participation components to OHS.

Despite the emphasis on participation, however, which is usually associated with the more progressive forms of management, there are strong subcurrents of authoritarianism in management that result in blame-the-worker policies. Wockutch and VanSandt (2000) presented several contrasts in high-profile OHS plans in the United States (DuPont) and Japan (Toyota), although they noticed some curious similarities also: "When there is an oil spill in a US factory we erect a cage around the spill, whereas in Japan they put up a sign warning workers to be careful. The critical fact from this perspective is that in neither place is the oil spill cleaned up" (p. 372).

Swiss Cheese Model

Even a well-designed system can have sensitive areas where safety risks can get through. The goal of safety management is thus to find those sensitive areas and provide additional defensive strategies that will limit the flow of risks. Reason (1997) likened the situation to a series of slices of Swiss cheese that are arranged in a staggered fashion like the squares depicted in Figure 10.8. Although other safety and risk analysis concepts considered hitherto emphasize the role of the individual, management is ultimately responsible for investigating risks and incidents and setting priorities as to what constitutes a risk that requires attention.

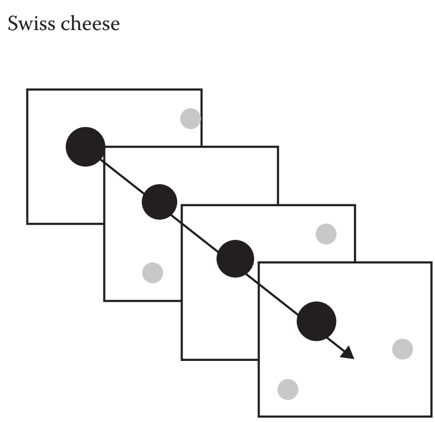

FIGURE 10.8
Swiss cheese model of risk management.

Reason (1997) suggested further that some of the fluctuations in an organization's accident rate could be traced to waffling in priorities that are expressed by various policies. Improper maintenance of systems may be at least as important as the actions of the human at the machine interface; the analysis of several major disasters such as Apollo 13, Three Mile Island, Bhopol Chemical Plant, the Claptham Junction train wreck, and several airplane incidents all showed significant contributions of poor maintenance or maintenance policies.

Resilience Engineering

The concept of resilience in human factors and ergonomics (Hollnagel, Woods, & Leveson, 2006; Sheridan, 2008) poses questions such as: How well can a system rebound from a threat or assault? Can it detect critical situations before they fully arrive? Building resilience into a system requires more than the analysis of accidents in hindsight or implementing a right-minded plan. It involves regular monitoring of the system for its proximity to critical situation, anticipating possible critical situations, and taking action to adapt as necessary.

The stress–strain diagram shown in Figure 10.9 depicts the placement of resilience properties within the broader scope of the system's functioning (Sheridan, 2008; Woods & Wreathall, 2008). Strain is an independent measure. In the low to moderate regions of strain, the outcome is a fairly consistent increase in the system's ability to meet demands of various types. Beyond a certain point, however, more strain is producing diminishing increments of capability in what the theorists call the "extra region." At the extreme of the

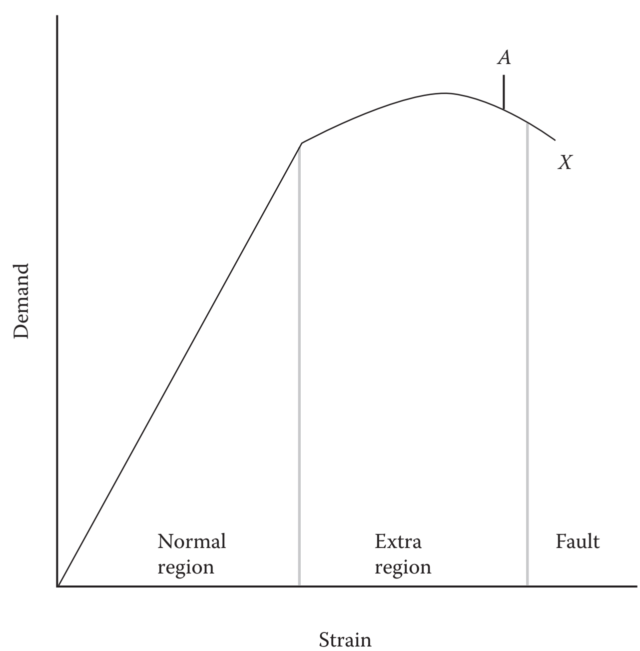

FIGURE 10.9
Stress–demand resilience function.

extra region the system faults. The spot marked "A" would be a good place in the process to introduce an adaptation that would stretch the capability of the system.

Successful adaption to possible risks begins with a calibration of possible risks according to their level of severity and probability of occurrence. In a traditional risk matrix, the outcome categories would range from negligible to marginal, critical, and catastrophic. In a resilience-oriented risk matrix, the outcomes would range from negative (catastrophic in the extreme) to neutral in the middle, to positive at the high end. The task is then to assess all forms of variability or flexibility in the system. Insights regarding the events to be avoided can be forthcoming from understanding the processes that produce unexpectedly positive outcomes (Hollnagel, 2011).

The definition of a preemptive strategy requires a good sense of how a complex adaptive system (CAS) operates. One feature of the CAS essentially involves the use of sensors that take the form of information gathering capabilities that can inform organizational members of changing events in the outside environment or internal operations. The most useful information inflows result in the system-level equivalent of situation awareness. Woods and Wreathall (2008) characterize this form of situation awareness as *calibration*, which is to know where the system is located along the stress-demand function despite changing circumstances.

A second feature of the CAS, which is not independent of the first, is to have a clear sense of how the system is organized for work and information flows. Here one should look for *functional resonance*, which is how the variability associated with one part of the process crosses to the adjacent parts of the process and travels through the system (Hollnagel, 2012; Leonhardt, Macchi, Hollnagel, & Kirwan, 2009). It would then be possible to pinpoint bottlenecks and nonresilient or nonadaptive features of the process. The bottlenecks in this context would not be much different from those associated with parallel cognitive processes (Chapter 6). Nonresilient or nonadaptive features of the process often involve automated functions, which traditionally permit little variation in the process and output and thus work very efficiently for a limited range of circumstances. A resilience event would occur when the human intervenes to make an adaptation that either extends or overrides an automatic process after detecting that a supercritical event was on the verge of occurring. Adaptive automation processes (Chapter 2) can be interpreted as attempts to build a modicum of resilience into a person–machine system.

The third important feature is the production of effective adaptive responses. Recognition primed decision making was suggested as one medium for doing so (Woods & Wreathall, 2008). Some adaptations are more expensive than others regarding the cognitive resources required; for instance, an adaptation at a group level requires a greater expenditure of resources than an adaptation by a single individual. The adaptations themselves are essentially the dynamics of self-organization all over again (Wears, 2010).

As with other self-organized systems, there is an optimal level of variability associated with high levels of performance (Leonhardt et al., 2009) known as pink noise in other contexts. Although there is a folk tendency to think of high performance as synonymous with consistent performance, the relationship is not really true. A certain amount of variability in the system is needed to generate an adaptive response. The grain of truth that remains in the folk logic is that high performing systems, although variable, spend comparatively less time in the very poor regions of the performance spectrum (Guastello, Gorin, et al., 2013).

Before departing this topic, it would be useful to close two more loops in the connection between resilience engineering and other concepts already covered. First, the "extra zone" in Figure 10.9 is comparable to the adaptive range in workload response described

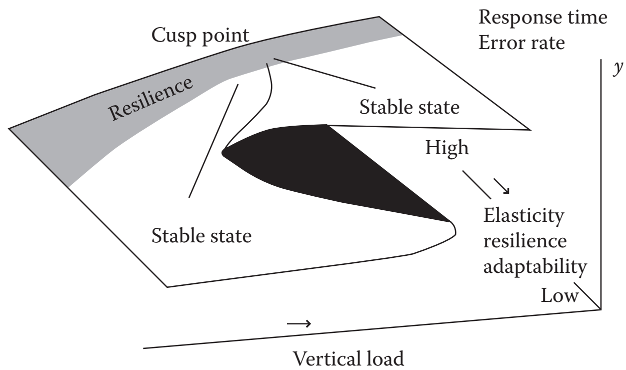

FIGURE 10.10
Cusp model for resilience.

by Hancock and Warm (1989), and the fault region of the demand curve corresponds to what happens when the adaptive range is exceeded. Second, the construct of resilience as it has unfolded corresponds to the rigidity–elasticity principle in the cusp model of mental workload. The cusp model is presented again in Figure 10.10 with markings for the system ranges that correspond to resilience. Here we see again that rigidity has the benefit of controlling system variability up to a point, after which a decisive fault can be expected. Elasticity and resilience, on the other hand, span the region of the surface that contains the cusp point, which is the most unstable point on the surface. It follows that too much flexibility, or resilience capability, can make a system unstable in the long run. Here one should look for features of the system that use too many degrees of freedom (or control) to get a job done well enough when fewer degrees of freedom would produce a result that is just as good.

Accident Prevention Programs

Despite all the effort and money that goes into accident-prevention programs each year, there is scant information available on the relative merits of the known accident-prevention strategies. Decision makers are thus destined to make important decisions based on unreliable or disorganized information. The lack of sufficient program evaluation information was first identified as a problem in the early 1990s (Guastello, 1993), and the number of subsequent researchers who have called attention to the lingering deficit has now become too large to mention them all. As an example of the severity of the situation, Lehtola et al. (2008) investigated the efficacy of accident prevention strategies in the construction industry. Their search produced 7522 titles, but only 5 reports that contained the necessary information to draw conclusions about the impact of a program on accident rates. The situation has become so strained that Pederson, Nielson, and Kines (2012) suggested using qualitative analyses in place of the more traditional combinations of numerical information, control groups, before-and-after comparisons, or interrupted time series. Although qualitative analyses can provide novel insights and useful information, it cannot answer the essential question of how well one type of intervention would perform in comparison to possible others.

This section of the chapter fills that information gap by identifying the salient program types, and by compiling conclusions about how well they worked to allow for comparative evaluation. The sources of review material were scientific journal articles that were identified through hand searches of the literature, computerized literature searches, and other articles identified through leads found in primary sources. Reports were also solicited through announcements that were published in relevant journals in the mid-1990s. As it turned out, the vast majority of material that was generated from the announcements was published somewhere; there was very little material hiding out in file drawers or fugitive literature. Occupational vehicular accidents were included in this expanded review, but nonoccupational vehicular accident studies were omitted. Because theories and technologies evolve and improve over time, this review was limited to techniques that were tested since 1977; earlier studies were included to the extent they pertained to techniques that were still under investigation. The review that follows was built on a review that was first published in *Safety Science* (Guastello, 1993), two interim updates that were presented at conferences (Shannon & Guastello, 1996, 1997), and direct searches of later literature.

There were 10 broad categories of intervention: personnel selection variables, technological intervention, behavior modification, poster and promotional campaigns, safety committees, medical management, near-miss accident reporting, comprehensive ergonomics, other management interventions, and governmental interventions. Studies were included in the analysis if they contained actual accident rates or counts as the dependent measure. As a result, studies that used safe or unsafe behaviors as their criterion were not included in the computation of effect sizes in Table 10.3 because of the strong attenuation effect between unsafe acts and actual injuries as reported earlier. Also, studies that expressed the results in monetary costs of accidents were not included because there are too many variables related to health-care costs, especially rising costs, that only become involved after the accident occurs. The concern here is, in contrast, for the means of preventing the occurrences.

A measure of effect size, D, was computed for all studies listed in Table 10.3:

$$D = [(R_{pre} - R_{post})/R_{pre}] - [(S_{pre} - S_{post})/S_{pre}]. \tag{10.3}$$

R_{pre} and R_{post} are the accident counts within a given period before and after the intervention, respectively. S_{pre} and S_{post} are the raw numbers of accidents before and after the intervention observed for the control group. In some studies the dependent measure was expressed as a rate. Because an accident frequency count is equal to the accident rate times the number of exposure units, a comparable D can be obtained by inserting a rate in place of R_{pre}, R_{post}, S_{pre}, and S_{post}. Unlike the evaluations of other types of safety programs, the personnel-selection studies were based on correlation coefficients. The translation of correlation coefficients into a measure comparable to D is explained later.

Personnel Selection

Personnel selection is a group of strategies that is directed at picking the right job applicants. Ability and personality tests of various sorts are commonly used. For safety applications, the objective is to identify job applicants who are not likely to have workplace accidents. Almost all personnel-selection studies were based on correlations between an employee's safety record and how they scored on tests of interest.

TABLE 10.3

Comparative Evaluation of Safety Programs

Program Type	Number of Coefficient	*D* (%)
1. Personnel selection	70	4.8
Personal maladjustment	21	5.7
Social maladjustment	8	6.4
Alcohol use	10	0.8
Drug use	4	0.0
Cognitive variables	24	6.5
Nonwork accidents (lifestyle)	3	0.0
2. Technological interventions	12	54.4
3. Behavior modification	11	53.1
4. Poster and promotional campaigns	3	–1.0
5. Safety committees	11	33.7
Quality circles	1	20.0
Safety committees, generic	3	36.0
Discussion groups	3	17.5
Joint labor–management committee	2	55.0
Two-group review routine	1	10.4
6. Medical management	6	39.8
Stress management and exercise	2	15.0
Employee Assistance Programs	2	32.5
Medical management	1	29.4
First aid training	1	34.7
7. Near-miss accident reporting	2	0.0
8. Comprehensive ergonomics	9	53.1
Musculoskeletal injuries targeted	2	77.5
All others	7	46.1
9. Other management interventions	5	55.0
Philosophy + Behavior modification for management	1	30.0
Autonomous workgroups	1	161.1
Self-regulation	1	46.0
Management–leadership training	1	71.5
Multifaceted safety campaign	1	–2.6
Multifaceted safety + management	1	24.0
10. Governmental interventions	29	9.7
Finnish national program	2	18.3
OSHA onset, regulations, inspections	5	2.6
OSHA industry-targeted inspections	12	18.9
Increase in workers' compensation costs	3	–46.3
Workers' compensation, experience method	3	25.0
Universal precautions for healthcare	1	0.0
Australian regulatory system	1	15.0
Constructions fatalities, USA	2	26.8

For correlational studies that were taken from the earlier review (Guastello, 1993), the correlations were transformed to percentage increase in relative efficiency by using utility expectancy tables. The tables required two additional parameters: the selection ratio and the base rate of success, which were set at .20 and .70, respectively. A selection ratio of .20 signifies that the organization can afford to hire only the most promising 20% of their job applicants. A base rate of .70 is equivalent to defining success on the job as an individual not incurring an accident in 4 years of exposure, based on the average accident rate of 8.30 for all private sector U.S. industries as reported by the National Safety Council (1989). For correlational studies dated after the first review, utility was calculated using Jarrett's (1948) continuous utility formula and an exponential distribution; the assumptions of a selection ratio of .20 and a base rate of .70 were retained.

Six types of personnel-selection variables have been tested: personal maladjustment, social maladjustment (which would include external beliefs about accident control), alcohol use, drug use, cognitive variables, and reports of nonwork accidents. Eight coefficients for personal maladjustment were available from the earlier review (Guastello, 1993). Arthur, Barrett, and Alexander (1991) produced seven, and six more were reported by Barofsky and Smith (1993), Jacobs, Conte, Day, Silva, and Harris (1996), and Lardent (1991). These variables included stressful life events, anxiety, a measure of distractability, trust, tensions, insecurity, even temperedness, concern with being on time for work and other life events, and regard for authority.

Seven coefficients for social maladjustment were available from the earlier review (Guastello, 1993). Jacobs et al. (1996) added another test of the safety locus of control variable. Arguably, there is a fine line between personal and social maladjustment. Nonetheless the average (unweighted) effect sizes for those two groups of variables were 5.7 and 6.4, respectively. In other words it appears that it is possible to reduce accident rates by 5.7% by using personal maladjustment variables in a selection context.

The studies of maladjustment variables, however, were always postdictive in nature, rather than predictive. Maladjustment could just as easily be the result of working too long or not enough in poorly controlled environments where ineffectual safety management reigns. Even if the correlations between individuals' personality characteristics and their accident rates were to be regarded as if they were as good as predictive values, personnel-selection techniques rank among the least effective methods of accident control of all the available options.

Self-report inventories or assays of blood samples are used to measure alcohol use. If such measures are to be used to select personnel, it must be assumed that alcohol use off the job was not associated with accident involvement while working. The eight studies reported in the earlier review (Guastello, 1993), and the two studies uncovered since that time (Dawson, 1994; Holcom, Lehman, & Simpson, 1993) show very little association between the two variables. The average effect size for this personnel-selection technique was 0.8%.

Drug use has been measured by self-report inventories or urine tests. Urine samples are usually tested for marijuana and cocaine, and the percentage of people testing positive for only marijuana ranges from 67% to 93%, depending on the industry (ACLU, 1999). Once again, if such measures are to be used to select personnel, it must be assumed that drug use off the job affects accident rates at work. The two studies reported in the earlier review (Guastello, 1993), and the two studies uncovered since that time (Feinauer & Havlovic, 1993; Holcom et al., 1993) found no association between the two variables. Manufacturing organizations that adopted urine-testing programs were found to have obtained no benefit from the program compared with other organizations in the same industries and geographic region (Feinauer & Havlovic). The effect size for this personnel-selection strategy is 0.0%.

Cognitive tests appear to have a better track record. Auditory selective attention tests showed positive results in one study (Arthur, Barrett, & Doverspike, 1990). More broadly defined tests of selective attention showed positive results in 13 other occupational samples (Arthur et al., 1991), and tests of field dependence and field independence showed positive results in 9 occupational samples (Arthur et al., 1991). General cognitive ability did not show a relationship to accidents in a large sample of bus drivers, however (Jacobs et al., 1996). The average effect size for the 24 correlations was 6.5%.

The last group of personnel-type studies was reported by Salminen and Heiskanen (1997). Three extensive stratified samples of Scandinavians responded to a survey in which they were asked about their involvement in work accidents, home accidents, and public situation accidents. The samples were taken in 1988, 1990, and 1993. The hypothesis was that there would be such a relationship, which would in turn mean that accident proneness could be traced to a person's lifestyle. The average correlation between work and other accidents was .02 in each sample, thus rendering an effect size of 0.0%.

The emphasis on individual-level explanations for occupational accidents (and illness) continues to be fueled by claims that the same people who are involved in accidents are also those with higher absenteeism rates. Verhaegen (1993) reported, however, that no such correlation exists in an economically suitable environment. Furthermore, it is only when the environment is economically degraded that one might observe correlations between personality measurements and accidents; an economically sound workplace would leave no room for such individual differences to influence risk. According to Quinlan (1988) the unwarranted individual focus is an outgrowth of practices in industrial psychology, dating back to the scientific management era, where psychologists are called in (and paid) by management to solve so-called work problems. The right answer for the client was one that blamed the victim in some way. Such viewpoints overlooked collective causes of accidents and management's role in them, particularly when the pace of production is stepped up beyond safe working limits.

Technology Interventions

Technology interventions may be broad in scope, such as introducing robots to perform dangerous work (Karwowski, Rahmi, & Mihaly, 1988), redesigning an entire manufacturing facility (Harms-Ringdahl, 1987; Kjellen, 1990; Mohr & Clemmer, 1989), outrigging loggers with improved personal protection gear and machine guards (Klen & Vayrynen, 1983), or introducing vessel-tracking systems into a maritime operation (LeBlanc & Rucks, 1996). Technology interventions could also be more event-specific, but nonetheless valuable: preventing needlestick injuries in hospitals with a new needle cap design, using visible tent stakes on camping expeditions, redesigns of oil-drilling equipment, rollover protection for farm vehicles, and introducing a barrier between train tracks and a roadway (Zwerling et al., 1997). Of all the occupationally traceable deaths to astronauts reported by Peterson, Pepper, Hamm, and Gilbert (1993), 47.1% of the known causes have been eliminated by improved equipment or materials, such as the O-rings in the construction of the space shuttle. The average unweighted effect size for these contributions is 54.4%.

Before continuing further, it is noteworthy that robot systems have the potential to reduce injuries, but they also have the potential to create new hazards because of the fast and forceful motions of the machine parts. Relevant remedies include emergency kill switches, radar detection of humans in the work envelope, and the placement of control stations for the humans well outside the machine's work envelope (Karwowski et al., 1988). Comparable values of *D* for this group of technological interventions are not available, however.

A new genre of technology intervention involves software solutions for accident prevention, and a crop of examples has come to the foreground recently in conjunction with fatigue-related accidents. It was noted in Chapter 9 that fatigue, the loss of work capacity over time, can arise from working too long on a mental or physical task or sleep deprivation. The latter case has been a concern for accident prevention in the transportation industries—long-distance truckers, railroad workers, and airline pilots—but also health-care settings where emergency response and extra-long work shifts are required to respond effectively to critical situations are the norm. One technology for addressing the matter is the use of "universal" work schedule controls that require operators to have sufficient time between shifts to get enough sleep so as not to incur sleep debt (Balkin, Horrey, Graeber, Czeisler, & Dinges, 2011; Dawson, Noy, Härmä, Akerstedt, & Belensky, 2011). Schedules are based on algorithms that combine wake–sleep cycles and body temperature cycles in different ways to predict susceptibility to fatigue events, slow response time, and errors. At present, there is no strong evidence that one algorithm is better than another for this purpose. The effectiveness of algorithms for schedules is limited by whether the individual actually obtains the necessary hours of sleep when there is opportunity to do so.

The translation of fatigue conditions based on sleep sufficiency into accident rates requires the analyst to separate sleep deprivation, time on shift, and workload effects such as traffic density (Gander et al., 2011). In addition, there is the observation that errors could be more likely in the early part of the shift and not necessarily toward the end. Furthermore, although errors can be serious enough, translation into accidents is mitigated by all the usual factors. As a result, the evaluations of scheduling algorithms that could provide effect sizes for Table 10.3 have not been forthcoming.

New technologies that are being explored for motor vehicles include EEG monitors, percent eye closure with eye-tracking performance monitors and adaptive workload systems. Concerns are intrusiveness, data loss in the online hook up, calibration, ability of the technique to make predictions of near-future driving quality based on the metrics, and conflicts between readings of fatigue events compared with nonfatigue driving errors (Balkin et al., 2011). For instance, alert drivers make more micocorrections for road curvature for maintaining lanes than fatigued drivers, while fatigued drivers make more macrocorrections; yet a macrocorrection could be a legitimate response to a road hazard such as a bad move by another driver. The efficacies of these techniques are considered promising by their proponents, but not yet established. A lot will depend on what decisions are made by, or based on these new data collection devices and what actions are taken as a result.

A different type of technological intervention takes the form of software to assist management in allocating safety resources (Shakioye & Haight, 2010). The thesis is that risks are dynamic over time, and not all safety management concerns can be attended simultaneously. Thus, it would be advantageous to predict risky events in the near term from the allocation of time spent on different classes of activities, and forecast which combination of activities would be most effective in the future time interval. The four classes of activities incorporated into the model were (a) safety awareness and motivation, (b) skills and training, (c) tool and equipment design and implementation, and (d) other equipment-related activities (p. 47). The central algorithm was a 15-variable polynomial function, which was calibrated on 30 weeks of incident and safety resource allocation data. It was not clear whether the incidents were all reportable injuries, or included near-miss events. The R^2 for the model was .70 predicting incidents; R^2 adjustment for the number of variables in the model was .36. The model recommended that 11% of person-hours (a forestry operation) be allocated to safety activities to meet the organization's safety targets. There was no particular claim that the tool provided a reduction in accident rates when used in this fashion. The study did indicate, however, that the safety targets could be achieved with a 4% reduction in person-hours invested in safety tasks by using the activity-switching recommendations from the program.

Behavior Modification

Behavior modification programs are based on operant learning theory. Safety officials, who could be workgroup supervisors, identify specific desired behaviors, such as wearing one's personal protection equipment properly, or compliance with particular safety rules. The identified behavior could be the number of reportable accidents themselves. The group of employees would then receive feedback on their behavior, often at the group level, with the goal of 100% rule compliance or zero accidents for the workgroup in a unit of time. Feedback is often presented in the form of highly visible posters; informational feedback acts as a form of reward to reinforcement in these circumstances. Some implementations of the program may also include a lottery for prizes, wherein only employees in units that have met the goal receive chances to win. Other versions will include censures for employees who repeatedly violate safety rules. McAfee and Winn (1989) published a thorough compilation of programs and types.

The previous report on the effectiveness of behavior modification programs for controlling actual accidents (Guastello, 1993) showed an average effect size of 38.6% for six implementations. A seventh program that had been in place for over 10 years in a uranium mine brought the average effect size up to 59.6%. Since that time four more reports have been collected (Gregerson, Brehmer, & Moren, 1996, giving two examples; Martinez & Brito, 1994; Zwerling et al., 1997). The average effect size for the 11 studies was 53.1%.

Poster Campaigns

Safety posters are frequently found in industrial sites that typically give generic advice to employees to be mindful of safety matters. Better posters will give specific safety information. In the previous report, one safety poster program had a net positive effect on workplace safety, whereas another actually had a negative impact. In the latter case, the control group improved its safety performance faster than the poster group. Another similar case of negative impact was reported by Gregerson et al. (1996). The average effect size for the three poster studies is thus negative, –1.0%.

Safety Committees

This category is new since the previous report, although safety committees of various sorts have been used for an indefinitely long time in industry. Previously, only one report had been identified in which the safety committee was characterized as a *quality circle* (Saarela, 1990). A quality circle is a committee of employees who perform similar types of work who meet voluntarily to solve product quality, productivity, and cost-reduction problems. Past research has shown that the quality circle technique has been successful in those objectives in addition to improving the quality of work life and reducing absenteeism (Marks, Mirvis, Hackett, & Grady, 1986).

Since the initial report several results for broader varieties of safety committees have been uncovered (Havlovic & Feinauer, 1990; U.S. Department of Labor, 1988; Zwerling et al., 1997). The variety with the greatest power to enforce recommendations appears to be the joint labor–management committee. The next most influential is the employee safety committee. The least effective of the known varieties is the discussion group.

Menckel, Carter, and Hellbom (1993) reported an unusual variety of group procedure that involved two decision teams. One group was responsible for investigating hazardous situations and soliciting corrective ideas. The other group was responsible for effective implementation of corrective procedures. Another underlying goal of the procedure was to develop safety knowledge and prevention skill among a wider range of employees at the site. In addition to the effect size of 10.4% for accident reduction, Menckel et al. also reported a 35% reduction in accident severity for this procedure.

When all varieties of safety committee are considered together, their average effect size is 33.7% based on 11 reported outcomes. The highest marks to go the joint labor–management committee with a 55.0% reduction in the accident rate.

Medical Management

This category started with two studies of stress management and exercise programs in the first review. Since then four new reports have surfaced on the effectiveness of an employee assistance program (EAP; Nadolski & Sandonato, 1987; Wickizer, Kopjar, Franklin, & Joesch, 2004), medical management (Zwerling et al., 1997), and first aid training (Zwerling et al., 1997) in reducing accident rates.

As a general rule, employee assistance programs address a wide range of psychological issues, of which stress and anxiety are a part. The study by Wickizer et al. (2004) was one of the five programs from the construction industry evaluated in Lehtola et al. (2008). It was actually characterized as a drug-free workplace program that included "a formal policy; drug testing; treatment; worker assistance; education of workers, supervisors, and managers; information; education' facilitation (financial incentives); and enforcement (drug testing)" (Lehtola et al., 2008, p. 80). By all appearances, it was an EAP added on top of a drug-testing program; the EAP feature was necessary to be compliant with the provisions Americans for Disabilities Act (ADA) of 1990 for the proper response to employees with substance abuse problems. Given that alcohol-related problems greatly outnumber drug-related problems in the general population and fall under the purviews of APA and EAPs, it would follow that most of the effective interventions for individuals were alcohol-related. There is a stark contrast in effectiveness for this program (D = 25.5%) and the effectiveness of drug testing at the point of personnel selection. This observation brings us to the broader issue that complex interventions by definition have a lot of parts, and some parts could be producing a much larger portion of the positive effect than other parts.

The medical management program that Zwerling et al. (1997) reported was designed to help employees make self-directed improvement on their personal health issues. One would guess that such an intervention had an impact on stress and particular health issues that affected fitness for work; further research would be needed to determine how the management of particular health affected any known intermediate variables that are already known to affect accident rates.

First aid training has an interesting effect on reportable accidents. If more employees knew how to respond to low-grade occurrences and had kits available, then off-site medical attention would not be necessary so often. By definition, the incidents would not be reportable on the OSHA-200 form.

Hurrell and Murphy (1996) noted that stress-related illnesses comprised more than 11% of worker compensation claims during the 1980s. They regarded stress management programs as "secondary interventions." A primary intervention would alleviate the source of stress in the workplace, rather than shift the burden of responsibility to the employee. Nonetheless, the medical management group of interventions produced an average effect size of 26.7%.

Near-Miss Accident Reporting

The near-miss reporting program is based on the principle that for every real accident that occurs, approximately 10 near misses have also occurred. By investigating 11 times as many incidents, a greater number of ideas for preventative measures can be generated. Two examples (Carter & Menckel, 1985; Menckel & Carter, 1985) from the earlier report (Guastello, 1993) did not result in any reduction in accident rates. The probable explanations for the lack of results might be traced to the quality of corrective ideas and whether the best ideas were actually implemented. On the other hand, the Federal Aviation Administration routinely examines near-miss air disasters for possible explanations and prevention of actual disasters. The same is true for chemical process industries (Gnoni et al., 2013). The probable difference between a successful and mediocre near-miss program is the extent to which action is taken on the basis of the findings from the near-miss incidents. Although new strategies for the design and implementation of a near-miss analysis program have been reported, no new evaluative statistics are available yet that produce a measure of *D*.

Perhaps one of the barriers to success for near-miss accident reporting programs is the tendency for individuals who evaluate risks and outcomes to equate a near-miss incident with a success, rather than a system failure (Dillon & Tinsley, 2008). Similarly, managers of projects with a serious near-miss incident are also regarded as approximately as successful as managers of similar projects without a near-miss; both are regarded as distinctly better than managers of projects with an actual failure. Part of the explanation for this thinking, according to Dillon and Tinsley, lies in the gap between statistical odds of failure and subjective odds of failure. In one of their experiments participants were given stated odds of event and asked to indicate whether they would engage in the target activity, for example, driving through an area where a hurricane is expected to hit relatively soon. Those who were also provided information about near-miss events gauged the situation as not so dangerous.

Comprehensive Ergonomics

Comprehensive ergonomics programs involve a full-scale assault on any and all human factors and ergonomics problems in the workplace. Each organization will have its own list of problems requiring attention. Problems may be identified through use of a survey that is filled out by employees in all work locations (offices are typically excluded), or identified by safety engineers or outside auditors who make a thorough search for potentially hazardous situations, especially where human factors could be involved. Some organizations do not use an audit; instead they implement every program they possibly can, which, with a little luck, could cover a substantial range of human factors or ergonomic situations.

Studies involving OSHA inspections are not included in this category. OSHA purview has historically paid little attention to human factors issues, although they have established allowable limits to noise and heat. Most of the inspections focus on structural issues such as proper electrical service for machines, the presence of machine guards in all the right places, and the availability of proper personal protection equipment. In the last decade, however, OSHA has expanded its purview into ergonomics-related consultations and inspections (OSHA, 2002) on an advisory rather than an enforcement basis.

The first report of this series on accident-prevention programs (Guastello, 1993) identified three examples of comprehensive ergonomics programs, although the specific contents of one of the interventions was unclear. Four new examples have surfaced since that time (Doos, Backstrom, & Samuelson, 1994; Helander & Burri, 1995; Lipsomb, Li, & Dement, 2003; Macek, 1987). The latest arrival (Lipscomb et al., 2003) was an evaluation of a program that was motivated by compulsory legislation directed toward preventing falling injuries in the construction industry. Because the program covered a range of activities usually associated with comprehensive ergonomics, it was included in this category.

Musculoskeletal injuries comprise a large portion of total occupational injuries as noted in Figure 10.1, and up to 67% of incidents in some specific facilities. Not surprisingly, some comprehensive ergonomics strategies have targeted musculoskeletal injuries, and two new reports are included in a new subcategory in Table 10.3. Moore and Garg (1998) reported a "participatory ergonomics" program in a large meat-processing industry, which historically has been one of the most hazardous industries in the United States. The particular facility had a baseline rate that was 20% less than the industry average at the time, and a lost-time incident rate that was 25% higher than the industry average; musculoskeletal injuries were mostly in the subgroup of lost-time injuries. The program itself involved auditing, targeting, and correcting as many sources of musculoskeletal injury as possible. Employees were involved in all phases of the program, and an ergonomics specialist was primarily responsible for its design and implementation. The interventions included the invention of new patented mechanical devices for certain meat packing operations. The program began in 1986 in which there was an immediate drop of 50% of lost-time injuries relative to 1994; there was a 17% drop (possibly spontaneous) from 1994 to 1995. Lost-time injuries dropped to an average of 16% of baseline for 1991–1993. Thus, the D value for this program was 76.0.

A similar report was published a year earlier in an industry that was centered on machine sewing tasks (Halpern & Dawson, 1997). Again musculoskeletal injuries, which had a substantial representation of repetitive motion injuries, were targeted. The program had most of the features of the meat packing program, but without patented machinery, and a noted use of force measurements and discomfort surveys. Ultimately there was a 79% reduction in lost-time injuries recorded for 1995–1996 compared with the baseline of 1992–1993 years. There was also an 11.1% reduction in all Workman's Compensation claims, which included the nontargeted sources of injury.

The average effect size for the seven examples is 46.1%. It is noteworthy, however, that the effectiveness of a comprehensive ergonomics program is limited to the extent that the organization actually carries out the recommendations from its diagnostic analysis. Depending on the complexity of the list of recommendations, organizations might need more or less time to complete the improvements or for the improvements to realize their full effects.

In what could be an interesting twist on the theme of comprehensive ergonomics, some insurance company representatives are now advocating accident prevention through better product designs (Braun, 2008). The former emphasis on "human error" as the culprit has apparently run its course and the sources of human error are now receiving attention. The history of the automobile feature was given as an example of what could be possible in other settings. Of course an evaluation of a new product design might eventually register as an effect size in the category of technology intervention. The source of design ideas, however, might reasonably emerge from information uncovered by comprehensive ergonomics programs.

Other Management Interventions

This group of interventions was directed at improving management effectiveness, rather than the quality of person–machine or person–environment interactions directly. Woodhill, Crutchfield, and James (1987) reported a form of behavior modification whereby management's bonuses were tied to their safety performance. The program was coupled with a widely disseminated corporate philosophy that placed precedence on workplace safety. A learning theorist would probably remark here that the performance–bonus link was responsible for most of the accident-reduction results and that philosophy exercises might help managers understand clearly the contingencies of reinforcement; it is doubtful the philosophy alone would have had much impact.

The autonomous workgroup intervention (Trist, Susman, & Brown, 1977) had an enormous impact on safety: Accident rates improved 161.1%. This unusual level of improvement was the result of substantial progress in the intervention group combined with a deteriorating safety record in the control group. In an autonomous workgroup intervention, work that was once organized as a fixed sequence of specialized processes, as in an assembly line, is now given over to a group. The group establishes its own sequence of people and tasks, including the supervision or managing functions that used to be allocated to a supervisor. In other words, management was more or less removed from the situation. Rees (1988) reported a self-regulation intervention that seemed to fit the autonomous workgroup definition also.

Fiedler, Bell, Chemers, and Patrick (1984) found a substantial safety advantage from an intervention designed to enhance managers' leadership skills and assign managers to workgroups that were better suited to their management style. The control group did not show any such progress. Although the specifics of their leadership training are outside the scope of this book, it would be fair to say that there is an alternative explanation for the results: A good many workers might have been simply happy to get rid of their managers and start the relationship over fresh with a new one. A careful test of this particular point has not been reported yet. Altogether, the average effect size for the four studies in the category of other management interventions was 77.2%.

A new addition to this group of interventions (Spangenberg, Mikkelsen, Kines, Dyreborg, & Baarts, 2002) was a "multifaceted safety campaign … including attitudinal and behavioral aspects (e.g., newsletter, best practices, safety inspections, financial safety award, themes on injury risks); information; facilitation (feedback); enforcement (inspection)" (Lehtola et al., 2008, p. 80) that was directed toward construction workers in Denmark. The safety record 3 years after the intervention was regrettably a bit worse than the record in the three years prior to the intervention. Another new addition, which also entangled multiple intervention types plus other nonsafety management changes (Bunn et al., 2001, reported by Robson et al., 2007) produced a D value of 24.0 for all reportable incidents.

One could regroup the five studies into two groups: the autonomous workgroups that involve the elimination of management roles (average $D = 103.6$) and the three that target the existing management (average $D = 33.0$). Ironically, the values of D favor autonomy versus management. The number of reports is relatively small, however, and a considerable amount of variance in the effectiveness of either subgroup of strategies has yet to be recorded.

Governmental Interventions

Twenty-seven intervention results were obtained in this category in addition to the two that were previously on record for Finnish government initiatives (in Guastello, 1993). Most of the reports pertained to some aspect of OSHA. They included the onset of OSHA (Butler, 1994), which actually met with a slight increase in nationwide accident rates, inspections for compliance with regulations (DiPietro, 1976; Mendeloff, 1979; Robertson & Keeve, 1983; Smith, 1976, 1979), and attempts to influence accident rates by various manipulations of worker compensation costs to the employers (Bruce & Atkins, 1993; Butler, 1983; Robertson & Keeve, 1873). Specific regulations were evaluated in two cases. A law requiring rollover protection equipment on farm vehicles was particularly effective (Zwerling et al., 1997); because this intervention involved a specific technological requirement, the results (D = 92.0) are classified with technological interventions. OSHA's Universal Precautions for preventing needlestick and sharps injuries in hospitals (and other exposures to blood-borne pathogens) had no effect on reportable needlesticks, sharps, or other injuries (Guastello et al., 1999). The onset of a regulatory agency in Australia had a net positive effect, according to Gun (1993).

The two latest additions to the category resulted from legislation designed to prevent fatal falls in construction (Derr, Forst, Chen, & Conroy, 2001) and fatalities in the excavation phase of construction (Suruda, Whitaker, Bloswick, Philips, & Sasek, 2002). The rate of fatalities from falls dropped 3.5%, and the rate of fatalities in excavation dropped 50.0%.

Emergency Response

The mainstay of what we know about emergency response comes from public situations that have experienced earthquakes or other natural disasters. The essential aspects of emergency management are considered next with the understanding that organizations experience internal emergencies also, and that emergency response is usually part of someone's job and one that has a great deal of hazardous exposure.

Hazard Perception

In an occupational setting, one of the important differences between experts and novices is that experts know how to recognize signs of a hazardous condition, and novices often attend to all the wrong cues. When the tsunami struck Southeast Asia in late 2004, the unusual behavior of the water, as seen from the shore, gathered a lot of attention, but too many people had no clue that the receding water would be followed by a huge rush of water over the shoreline very soon afterward. Figures 10.11 to 10.13 are photographs taken by an anonymous photographer, which were circulated through the Internet without further credit. Some important group dynamics are taking place in the photos. In Figure 10.11, the crowd is watching. In Figure 10.12, some people are backing away, while others are left behind. In Figure 10.13, the crowd is mostly running away, but some people insisted on staying and watching what would happen next; they found out. One can only wonder about the effectiveness of the umbrella.

Although it was a post hoc interpretation of events, it was possible to interpret them nonetheless as a result of a cusp catastrophe model (Guastello et al., 2008). The two stable states were "continue gawking" and "running away," and there was a discontinuous change between the two states. The asymmetry parameter was knowledge of the situation; some people had prior knowledge of what a tsunami would look like when it unraveled,

FIGURE 10.11
Risk perception of the crowd before the tsunami reached the shore. (Reprinted from Guastello, S. J., Koehler, G., Koch, B., Koyen, J., Lilly, A., Stake, C., & Wozniczka, J., *Theoretical Issues in Ergonomic Science*, *9*, 95–114, 2008. Photograph in the public domain.)

FIGURE 10.12
Risk perception of the crowd when the tsunami reached the shore. (Reprinted from Guastello, S. J., Koehler, G., Koch, B., Koyen, J., Lilly, A., Stake, C., & Wozniczka, J., *Theoretical Issues in Ergonomic Science*, *9*, 95–114, 2008. Photograph in the public domain.)

while others were clueless or did not make the connection between what they saw and any public service announcements that they did receive.

The bifurcation factor is social facilitation. The very presence of a crowd affects judgment. One might respond quickly if other people were not around, but seeing how other people were responding affected the responses of the target individual. As we saw in other arousal–performance dynamics, the presence of other people is arousing also and predisposes people to one or another response very distinctly. Thus, in the presence of danger signs, groups are slower to react than individuals; the nonreaction of some delays the reaction of others.

FIGURE 10.13
Risk perception of the crowd after the tsunami reached the shore. (Reprinted from Guastello, S. J., Koehler, G., Koch, B., Koyen, J., Lilly, A., Stake, C., & Wozniczka, J., *Theoretical Issues in Ergonomic Science*, *9*, 95–114, 2008. Photograph in the public domain.)

Time Ecologies

Emergency response (ER) systems, like other types of public policy, operate on multiple time horizons or *time ecologies* (Koehler, 1999). At the slowest time horizon, senior management is identifying and interpreting risks of an outbreak of a natural or other type of disaster. The time horizon is occupied by foresight and action planning and could extend for many years. Sociopolitical systems that fail at this level are seriously impaired when an actual disaster strikes and the focus of attention shifts to the more immediate time horizons (Comfort, 1996; Pauchant & Mitroff, 1992; Reason, 1997), as when hurricane Katrina struck New Orleans in 2005 (Cigler, 2007; Derthick, 2007; van Heerden, 2007).

The midrange time horizon initiates when the disaster actually strikes. The horizon for rescuing people from an earthquake region is about 5 days (Comfort, 1996). The majority of people rescued who survive are rescued within the first 2 days; the odds of survival given rescue decay sharply afterward. Meanwhile, all the shock elements of unplanned physical locations, time of day, availability or impairment of medical or transportation resources, fires and explosions, and a generally fast-changing situation require instant adaptive responses.

The chaos of the situation should be taken literally in the NDS sense, with sensitivity to initial conditions figuring very prominently in the unfolding of events (Farazmand, 2007; Koehler, 1995, 1996). Many, sometimes hundreds of formal and informal organizations, and citizen groups mobilize their capabilities over the short time horizon (Comfort, 1996; Morris, 1906; Morris, Morris, & Jones, 2007). Coordination among them is especially challenging, and complex systems in a high level of entropy can produce surprises of their own accord (McDaniel & Driebe, 2005; Sellnow, Seeger, & Ulmer, 2002). The skill for managing chaos is thought to be in short supply in the general population of management personnel (Guastello, 2002). Nonetheless, Morris et al. (2007) gave high marks to the U.S. Coast Guard and U.S. Air Force for their coordinated actions in the Katrina disaster, as did Morris (1906) to the mayor of San Francisco for managing that historic earthquake.

While the activities at the midrange time horizon are getting started, events are occurring at the microlevel time horizon, operating at the scale of hours and minutes. Koehler

(1996) emphasized the critical and problematic nature of timing at this level of activity. For instance, one decision-maker can ascertain that a hospital emergency room has a certain amount of carrying capacity at a particular moment, and then dispatch some casualties to that hospital. By the time the casualties arrive, other decision-makers had the same idea and dispatched more casualties to the same location, producing a bottleneck. Other critical events are connected to the discovery of new casualties or the prevention of concomitant disasters, such as fires in the wake of an earthquake, or the change in the path of a forest fire caused by a sudden shift in the winds. Human communication and the physical movement of people and equipment are not always fast enough to compensate. The psychological representation of time by disaster respondents and victims is strongly constricted to the needs of the present moment. The ability to see the future, even in the short horizon of a disaster response, is greatly impaired.

Situation Awareness and Sensemaking

Situation awareness is usually regarded as a *process* of perception and interpretation of events that can be assisted by technology, rather than a particular outcome (Durso & Sethumandhavan, 2008; Endsley, Bolté, & Jones, 2003; Wickens, 2008), as discussed in Chapter 6. Communication technologies are vital to an effective ER. For instance, satellite photography can provide ground personnel with information about which areas are affected to what extent and which escape routes are operable and which ones are not. Cellular phones can keep ER personnel closely connected—until the battery dies.

Sensemaking (Weick, 2005) places joint emphasis on the process of gathering relevant information and the cognitive integration process that occurs shortly afterward. Expectations that are only based on known information can produce some automatic actions that misfire if the interpretation of the situation is wrong. Preparedness for the unknown, surprising, or emergent events could produce more advantageous results. The Centers for Disease Control and Prevention's correct diagnosis of West Nile virus was only obtained after the center became aware of laboratory tests that did not fit the original hypothesis and new information about West Nile virus that was not previously on record. Arrival at the correct diagnosis was facilitated by coordinated communications among the responding agents.

The center's experience raised the issue of how best to prepare for an emergent disease epidemic or bioterrorist attack. One does not prepare for the new disease exactly. Rather one prepares a reasonable *strategy* for finding out what it is and formulating a coordinated response with relevant agents.

ER decisions are also dynamic in nature: The decision made at one moment affects the options available soon afterward. There are other group dynamics to consider in conjunction with emergency response, notably group size and coordination, which are considered in Chapter 13 along with other practical implications of complex systems.

The principles of resilience and adaptive responses introduced earlier in this chapter are highly relevant as well. An effective ER actually requires an interplay between well-rehearsed scenarios and responses and recognizing when the original plan does not cover what is happening at the moment. For instance in the first responses to an earthquake or massive flood, the first steps are to shut off the gas lines to prevent fires, shut off electricity from the power plant feeding into damaged transformers, and limit telephone communications in and out of the affected area. The next steps are to identify and interrupt domino sequences and launch rescue teams. Other task forces are then simultaneously working on restoring or importing essential services. The chaotic elements of the disaster present a challenge to recognition-primed decision making, but a wide repertoire of the latter is essential. For this reason, wise municipalities in high risk areas stage practice drills.

DISCUSSION QUESTIONS

1. Consider a situation where someone dove into shallow water and suffered a spinal cord injury as a result. Construct a chain of events leading up to this incident. Can this same situation be conceptualized as a factorial model? How about a fault tree?
2. Consider the Claptham Junction train wreck. Does the Petri net concept convey anything differently from a standard fault tree analysis? Is one method more informative than the other in this situation?
3. Retrace the major disasters at the Three Mile Island nuclear facility or the chemical factory in Bhopol. What errors occurred at the human–machine interface? Was stress a factor in any known way? Did management contribute to the problem in any way? Hint: It would be necessary to look up key documentation on these incidents. Reason (1997) recommended Kemeny (1979) with regard to Three Mile Island and Meshkati (1989) with regard to Bhopol.
4. In January 1986, the space shuttle Challenger exploded shortly after liftoff, killing all persons on board. The technological culprit was a failure of O-rings, which were simple sealing devices in hydraulic equipment. Was that the only safety factor that was involved?
5. In February 2003, the space shuttle Columbia exploded when it reentered Earth's atmosphere after a 16-day mission. NASA and other investigators are trying to figure out why the incident occurred. What can you uncover about the accident in the categories of technology or engineering, operator error, stress, or management error? It may be desirable to revisit this question again after reading the final chapter on human factors in outer space.
6. The current culture of technologists often takes the position that the vast majority of accidents or other unwanted events are the result of human error, and therefore new forms of automation should be developed to minimize human input. Is this a wise perspective?

11

Human–Computer Interaction

Many of the principles of human factors that were discussed up to this point concerning controls, displays, and work under stress apply to human–computer interaction as they would to more conventional machines. At the same time, the development of computer systems has generated numerous phenomena that did not have counterparts in the age of conventional machines. Thus, this chapter on human–computer interaction captures the issues that are germane to the special features of computer-based systems that distinguish them from other machines.

Not making matters any simpler, modern industrial equipment has become increasingly computerized so there is only a small difference between the two types of machines. In actuality, however, there is a critical difference in that the modern machine is still doing work that has a physical effect on materials somewhere, whereas the computer in the strict sense is moving symbols around and not much else. In the not-so-strict sense, there is an interface between the information component of a system and the electromechanical component that makes physical events happen; the programmable lathe in Chapter 2 is one of many examples. Furthermore, there is a growing use of networks of computer programs where a manipulation on one affects the operation or database of another; this class of complex systems is considered in Chapters 12 and 13.

It has been challenging to prepare textbook chapters on the topic of human–computer interaction during the past 30 years because the technology changes rapidly. Not only do HFE efforts lag behind the development of products, the technology changes before HFE research has run its course on a generation of equipment, often rendering the HFE research findings pointless. For this reason, Newell and Card (1985) encouraged human factors specialists to become involved in the early stages of the design processes much more often than they had been before.

In light of the historical developments, this chapter considers the controls and displays that have been most prominent in human–computer interaction. The next section contains some observations about the changing nature of the human–computer interface itself. It is followed by control devices, displays, computer memory dynamics, and the growing use of virtual reality.

Changing Nature of the Interface

Developments in computer technology have changed the nature and scope of the human–machine interface itself. In the conventional machines that were discussed to this point, the control and the display are two separate entities, even though they are often juxtaposed on a control–display panel for optimal use. In the earlier phases of the computer's commercial existence, computers were heavy on controls and control sequences, light on displays and display intelligibility, and abominable for contemporaneous juxtaposition.

When the VDT, CDT, or monitor became the primary display medium, it became a large part of the control medium as well. Think about it. What do we do? We move a mouse that moves a pointer—a virtual finger—over a picture of a button. The button activates when we click the mouse. Some equipment configurations allow us to skip the mouse and poke a finger at the computer screen. We do these things while working on a project of some type that is displayed on the same screen. And at special moments, a box pops up that could be a display of information, or it could ask us to click something in order for something else to happen, or it could be irrelevant and annoying.

The human–computer interface has changed so substantially since its inception that the changes formed the proverbial big circle. The first interface was actually an entire room that had portions of the computer equipment lined up against the walls that surrounded the human. Built in 1946 as the first prototype of an automatic computational machine, it surrounded a team of people that was required to operate it (Heppenheimer, 1990). That was a lot of hardware for a device with a memory of 1 kilobyte!

The operator set myriad controls on the ENIAC every time a new task was required. The concept of programming had not been introduced yet. The operator consulted an abacus during the control process; the handheld calculator did not exist either. By the 1970s, the handheld calculators that were widely available had computational memories greater than ENIAC.

In the post-ENIAC era, the interface became a desktop device of one sort or another. By the 1990s, the entire computer and display device had become small enough to carry in a briefcase. Some varieties of computer devices can be made so small as to be worn in a shirt pocket—wearable computers. The smart phones that so many people carry today are ubiquitous examples that combine the telephone, e-mail access, and access to a wide variety of computer programs and popular web sites. Biomedical engineering continues to develop implantable devices such as heart pacemakers, neuroregulators that can predict and prevent epileptic seizures, and prosthetic devices that can operate from the host's intention as we would move our natural limbs (Nathan, Prost, Guastello, & Jeutter, 2012).

Identification chips can be inserted into pets containing the pet owner's identification and the name and location of the veterinarian who installed it. The goal is to be able to identify the dog and owner if it were lost; most people who find a lost dog dial the phone number on the dog's human-readable ID tag. Efforts have been made to make chips that can be used to actually locate the dog, but they seem to need an energy source that has not been designed yet. Students in some school districts are required to use ID tags that have an identifier chip. One school district in Texas is requiring high school students to use them so the district can see who is present in school and who is not, crack down on truancy, and improve their state-funding levels, which are predicated in part on student attendance levels (*Huffington Post*, 2012). The daily tabulations would interface with a telephone call list that counselors would use to ascertain the whereabouts of missing students. Students who do not comply with the ID program would not be allowed access to common areas such as the cafeteria or library.

Yet other combinations of technologies can once again surround the operator—virtual reality chambers—although with very different purposes and results compared with ENIAC. Another way of surrounding the operator is with the use of sensor textiles. Gloves have been designed that contain electrodermal sensors in the fingertips to detect physiological arousal (Valenza, Lanata, Scilingo, & De Rossi, 2010). Gloves are more portable and versatile than the exposed sensors that have been used historically in electrodermal response studies (Guastello et al., 2006; see Chapter 2). Valenza et al. (2010) have, so far, been able to isolate four NDS patterns of electrodermal response that supposedly correspond to personality patterns, although the latter were not divulged. The sensor gloves were also suggested as a step toward smart textiles that could read biometric information from the wearer. The possible uses of biometric clothing and sensor gloves are not yet clear; this could prove to be one of many engineering strategies that make the object first because it can be done, then figuring out what to do with it later.

Controls

Keyboards

The QWERTY keyboard that stares back at all of us was initially designed for use with the first typewriters in the late 19th century. The first typewriters were mechanical devices (i.e., they did not need to be plugged into walls) that bore a great similarity to the mechanism of a piano. Pressing a key would activate levers. At the end of the last lever was a die that was cast with two characters that we know as uppercase and lowercase (capital and small letters in most cases). The die would strike a cloth ribbon that was loaded with ink. The ribbon was positioned just in front of the paper on which the print was going. When the die hit the ribbon, ink from the other side of the ribbon would impress on the paper. The paper was manually fed into the carriage portion of the typewriter. The carriage was fitted with grippers for the paper, adjustments for single and double spacing, margins, and tabs. The shift key on the keyboard would connect to the carriage and raise it up so that the hammer with the die would strike the uppercase letter, rather than the lowercase letter.

Any creature from outer space who knew our alphabet and saw a QWERTY keyboard for the first time would have to ask how we got from one system to the other. The answer is not historically clear in the sense that there were any human factors experiments to rely on, but some HFE sensibilities were apparently involved. In one saga, the concept of the typing phenomenon would involve the use of all eight fingers and the thumbs on the space bar. In another saga, all the letters of the word *typewriter* appear on the top row of letters as an aid to potential salesmen who wanted to show off the device without actually having any typing skill. The third saga seems to be the one that is the most strongly corroborated: The QWERTY design was actually intended to *slow down* the typist so that the letters would not be typed faster than the capacity of the report action of the mechanism. If the typist typed faster than the machine could handle, the keys would jam and a mess might appear on the typing page.

Typewriter mechanisms eventually improved for faster report times. A prototype keyboard was introduced around 1915 in which the letters were rearranged to maximize left–right alternations between the hands, thus allowing faster typing, especially for two-fingered typists. The new keyboard arrangement did not become popular. The user population had already become accustomed to the QWERTY arrangement, and relearning time was prohibitive. Thus, the preference for QWERTY persists today even though the lever mechanisms behind the keyboard have disappeared.

Electric typewriters were first introduced in the 1950s. The essential design was to replace part of the lever mechanism with an electrical contact. The intended result was to reduce the physical forces needed from the typist, thus improving typing speed, and to deliver an even impression on the paper from each keystroke. Anyone who bought the first models got a surprise, however. The keys had very little resistance to them, and keys would be depressed at undesired moments. If a typist lingered on a key a fraction of a second too long, or hit a key a little too hard, multiple *keeeeeeeeeeeeeeystrokes* would appear suddenly. The obvious remedy was to reintroduce a little more resistance to the keys and to give additional resistance, if not to stop altogether, multiple automatic keystrokes except on glyphs that the user might actually want to strike repeatedly such as *X*, periods, hyphens, and underscores.

Perhaps "obvious" is a poor choice of words. A colleague bought a 2010 model of a popular brand of laptop computer. He was no stranger to laptops at the time. Imagine his surprise and frustration when he got the *keeeeeeeeeeeeeeystroke* effect. If anything should be obvious in retrospect it is that engineers who forget history, or never learn it, are doomed to repeat it.

Error correction on a typewriter was simple for decades: Do not make them, or you will hate yourself afterward. Corrections were made with erasers. The skill was to erase the unwanted characters without ripping a hole in the paper and to retype the new characters in exactly the same position in which the originals had been placed. The paper feed and adjustment mechanism—and the person using it—had to be well controlled, or else the corrected text would not fit into the erased space exactly as intended. Correction fluid and correction tape eventually appeared in the 1960s. They were very popular.

The next level of evolution for the typewriter took the form of the IBM Selectric®, which was widely adopted for use in commercial offices in the 1970s. The Selectric was noteworthy for several features (besides weighing a proverbial ton). A revolving sphere that contained all the type characters replaced the hammer-and-die mechanism. Spheres were interchangeable, and the common preferences were regular versus italic type, larger versus smaller size, and math symbols. In other words, the office typist was now able to change fonts. The Selectric was also the platform for building the first word processors; we return to this topic a bit later on in this chapter.

Ergonomic keyboards were first tested in the 1920s, then forgotten, revisited in the 1960s, then revisited again in the 1980s when the reports of carpal tunnel syndrome and other musculoskeletal injuries were accumulating among people who spent the better part of their workdays entering data (Rempel, 2008). The concept is that hands, which are attached to arms and shoulders, approach a keyboard at a 25° angle. Thus, one feature of the ergonomic keyboard was to split the keyboard accordingly as shown in Figure 11.1a and 11.1b. The design in Figure 11.1a included a rest pad for the palms of the hands and some elevation in the back of the keyboard. The design in Figure 11.1b splits the keyboard with some control over the separation angle and more control over the vertical angle. As a result, the wrist pad could not be attached, but it is included as a part that can be freely positioned on one's desk.

Some performance studies for the earliest designs showed fewer typing errors with the ergonomic keyboard compared with the normal straight keyboard, but typists remained more comfortable with the straight keyboard for many years, and many still are. Performance studies on the newer series were showing that performance was actually better on a standard keyboard, but the gap in performance disappears after an hour. Ratings of pain and discomfort with prolonged use favor the ergonomic design, however. Ergonomic keyboards are seldom included as standard equipment at the time of purchase, but they can be purchased separately.

FIGURE 11.1
Keyboard designs: (a) and (b) are ergonomic designs, (c) is a side view of keyboard supplied with a desktop computer, and (d) is the keyboard portion of a laptop design.

The keyboard in Figure 11.c is a side view of a standard desktop keyboard. It has two increments of elevation that are manually operated, the keyboard is not split, and the palms rest on the desk in front of the front edge so that the thumbs naturally fall on the space bar. The keyboard in Figure 11.1d is a standard laptop configuration that has no rear elevation. It is expected that the two hands approach at a 25° angle, but the only place to rest the palms is on the keyboard surface to the left and right of the position pad. Some thumbs might align with the space bar, but for other people the thumbs drop further down around the position pad and click-buttons, causing unwanted clicking events. This could be another situation where the protection against the accident operation of controls could be improved.

Keypunch Machines

The QWERTY keyboard, now augmented with specialized controls, was harnessed for computer operation during the 1950s in the form of a keypunch machine. Computer programming and operation at that time usually did not allow the operator to give commands to the computer directly. Instead, the operator had to prepare a stack of punched cards to execute the programs. The cards were prepared with a keypunch machine.

An example of a punch card appears in Figure 11.2. Each card had the capacity for 80 keystrokes, and represented a short line of type comparable to what appears on our computer screens today. The operator loaded a set of new cards into a bin that was located on one side of the machine. The feeder mechanism moved a card into a position where typing and punching would occur. The typing operation would result in a set of holes in the card that would eventually allow electrons to pass or not pass through each particular location. This binary code corresponded to what was printed in human-readable characters at the top of the card. When the card was finished, the operator pushed the release button, and the finished card would move to the collection bin.

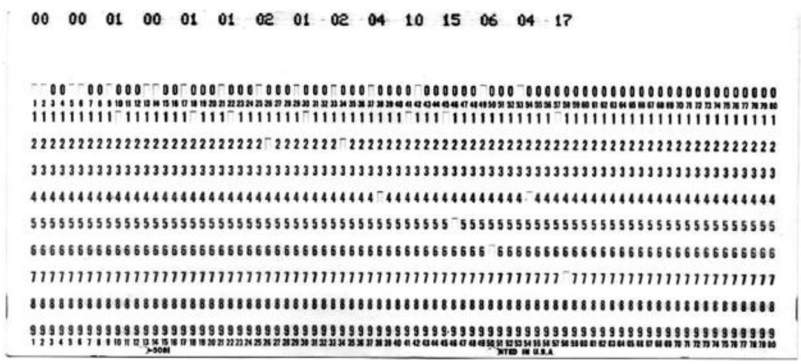

FIGURE 11.2
Computer punch card.

We should pause to reflect on what was taking place. Typing errors occurred often, but the only way to correct them was to eject the card and catch it before it reached the collection bin or to mark it for later removal. As the keypunch machine aged, the mechanism would go out of alignment, and some of the cards would be punched a little off target. It was not usually possible to detect an off-punch card until the batch had been executed.

Once a set of cards was finished, the operator took the stack of cards to a device known as a *card reader*. The top card identified the user and the start of a new job. Card readers were typically located in inconvenient and cramped spaces in the computer facility. Someone would bump into someone else who was holding a stack of 500 cards and the cards would go flying and would need to be reassembled in the correct order.

Was the deck of cards reassembled correctly? Were the error cards removed? Were there any lingering errors on any of them? Were any off punch? The users would know the answer to these questions in about 45 min. A typo that we would discover in less than a minute today, and one that could be fixed in a few seconds, would require as much as 45 min to reach the user in the form of an error notice on printed output. The reply time could be as short as 15 min during times of low computer traffic volume. Thus, a culture of people evolved who did their work at 3:00 a.m. in order to work more and wait less.

Numeric Keypads

The computer keyboard has a numeric keypad sitting on the right. The numeric keypad was introduced during the 1950s, and it is strikingly similar to the keypad on a touch-tone telephone. The latter was a classic example of solving a design problem with a human factors experiment. Deininger (1960) studied the best means of configuring the 10 digits—rows, columns, a circular arrangement similar to that of a rotary telephone, and the square configuration that was eventually adopted. The participants in the experiment were asked to dial a series of phone numbers, and their speed and accuracy were recorded. The rest is history.

The circular arrangement was a reasonable attempt to capitalize on a population stereotype associated with the rotary telephone, but apparently the population stereotype was not powerful enough to overcome a better design. This outcome was quite the opposite of the attempts to redesign the QWERTY keyboard.

One might ask why the arrangement of numbers on the numeric keypad places the 1–2–3 row at the bottom whereas it appears at the top on a telephone. During the early days of computer programming when the numeric keypad was first introduced a large amount of typing consisted of series of 1s and 0s. If the 1 and the 0 were placed close together, it was easier for the keypuncher to twiddle two keys. The widespread use of touch-tone phones occurred many years later as binary typing became extinct; thus, confusion of population stereotypes was not a problem.

Membranes

Membrane keyboards were first introduced in the early 1980s. They were piloted for inexpensive home computers and several commercial applications. The first QWERTY membrane keyboards were flat surfaces. One only had to tap the spaces that indicated the letters, and thus minimal physical forces were needed. The attempt to improve speed, if not also to make an inexpensive device, backfired as users made more errors because they were not used to the lack of tactile feedback that lets them know when they have hit a key properly. Manufacturers then responded by making ridges in the plastic membrane to signify the location and space associated with a key. The additional feedback seemed to help, but not enough to replace the button-type keyboards.

Membrane keyboards found a few homes in applications where keypads were needed, but the kind of typing that lends itself to the QWERTY configuration did not make the transition to membranes very well. Figure 11.3 is a membrane keyboard from an electrical power plant operation that was installed around 1990. Note the absence of a QWERTY pad. The key panel to the right is an alphabet pad, but the letters are in alphabetical order, and it appears that only single letters were needed to input data. The next generation of power plant equipment from the 2002 era reverted to a standard desktop-style keyboard with a QWERTY configuration.

Membrane keyboards are easier to clean than the standard variety. Thus, membranes have survived on the interface of the cash registers in fast food restaurants where the environment has cooking oils in the air. In the same application, however, the touch of one button for *cheeseburger* is all that is necessary to sell a cheeseburger; the price is already

FIGURE 11.3
Membrane keyboard, 1990 vintage.

programmed into the machine. Menus and prices need to be updated periodically, however. Thus, the relative merits of the membrane have to be understood in the context of the control operations that are required.

Positioning Devices

The mouse, the four cursor arrows on the keyboard, trackballs, and joysticks are perhaps the four most commonly occurring positioning devices. At one point in history, there was some interest in which was the best device. This question regrettably falls as short logically as asking which type of control was best for the conventional machine. The ultimate answer is, once again, to pick the right tool for the right job, and a situationally specific study may be needed. All other things being equal, Fitts' law seems to apply well to the response times associated with positioning movements, and the mouse is fastest for most tasks (Card, English, & Burr, 1978; Landauer, 1987).

To complicate matters a bit, the motor skills of the user will affect the speed and accuracy of the use of any particular device. Adults who learn to use the mouse for the first time in their lives (there are not too many of them anymore) experience some lack of control or jitter in their mouse placement; they might prefer the predictability and simplicity of pushing the buttons that correspond to the arrow keys. For the accomplished users, however, the mouse is usually the preferred device, and may be the only way to activate icon controls on a program. The arrows, on the other hand, are good for moving the cursor a small number of discrete steps in the four orthogonal directions.

A quick reminder of Fitts' law would be helpful here. Movement time is a function of the size of the target, larger is better, and the distance travelled to reach the target. For mouse and trackball operations, the trajectories all fall in the same 2-D plane of the computer screen. For touchscreen interfaces (see below), the movement is more similar to the original context of reaching a button on a control panel from somewhere outside the plane of the control panel.

A trackball is essentially a mouse that has been built upside down. It sits in one place on the desktop, and the users roll their fingers over the ball to move the cursor on the screen to the desired point. The trackball would have an enter button that corresponds to the left click on the contemporary mouse. The trackball was introduced as an alternative to a mouse for a desktop computer that was placed in a small space where there might not be enough space to accommodate a mouse (Shneiderman, 1987). The trackball requires fine motor control over the fingers as opposed to gross motor movement of the arm. The reach for the enter button was probably more convenient for the user with the mouse as the left-click button is located under the index finger. As additional uses for a right-click button were introduced, the mouse gained superiority because its other control is conveniently located under another finger, and the operators do not have to move their hands to operate the buttons.

On the other hand, the trackball design may have found a home on most laptop computers. There is no mouse because there is no desktop to roll it on. Instead there is a little red nubbin tucked in the middle of the keyboard with enter buttons to the left and right of the space bar. The nubbin was replaced by a touch-sensitive pad (e.g., Figure 11.1d) with two buttons underneath corresponding to the left-click and right-click control movements. Wireless mice became available, nonetheless, for people who were having difficulty making the psychomotor transition from the desktop to the laptop computer. Users' preference can probably be gauged in terms of the numbers of laptop and desktop computers sold every year. There is obviously a trade-off taking place among the conveniences of one feature over another.

The joystick originated in equipment where a 2-D controller was needed, such as an airplane or a stick shift in an automobile. Joysticks also showed up in arcade games and remote control toys where the control of the object in a desired direction was not complicated by a positioning device on a computer screen: Move the joystick, see the results, and click the button on top to enlarge something. See Figure 7.13 for an example positioning device for a remote control security camera.

Touchscreens

Touchscreens bring the simplicity of a single touch to the display, rather than the keyboard. Here we have a true example of the display being the control panel simultaneously. Figure 11.4 is an example of a touchscreen located in a power plant. The user only needs to touch an item that is depicted on the screen, and the screen changes to produce a new image of what is inside the original small target space.

Although some touchscreen applications have endured since they were introduced in the 1990s, others have been replaced by more conventional keyboard and mouse configurations. The touchscreen seems to have reached its limit in safety-critical environments when it induced too many chances for error by hitting the wrong target or the amount of pressure on the screen not being sufficient to produce a response as soon as it was needed. Another limitation is that the amount of time the operators had to stand or sit by the screen with their hands and arms in the air ready to poke a target produced fatigue that was not generated by keyboard and mouse controls.

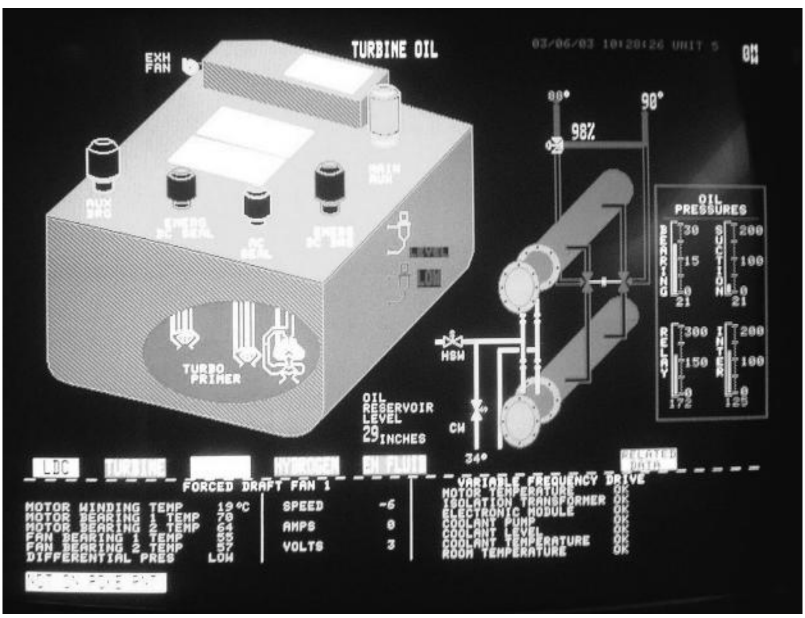

FIGURE 11.4
Touchscreen control with graphic user interface, 1990 vintage.

Styli

A stylus has the look and feel of a pen. It operates when it is touched to the screen, but it allows for more control precision. It was a device of choice in the graphics industry years ago because it capitalized on the artists' control precision with a handheld drawing tool, yet allowed for data entry. Fingertip touchscreens, on the other hand, appear to produce their best results when the targets are large and few in number. The equipment shown in Figure 11.4 was eventually replaced by a mouse-driven system.

Figure 11.5 illustrates the use of a stylus for 3-D data entry. Here the operator wants to encode the shape of a model frog into an animation or virtual reality program. The frog will be used as an object within the program thereafter.

Gestural Interfaces

A gestural interface allows the operator to command a machine with a wave of a hand. The first examples of a gestural interface were introduced in the 1920s with a musical instrument, the *theremin*, which was named after its Russian inventor Leon Theremin (Pinch & Trocco, 2002). The theremin was a box-shaped device with metal plates on its upper surface. The metal plates exuded an electronic beam above the plates. The motion of the hand interrupted the beam, and the interruptions were translated into sound. One hand controlled pitch, and the other controlled volume. The theremin did not gain a great deal of popularity when it was first invented, but it was featured in the soundtracks of some B-grade horror movies in the late 1940s and early 1950s. It experienced a resurgence in the mid- to late 1960s when it was discovered by some musical ensembles specializing in the psychedelic genre with electronic sounds (most notably Ultimate Spinach, Lothar and the Hand People, and the Beach Boys' hit "Good Vibrations").

FIGURE 11.5
Microscribe® Digitizer used to encode a 3-D object. (Photograph copyright 2005, Immersion Corporation. All rights reserved. Reprinted with permission.)

Although the unique fluid sound of the theremin was very attractive to experimental musicians and virtuosos did exist, the overall lack of tonal control was a serious limitation for professional musicians who wanted to adopt the new sound relatively quickly. Most theremins sold during the 1960s were manufactured by Robert Moog, an expert performer himself, who eventually combined the theremin's electronics with new controllers and a piano-type keyboard to make the famed Moog synthesizer.

One motivation for developing the gestural interface in the late 1980s was to assist people who were working in challenging environments, such as Mars, so they could operate their equipment without having to remove the glove portion of their environmental suits. They would not have to fiddle with the complication of low gravity when turning knobs or throwing switches. Martian apparel is still in very low demand, however.

A more viable application of the gestural interface concept is in virtual reality systems. The system requires the use of a glove that contains a system of sensors, as in Figure 11.6. Moving one's hand in the air will interact with a virtual object, which the computer will interpret according to the laws of physics—or not as the designers prefer (Greenleaf, 1996). The basic limitation of the gestural interface is that the hand motions must be executed with stereotypic precision. The motions for a specific command must be executed exactly enough so that the machine recognizes the command and distinguishes it from other gestural commands.

The gestural interface and the touchscreen have been combined in the operation of the current generation of smart phones. The most common motions are the flip-up, down, left and right to scroll, and squeeze the thumb and index finger inward or outward to zoom in or out. The fingers maintain contact with the screen thus circumventing the need for a sensory glove and reducing the degrees of freedom in the motion that the device has to read. The stereotypic smartphone motions seem to have been learned easily. There have been informal reports, however, of users accidentally activating a program or deleting

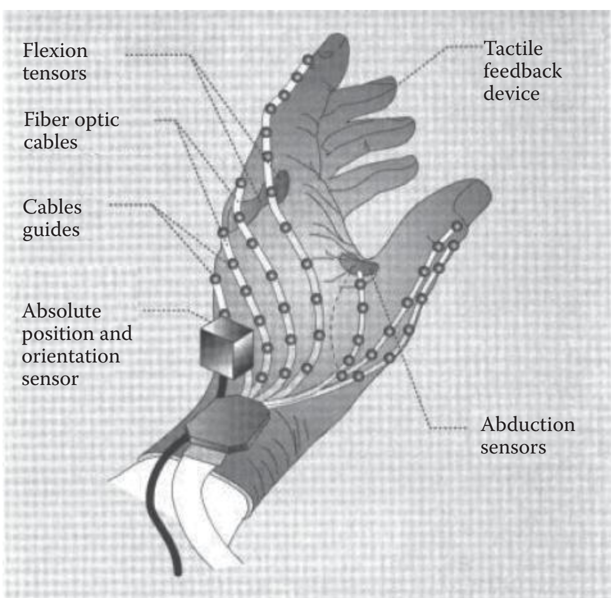

FIGURE 11.6
Data glove for VR applications. (From Greenleaf, W., *Virtual Reality*, 3(1), 13–17, 1996. Copyright 1996 by Walter Greenleaf, Greenleaf Medical. Reprinted with permission.)

something while making the scroll motions. Designers might want to consider how the newer devices could be requiring comparable advances in the protection of controls from accidental operation.

Another application for gestural interfaces is to manipulate very large, wall-sized displays. The operator makes a stereotypic motion to capture a section of the image and move it from one place to another. This style of operation was popularized in the movie *The Minority Report* (Spielberg, 2002). The real-world applicability is actually very limited as is the repertoire of motions that can operate such a system. The 3-D mouse discussed below is more versatile for the manipulation of large images.

Mobile Devices

The laptop computer and smart phone are two of the most commonly used mobile devices. There are others, however, such as devices used by paramedics and those in specific industries where, on the one hand, observation and recording of information from the work context is a substantial part of the operator's job. So far, any ideas in this chapter concerning the usability of any of these devices have been confined to control operations in a relatively stationary context. Mobile devices, by definition are meant to be used as the operator moves around, however. Thus, usability testing for the controls, displays, and program layouts needs to take context into consideration.

One of the biggest advantages of the mobile devices is also one of its biggest limitations: they are small. How many controls can be reasonably condensed into a small space? One rule of thumb is to organize the screens or control panels to accommodate the 95% percentile male finger. The frequency with which this rule is actually observed in mobile designs is an open question.

Baber (2008) took the position that virtually every task performed with a mobile device is a dual-task problem. The operator is walking, paying attention to other things and people, and possibly manipulating other devices simultaneously. Thus, the designers need to visualize all the contexts in which the device might be used as well as the features and operation of the device. How does the effectiveness of the operation compare at a quiet desk with office lighting to a train station with noise, varying lighting, and other people present? How does the operation of the device affect the other activity? For instance, people walking on treadmills walk 30% slower while using a personal data assistant than walking without the device. Are the pedestrians on a busy street as attentive to traffic as they might be otherwise?

Individual differences among users can be expected, some of which could be extreme. For instance, a colleague reported that he acquired an interest in cell phone use while walking when he saw a young woman walk into a public fountain while using a cell phone.

Gaze Control

Chapter 4 explained the uses of eye tracking for determining the patterns of attention to portions of a visual display. Researchers are now experimenting with eye tracking systems that can be used as a form of control. In close-space interaction, gaze could be used as a substitute for a mouse and keyboard for people with motor impairments (Koesling, Zoellner, Sichelschmidt, & Ritter, 2009). In principle, the technology could be adapted to longer distances of operation where the person or robot is too far away to reach by pushing buttons or a touchscreen. The eye tracking sensor, once they are calibrated to the individual, can pinpoint the position on the screen that corresponds to something the user wishes to control. The limitation of gaze control, however, is that the amount of visual attention to a location does not necessarily mean that the user wants to interact with the position. The user might instead be monitoring the display without seeking any changes or trying to figure something out. This problem has become known as the *Midas touch* problem.

An alternative control design registers gaze position but does not activate the control until the user blinks. Koesling et al. (2009) compared gaze and blink controls with research participants performing a typing task on a virtual keyboard. The blink system resulted in shorter response times and fewer errors. The subjective reaction from the participants, however, was that the gaze control *felt* like it took less effort and worked more accurately. The implications for fatigue with longer-term eye blinking have not been ascertained.

Memory Enhancements

Somewhere between the controls and the displays were innovations in computer memory and programming. Note the shifts in the allocation of functions between person and machine as these innovations occurred.

Word-Processing Challenge

The first dominant silicon form was the mainframe computer that worked on a timesharing system. All users would submit their jobs, and the system would allocate chunks of time to each job up to a working capacity limit. The theory of limited human cognitive channel capacity, which was introduced in Chapter 6, was modeled after this principle.

For systems that worked on punch card operation, word processing as we know it today was not a viable option. Electronic word processing required devices that were dedicated to single user tasks and allowed for a quick viewing of the results from keyboard entry. One of the first models for the office worker (as opposed to folks in the printing and publishing industry) was the MagCart® typewriter by IBM. MagCart was built from a Selectric unit, and it had a small memory device that sat on the desk and plugged into the main unit. Typists could now save, retrieve, and reuse typewritten documents. The memory mechanism recorded the information on plastic cards that contained a magnetic strip on one side—not unlike the credit card, which had already become popular in the mid-1970s. They were very expensive relative to their times and did not gain much popularity before the advent of personal desktop computers circa the 1980s; they were, after all, intended for high-volume business purposes.

Type is not print. Type allocates equal spacing to each letter, number, or nonalphanumeric character. Print, on the other hand, allows variable spacing for narrow and wide characters. Different fonts often have different character spacing requirements. The word processor as we know it still had not come to life. The printing industry could purchase a transition technology in the form of the IBM Compositor® for a mere $10,000 new. The Compositor was an overgrown Selectric that had a working memory of 8,000 characters. The operator could use print fonts, and could change the font within a document by inserting stop codes through special keys on the keyboard. As the Compositor replayed a stored document (on paper, not on a CRT), the replay would stop where indicated, and the operator would change the font. The fonts were designed like the font balls in a Selectric. Additional memory like those that accompanied the MagCart typewriters could be attached.

The Compositor was only capable of font sizes up to 12 points. For larger type one needed to use photographic enlargements or press-on type that was set by hand. The organization of pages containing graphics and text had to be composed by hand in an operation known as *keyline paste-up*. Here we encounter the original meaning of the word-processing verbs, *cut* and *paste*. Cut really did require scissors or a razor blade, and paste really required glue or hot wax.

Desktop Computer

The Apple Corporation gets the popular credits for introducing the personal computer to the world. According to Turkle (1984), the personal computer was motivated by the needs of folks with engineering backgrounds who loved computers and wanted their own, but who obviously could not afford a mainframe, did not have space for it in their homes if they could afford it, or could not use their employers' equipment for tasks they had in mind. The mainframe user who was used to the norms of speed and computational power may have regarded the personal computer as a bit silly. The original varieties had no memory storage capacity. The device was operated using two floppy disks—one contained the program, and the other contained the source data that the user wanted to operate on and the eventual output file. Hard drives were not part of the desktop until later when IBM introduced the 286-series computers.

Floppy disks in those days were made out of soft material and were indeed floppy; they had a capacity of 0.75 MB. They were eventually upstaged by the next generation of disk, which by then was only used for storage, had a harder plastic shell, and a capacity of 1.25 MB.

The 286- and 386-series computers were still slow. A colleague who required massive number crunching in his work timed a job performed on a 386 with the same job performed on a mainframe to which he connected by telephone link-up. He reported that the mainframe was 100 times faster than the desktop when operated after midnight.

Substantial improvements in the user's experience occurred when Apple introduced the Macintosh circa 1988. It performed the basic word-processing functions that we know today and contained some capability for fabricating graphics. It was a Windows-based system, rather than a one-screen system. Microsoft Corporation introduced its version of a similar interface for IBM-platform operating systems in 1993 and called it *Windows*.

Desktop computers, despite their early primitive nature, introduced three new concepts in human–computer interaction: the CRT display system, the concept of user friendly, and the metaphor in programming. By 1981, the vast majority of keypunch machines had hit the junkpile, and mainframe systems were equipped with CRT displays. The user could now see the input and the output without having to go through the print-and-wait process. A 45-min exercise in error correction was now reduced to 2 min or so. There would always be some variability in the time required to figure out what to fix, which is still true today.

User friendly quickly acquired a lot of meanings, including but not limited to the advantages of a well-organized CRT display system. The broader use of the phrase indicated that the program and hardware support were adapted to the needs and preferences of the user, rather than requiring the user to learn and think like a computer and adapt to its limitations. In other words, any ideas that were inspired by good human factors thinking could qualify as steps in the user-friendly direction.

The concept of user friendliness eventually morphed into usability, which was the recognition by designers of products of all types of the role of human factors and ergonomics in their work. The concept extended beyond human error and performance into matters of convenience versus annoyance (Reiss, 2012). The most annoyed user of anything in the present era is probably not subjected very often to the limits of the machine's or system's capabilities or its inability to bridge the communication gap between the human and the rest of the system. More often the gap lies between the user's goals, knowledge, and experience and the questionable judgment of the product design team. Convenience matters as well as functionality, and too often, systems are rigged for the convenience of the producer at the expense of user frustration.

One persistent gap between the human and the machine was the effort that it took for the human to understand enough of the intricacies of a computer program to operate it. Programmers came to rely on metaphors for this purpose. By engaging concepts that were already known to the user, the program could bridge the transition between population stereotypes that were drawn from the concrete world to analogous functions in the virtual world. The *cut* and *paste* verbs were an example. Another example is the use of the desktop and file folders in current operating systems. Graphics programs use metaphors that echo physical artwork tools such as pencils, pens, paintbrushes, and paint buckets. In other graphics programs, a box has a complex set of meanings. Although some metaphors might only be meaningful to people who are entrenched in a particular line of work, the use of metaphors has been effective for conveying the operation of program features to the intended users. The role of the metaphor is revisited in a subsequent section of this chapter on icons.

Menu Structures

In the pre-Windows interfaces, menus of operation choices or files would be hierarchically organized. The broadest category would constitute the top-level directory. Groups of more specific operations would appear once an option from the top-level directory was selected. Tertiary and lower-order groupings were possible. A limitation of these menu systems, assuming that they were logically organized, was that if the users did not reach the desired item on the first try, they had to use a backup control to revert to a previous menu and try another avenue to reach what they were looking for. That process often required them to remember what they saw on the earlier screens and probably write down the pathway for later use. A significant advance was to provide a toolbar with the top-level menu listings and pull-down menus; thus, the user can take a quick peek at the options to see if the desired elements are present, and try several pull-down menus in rapid succession.

A strategic design question is how to organize the program elements under the first-level menu options. Landauer (1987) remarked that the principle of semantic categories would be applicable. A sample of probable users should be given a list of functions and asked to sort them into categories. One can then compile a matrix showing the number of times each program element was categorized with each other element. The elements should show reasonable clusters, and the clusters would be better to the extent that they were more obviously distinct.

In the early days of interactive systems, the advisement was that the number of clusters should remain small, and seldom more than seven in number, in keeping with the principle of the magical number 7 ± 2 (Galitz, 1993). The idea turned out to be convenient for systems that could not load up quickly due to limited processing power, but not optimal. The act of searching through menus turned out to be more similar to scanning an array of information for a target (Norman, 2008). The range 7 ± 2 might be good for describing how many chunks of information people could keep in their short term memories or how many categories they tended to impose on a system of elementary ideas, but scanning for the correct menu item is a different mental operation. Menu organization can have its best impact on work speed and error rates by imputing information by use of meaningful categories and spatial arrangements. If a hierarchical menu is needed, search time is much faster if all levels of the hierarchy are visible at once. The practicality of this recommendation could be limited by the size of the list of options, but it is effective when it is feasible.

Data Storage Capacity

The primary advantage of a hard drive compared with its predecessor desktops is that it can store all the program files and files produced by or for the programs without reloading with every use. The storage capacity is divided between two modes, passive storage of things not in use and random access mode (RAM), which is the space needed to operate the programs. The function of RAM is equivalent to that of working memory in humans. The hard drive was essentially a recovery of the function of the mainframe computer, but without the total processing capacity, at least for another decade. The main disadvantage of the hard drive is also that it can store everything. When it crashes, everything is potentially lost. Thus, in the culture of users, the users have historically reminded each other to back up files, which they did not do often enough.

As hard drives grew in capacity, so did the programs and files they made. Preservation required media with larger capacity, thus the zip disk and zip drives appeared. The zip disk had the look and feel of a floppy disk, but was a little larger and had capacities of 100 and 250 MB. Zip drives were prominent for only a few years before they were replaced by burnable compact disks (CDs) with 800-MB capacity. Within just as short a period, computers were no longer equipped with floppy or zip drives as standard equipment; the user had to acquire peripheral devices, which plugged into a USB port, to continue to use the stored material. When people started to make video products on their computers or just wanted to use their computers as video players, burnable digital video disks (DVDs) were introduced. DVDs have roughly five times the storage capacity if used for nonvideo files.

The limitation to the first generation of burnable CDs was that once they were burned it was not possible to add or remove files, even if there was a lot of unused storage space remaining. With rewritable CDs, it was possible to add or change files, but not to simply delete files that were no longer wanted. It was possible, however, to reformat the disk, which meant deleting all the files and starting over. Thus, the increased capacity came at the expense of built-in waste of virtual storage space, clutter, discard, or time spent working around these limitations by copying files from the CD to the hard drive and back again.

Just as it appeared that the usability issues with the rewritable CD had been worked out, a new form of annoyance arrived with the Windows 7 operating system (at least in the rendition in my office at the university). The desktop or laptop user has files that are organized into folders, all of which are visible in the user's file directory. The burnable CD has folders of files on it made by the user. Now the user wants to add a file to one of the folders on the CD. This was no problem at all in the XP series of operating systems. In Windows 7, however, the pop-up window says the device allows you to store a group of files [blah, blah]. What actually happens is that a single file cannot be moved to the CD without getting an error message about the disk space not being available. One must move the entire folder including any files in it that were not meant to be transferred. Although the workaround is obvious after figuring out that the CD burner really does work and that it just does not *want* to work, a workaround is still a workaround and a time sink, especially when it was not necessary in the recent past.

The next advance was the jump drive (also known as the flash drive or data stick). Smaller than a pack of chewing gum, it contained all the drive mechanism and storage space necessary to plug into a USB port and store files of any format. Capacities grew; 2 GB is a common size today. One can add, replace, and delete files with all the convenience of the old floppy or zip disks. People found that they could store all their important files on just one or two of them. Some users saw fashion potential in them and wore their jump drives on a necklace chain. (This writer, however, refused to wear all his class lectures around his neck; the metaphoric image of a noose or a yoke was not comfortable.) The greatest asset of the jump drive was also its greatest limitation: A user could store so much information on one of those little things, and they were easy to lose. When the jump drive was lost or chronically misplaced, so was everything on it. It might be advisable to back up the jump drive … on a desktop computer?

The foregoing chronicle of storage devices should highlight a tension that existed between the user's needs and manufacturers who were preoccupied with introducing new products in that some capabilities were gained, lost, and regained as designs iterated. Users wanted to store data, retrieve it, organize in a way they saw as personally meaningful, perform tasks of growing levels of complexity with it, and share the results with friends and associates on occasion, and maintain control over what they thought they owned.

Clouds

The latest innovation in computer memory seems to be making the full circle back to the early 1980s with devices that have no hard drives—the smart phones and tablets (Duan, 2013). Smart phones have the convenience of being able to deliver Internet, including movies and television programs in a handheld device. All the content is stored on a provider's web site and downloaded as needed. The tablet is mostly an overgrown smart phone with a larger display that is more suitable for reading text. Business-related files can be stored and used in the same way. Another advantage is that all the responsibility for updating and backing up files was put on someone else.

The cloud system together with the absence of a hard drive works for people who only send brief text messages, make phone calls, or use corporation data sources that they have no intention of altering in their work. The system works for library material as well as retrieving books from a library that one is only going to return later on. The comfort and readability of the books is another matter; digital eye fatigue was discussed in Chapter 9, and as far as the state of the science knows today, users regularly make a trade-off between one type of convenience and another type of annoyance.

Cloud systems for home entertainment purposes, from one perspective, do not seem very different from turning on broadcast television and radio. Rather than programs always being broadcast at times specified by the supplier, one can watch whatever whenever. One enhancement is to give the user the option to halt a download or stream, do something, then go back to the show without missing anything. These conveniences used to be one of the main reasons why people bought videotape machines—record the show, play it later, start and stop at will, and fast-forward over the commercial messages.

Programming the videotape machine was a source of error and a long-standing challenge to usability, but that did not prevent some households from accumulating three or four of them. Videotape recorders and DVD players eventually afterward were also good for owning a library of favorites that could be played repeatedly at will. Now all those annoyances can go away and the user can pay per view for streaming downloads.

Cloud facilities have become attractive to some users for storing and backing up files. One only needs to upload, or permit the cloud to enter one's computer and do the uploading on a regular basis. This method of file preservation appears to be a solution for users who have been on the short end of more than one hard drive crash that resulted in large losses in files and replacement costs. The downside to some, however, is that suddenly some other agents have access to all of one's material, private information, intellectual properties, and so on. Some people value their privacy more than others; some people hedge their uses of cloud systems by maintaining some control over some of their material and not uploading everything. Some individuals and businesses would not use them at all.

The cloud systems are actually making two full circles. Some tales from the trenches would be useful here. In the early 1980s, yours truly visited a major transportation company who, in the course of things, expressed their excitement about their new computer system that stored all their shipment information, much of which was geographic in a satellite system. Actually, this was not much different from communications satellites that facilitated long-distance telephone, emergency response, and so forth. The access to the database was done through a telephone-based system that was equipped with accelerators to accommodate all the data flowing back and forth. "Do you have the information backed up in case the satellite goes out?" "No need, the technology is very sound." Uh-huh. During that period, one military and political concern was for building the "Space-Based Defense Initiative," which included in part killer satellites that would cuddle up to an enemy's satellite and explode itself, or perhaps launch an anti-satellite attack from Earth. A different concern was that NASA and the RCA Corporation were doing a great business designing and launching satellites—and fixing malfunctions.

I never heard of anything bad happening to the transportation company's system (good thing), but not everyone in the late 2000s was so fortunate. The setting was a university clinic for psychological services that used digital video recordings of patient–therapist sessions for training purposes. In some systems for training, the sessions are saved on DVD and do not leave the premises in any way. In this one particular case, however, the clinic adopted a system of storing the recordings on a cloud server. There was a huge storm on the other side of the continent and the server went out. The clinic lost access to its files for three weeks.

Visual Displays

As with conventional machines, the first major design questions were to decide what should be displayed. Questions about how logically followed after a sufficient time lag. Here we capture the broad strokes of evolution from the first error messages to the graphic user interface.

Error Messages

In the days of the early computer languages such as FORTRAN, the first step that the computer would take in processing the program was to run the user's program through a compiler program. The compiler would determine whether the grammar, syntax, and specification of commands (and spelling) were correct. If an error were detected, the output would state the program line number containing an error. During those times the programmers were mostly from a population of experts. Thus, a small hint concerning an error would be enough to alert the programmer as to what needed correction. Novices and students could become terribly frustrated, however, and they would lose a great deal of time trying to determine what the error message implied. Eventually, the norms for error documentation became more generous with the explanations of what was wrong. Once the age of user friendliness arrived, the error messages might state some possible corrective actions that the programmer could try.

Here we saw the essence of the smart display. Instead of simply reporting what offended the compiler, the display framed the information in terms of what the user wanted to know or to do. This technique eventually carried over to displays that might appear on computer-based or conventional machines: The displays would frame their information in terms of what the user should do and not simply report the system status.

Sometimes what you see really is what you get from a computer program. The results of a control operation might not need any annotation. The user only needs to be displeased with the obvious results. For these moments, the *undo* function is your friend. This is yet another innovation made possible by the concept of short-term memory. The computer can hold the last version of the product, for example, a photograph that is going through an enhancement process, and revert to the original version when *undo* is clicked. It can also permit a *redo* command. Undo functions should be easily accessible in most any type of program.

Screen Organization

A user-friendly organization for a single screen of information draws on the best wisdom of conventional machine control and display panels and that of good text layout. Elements should be organized by function, sequence of use, frequency of use, and whether the elements are used together to perform an operation. Relative size of elements matters. Elements should be large enough to be seen from the normal viewing distance, and large enough to form an easy target for a pointing device if a control operation is associated with the element. On the other hand, the operator should not have to scroll up and down a screen to any great degree to complete the work associated with the screen. There is obviously a balance that needs to be struck with regard to the size of elements and the choice of elements to put on a particular screen.

Elements that are used less often, perhaps only for particular circumstances, can be aggregated to a second screen or a pop-up window that can be opened and closed as desired. The logic for organizing a multiscreen system is not especially different from the logic for organizing a menu of options that was sketched previously.

The combination of computer technologies over the years has been successful in reducing the morass of control and displays for a power plant control room, such as the mid-1970s vintage example in Figure 7.17, to a set of six monitors plus some ancillary items in the 2003 system in Figure 11.7, top panel (only four of the six monitors are shown). One of the displays shows the alarm status of sensors throughout the system. If one of the line entries is interesting, the operator only needs to point and click on the line and a detail of the sensor's history pops up in a window as shown in the bottom panel of Figure 11.7. Note that the information can be presented graphically in a line chart or in a table of numeric values. The graphic and tabular representations obviate the need for the types of historical displays that were used in decades past.

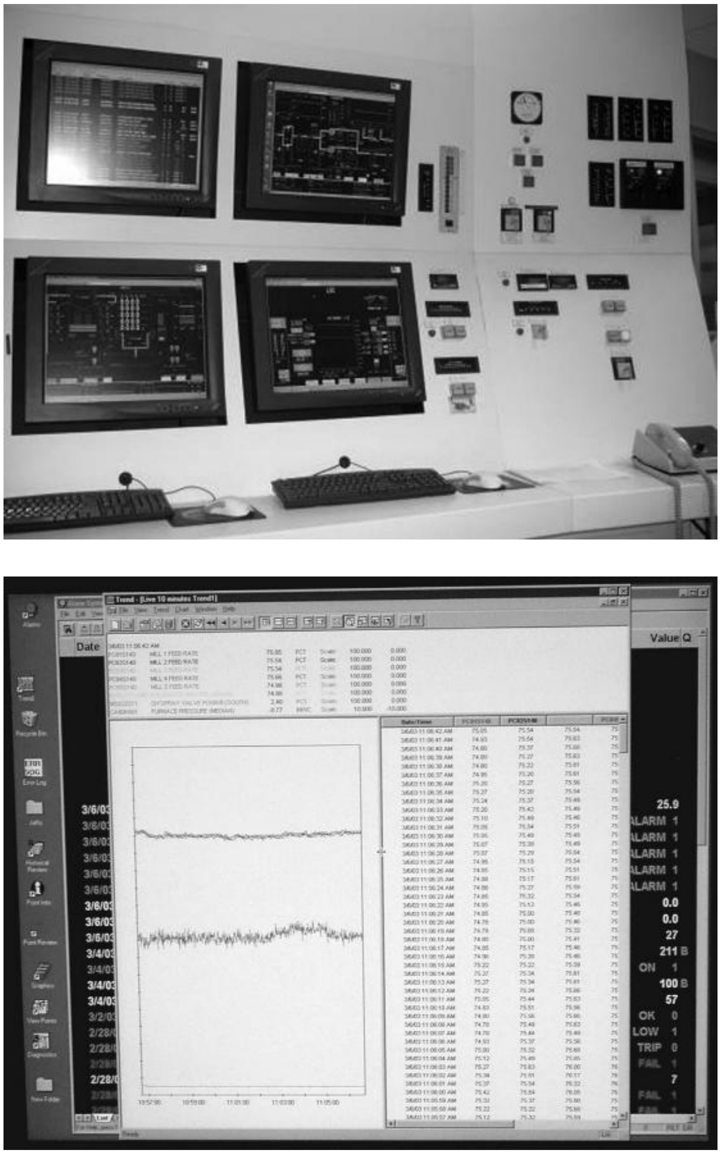

FIGURE 11.7
(Top) Four of six graphic user interface screens controlling a power plant with a few ancillary controls and displays, 2002 vintage. (Bottom) Screen 5 allows for clicking and expansion into a historical display.

Despite all good intentions, the designers sometimes miss something important such as relative demand. Figure 11.8, top panel, is showing a critical system stability parameter. The operators can only see what is going on when they get close to the screen and if they know what to look for. The big secret is shown in the bottom panel of Figure 11.8, which is a photographic enlargement of a portion of the main screen. The operators do not want the dot to cross the left curve (which is red in the original). Hopefully some other element of the system is calculating the proximity of the point to the curve and will display an alarm condition when the point crosses the curve.

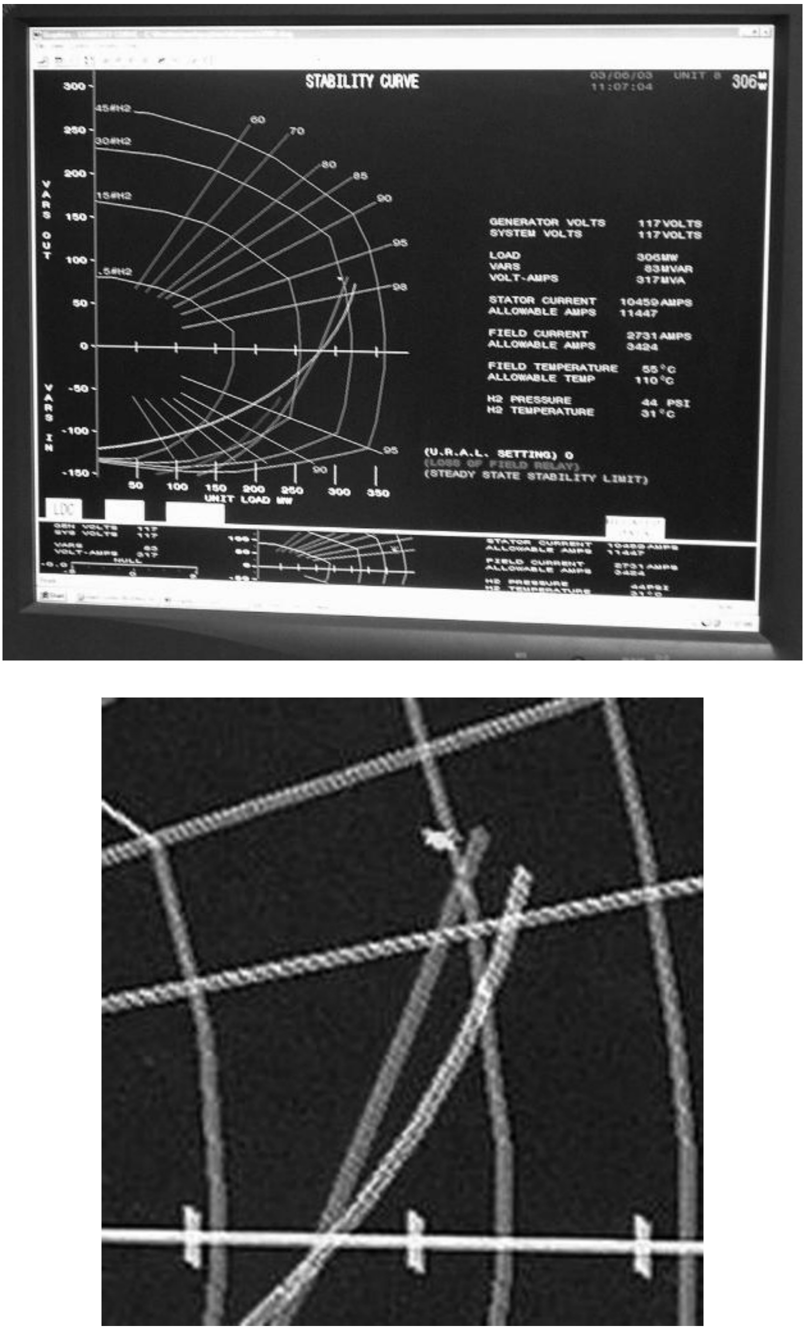

FIGURE 11.8
A problem in demand representation. (Top) Screen 6 from the power plant control system shows a critical system parameter. (Bottom) Closeup of the same screen shows where the critical item is located.

Screens pop up when the system needs to convey a message to the user about some aspect of the system status or the command that was just given. Where should such a window be placed relative to everything else on the screen? Should it be in a fixed location, or a variable location? According to a study by Haubner (1992), the message windows should always appear in a fixed location.

In cases where a small dialog window pops up in the course of ordinary use, some systems allow the user to move the pop-up window to a convenient location on the screen while working back and forth between the main window and the pop-up. Pop-up messages that pertain to something critical, however, do not allow the user to move the window or otherwise continue working until the issue has been resolved.

Human–computer systems are not as closed and self-contained as they once were. The Internet access is usually running constantly. Sometimes the only way to use a program is to go to a web site; it cannot be downloaded for private use with any degree of convenience. The Internet connection means the outer world has a connection to the operator as well. Pop-up windows are widely used for instant messaging, advertising, and malware, all with the intent of distracting users to make costly decisions with a click to the pop-up. Security software and Internet service providers do provide pop-up protection, although it is not completely effective. If anything, there is a regular cat-and-mouse game going on whereby the security system figures out how to counter the new threat and the threatening agent figures out how to bypass the security system. The security systems also have their own pop-up messages to deliver.

The typical user responds negatively to pop-up windows, ranging from mild irritation to strong annoyance, and quickly develops the heuristic of ignoring all pop-ups. Bahr and Ford (2011) made three recommendations to potential senders of pop-ups. (a) When the user is busy, do not interrupt. (b) Save an intrusive pop-up only for emergencies. (c) Present subtle alerts that the operator can go see later, rather than alerts that grab full attention.

Graphic User Interfaces

Graphic user interfaces (GUIs) are a form of screen organization that is applicable to large database and control units. The first-level screen depicts the entire database at an overview level. The graphic screen shown in Figure 11.4 is actually an example; the operator was working on a second-level screen at the time the photograph was taken. The operator would then click (touch) a desired component in the big picture, and a second screen would appear showing the contents of the specific area. The operator could then click an object on the second-level screen to zoom deeper into the information base. There is a hierarchical organization of information that is implicit in such a design.

Another example that has seen many applications would be geographic databases that could contain marketing and demographic information or locator information. The top-level graphic would show a recognizable map, for example, of the United States. The user would click a state and see the details at a statewide level, click a county to see what is going on there, pick a city within a county, and so on.

Shneiderman (2000) gave several examples of top-level graphics for GUIs that could be used to depict abstract databases such as patterns of customer sales in a large store, or the Library of Congress' catalog of holdings. Some of the designs may be more intuitively appealing than others, and readers are encouraged to browse the examples and see for themselves. Importantly, the logic of all the designs is the same: overview, select, zoom, repeat. This principle carried over rather effectively to geographic information systems in particular.

Use of Color

The computer offers the temptation to apply elaborate (and sometimes tasteless) color schemes because of the low costs associated with using color on a screen compared with print. The issues associated with color in conventional interfaces apply just as well to the computer interface: Higher contrast provides more information than lower contrast. Here *contrast* refers to both luminance and chromatic distinction.

According to Galitz (1993), the empirical studies were divided as to whether color enhances performance. Extreme contrasts or too many different colors in a display can cause discomfort and distraction from the content. In some cases, the users thought their performance was enhanced by color, but it really was not. Galitz advised that the number of colorations be kept to less than 10 and closer to 4.

Perhaps the critical question for computer interfaces is the same as it is for conventional machines: Does color add information, or is it frivolous and decorative? A productive use of color in a GUI would be to aggregate items by function or other relevant classification category and then color the backgrounds to signify where one category ends and the next begins.

A good example of adding high-contrast color to convey information can be found in systems that display biological information. A cell that is depicted in its original colorations would be of low contrast, and the elements of the cell would not be easily distinguished. Exaggerating the contrast of original elements' colors helps somewhat. Systems produce much more readable results using pseudocoloring. With this technique, the original elements are identified by the system, then displayed using colors that are easily distinguishable to the human although they might bear little resemblance to the original specimen's colorations. Pseudocoloring is used routinely in PET and fMRI brain images. Brains do not really light up in the way that the pictures on the screen might suggest.

Geographic information systems are natural applications for color graphics. For instance, the database of national medical data would be organized around a map of the country. The country would be segmented into irregular shapes corresponding to states and counties (e.g., in the United States). Each entry for a county (or other smallest unit) would contain a piece of categorical data, for example, prevalence rate of a condition. The rate scale would be broken down into 4 to 9 categorical levels, each of which is assigned its own color patch. The designer can then choose from a palate of colors to indicate the categories (Brewer, 2006; Brewer, Hatchard, & Harrower, 2005).

The question of how to choose the color palate could be answered better with some organized human performance research. The best suggestions at the moment are as follows: For a bipolar scale, the color series could range from dark blue at one pole to bright red at the other pole, with the other chromatic colors in between. Medical imaging such as the fMRI makes use of this scheme, so it would have some familiarity in other applications. For incidence rates from 0.0 to 1.0, a cascade from light to dark is a popular choice that can be found in chromatic sets, such as shades of blue, shades of green, and so forth. Keeping the colors for one display within a chromatic range would be useful when separate maps for several related variables are composed. The user would be able to remember "the red map" or "the yellow map" as a navigational cue. If the categories were really qualitative categories that did not have an underlying continuous measurement connecting them, then a set of colors that jumped chromatic ranges might be less misleading than shades of brightness; aesthetics might interfere here, however.

Just as there are limits to the number of colors a person can reasonably interpret in a display, there is also a limit to how small or finely grained the colorized data should be geographically. Data that are too fine-grained with color changes can lose the visual clarity of an underlying pattern such as a source point or a contour. For this reason virtual cartographers recommend the use of any of several statistical spatial smoothing strategies (Oliver & Berke, 2010; Waller & Gotway, 2004). The data entries can usually be expected to contain an element of error variance, either in the observed number, or in the spatial location to which the number should correspond. Smoothing (in one computational scenario) would involve calculating cascades of medians across a geographic contour (e.g., from a county and the five adjacent to it). Waller and Gotway noted that viewers of GIS data tend to regard the numerical or categorical values to be as accurate as the geographic markers on the map: "In short, a good map of bad data often seems more believable than a bad map of good data" (p. 69).

Aesthetics and color can play a role in the attractiveness of the product—any product—to the potential user (Ji, 2013). This is a marketing issue rather than a human factors issue because it does not affect the joint function of the person and the machine. It might attract enough attention in a (virtual) store, however, to induce a customer (user) to spend some time trying to interact with it or think about purchasing it. If the display hit the viewer at an emotional level in a positive way, all the better. Marketing people might call this phenomenon "usability" too because it affects whether the user *wants* to use it, even though the functioning of the system could be mediocre.

Pop Up and Wait

Despite advances in processing power and speed, it sometimes takes nonzero time for the computer to complete an operation before allowing the user to continue working. In these circumstances, it is advisable to use a pop-up window that shows the progress of the operation (Galitz, 1993). One of the usual glyphs is the hourglass, which just sits there until the operation is complete. Another type of design is the slidebar that changes color from left to right as the operation progresses; the color bar is often accompanied with a verbal message indicating the percentage of the job that has been completed at a given moment. These wait signals explain to the user why nothing appears to be happening, and dissuades the user from interpreting the nonresponse of the system as a crash.

Visual Icons

Icons are the little button-shaped areas on an interface that do something when clicked with a mouse. Despite their widespread use on interfaces, there is still little published empirical research on the effectiveness of their designs on human performance or recognition. The first distinction one might make is between verbal and pictorial icons. The rules for signs (Chapter 4) indicated that if the information was detailed or abstract in nature, a textual representation was preferred to pictograms. Pictograms were typically preferable for concrete nouns and verbs. Similar principles seem to hold true for icons: (a) Concrete designs trump abstract designs. (b) Population stereotypes, when they exist, facilitate recognition. (c) Text or text supplements greatly enhance the recognition of some icons; mouse-overs seem to be an ideal means of supplying the needed text without adding clutter to the visual display.

Some pictorial representations of objects might be found in industry standard glyphs in some topic areas, but the effectiveness of some sets of glyphs was not guaranteed. They might have an advantage, however, for any population stereotypes that are in operation. Guastello, Korienek, and Traut (1989) compared users' responses to icon designs for four industry vocabularies: building automation, finance, data processing, and engineering. There were five types of icons for each object: a short verbal abbreviation, a longer abbreviation or full word, the industry standard pictogram, a custom-made pictogram that was a more concrete representation of the object than the industry standard, and an icon design that included a concrete pictorial representation and a short verbal abbreviation. Human participants were presented with a real word for the object and asked which of the five icons was closest in meaning to the target. An example of a set of experimental icons appears in Figure 11.9.

Brain lateralization theory suggested that the mixed-modality icons would be more meaningful than the other types. The human brain contains two cerebral hemispheres. One side contains the verbal-processing center, and the other side contains the picture-processing center. The two sides of the brain are thought to work in parallel, although a particular person might be faster at processing pictures than processing words. With parallel processing for everyone, however, all system users would have access to their faster RTs, assuming the verbal and pictorial content was compatible and not conflictual. The principle of redundancy was also applicable here: Two brain mechanisms in operation are better than just one.

As it turned out, the mixed verbal and pictorial icons were rated as the most meaningful by the participants in the study. Although meaningfulness is not a terminal performance criterion, it is well known that meaningful stimuli are stored and retrieved from long-term memory more efficiently than nonsense stimuli. The concrete pictures and the longer abbreviations were equivocal over the four content domains.

Some pictures were more concrete than others. In a later examination of performance on an interface with pictorial icons, Blankenburger and Hahn (1991) found that pictograms that were more concretely representative (their phrase was "small articulatory distance") of the target produced more efficient actions with regard to speed and accuracy. No one is sure yet how concrete the icons need to be in a given context without running some tests with human operators. The mouse-over technique, however, does allow a comprehensible label to pop up, usually in a yellow box, to explain the meaning of the icon in contemporary popular programs, thus permitting simultaneous verbal versus pictorial processing.

Air traffic control systems are currently undergoing re-design to accommodate data communication links from the airplanes to the control systems. Air traffic controllers in different regions of the country do not always use the same system software presently, hence different sets of icons. Thus, Ngo, Vu, Thorpe, Battiste, and Strybel (2012) considered the question of how any industry standards should carry over to the new systems. The icons sets they tested tended to be more symbolic than concrete. Air traffic controllers were more likely to associate a double arrow pointing upward or downward as a symbol of increasing and decreasing airspeed, and there were a few other consistent associations with symbol shapes and commonly used pieces of information. Novel information units

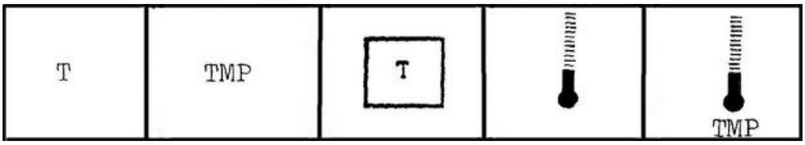

FIGURE 11.9
A sample visual icon for a temperature sensor evaluated by S. J. Guastello, G. Korienek, and M. Traut (*International Journal of Man–Machine Studies*, *31*, 99–120, 1989).

germane to the incoming systems, such as whether the airplane had particular data link equipment, did not have consistent icons. Color was good for distinguishing flight rules associated with an aircraft, but did not affect any other information parameter, nor did open versus filled shapes. Text icons were best for conveying heading and several other information parameters.

Some system designs confront two possible avenues for capitalizing on population stereotypes. One option, when developing a new program, would be to utilize icons that are very similar to those used in related programs already by the same user population. The second approach, which might be attractive to vendors who already have programs on the market, is to make the icons consistent with those in existing programs *within* their brand, and different from their competitors' choices. The latter approach, besides avoiding charges of copyright infringement, puts the effort into developing brand-centered population stereotypes and creates frustration for users of competitors' products (Caplin, 2001).

The technique of frustrating the users of alternative programs might not attract users of competitors' programs immediately because the competitors' users do not have a need to switch. It would, however, inhibit one's own users from switching to the competitor's products. If one culture of brand-specific users grows faster than the others, however, that brand is likely to become dominant in the industry. Dominance in the industry can command preferential treatment from other software and hardware manufacturers who are seeking compatibility for their products with as much of the extant industry as possible. This sequence of events occurred in the graphics industry during the early 2000s, and can be expected in other industries as well.

Auditory and Multimedia Displays

Auditory Icons

Auditory signals, such as beeps, squeals, bells, whistles, and sirens, are well known in conventional circumstances (Chapter 5). Eventually, they appeared on computer interfaces. Gaver (1986, 1994) proposed a theory of auditory icon design that was made possible by digital sound synthesis. His taxonomy contained three types of auditory icon. *Symbolic icons* are those where the sound pattern has an arbitrary association with the machine process in progress. *Metaphorical icons*, however, bear some literal similarity. As with metaphors in literature, these icons capitalize on one feature of the process in question. For instance, a hissing sound might represent a snake. (Of course, the widespread utility of a snake to computer programs is another matter.)

Nomic icons, which Gaver (1986, 1994) preferred, are those that have a literal similarity between the process and object of the icon. For instance, the sound of breaking glass might represent a glass object, the sound of wood hitting wood would represent wood, and so forth. Sounds such as glass and wood have a distinctive wave envelope, which would help in their fabrication. Again, the concrete rather than symbolic icons are preferred. No one really knows for sure, however, how auditory or visual icons acquire their iconic quality (Jacko & Sears, 2003), but if features other than concreteness are involved, iconic quality could be related to both the distinctiveness of the image or sound and its unique association with an object or event.

Are auditory icons preferred by their users? The answer to this question is regrettably informal, but should suggest where formal research should head. Many system users prefer to work with the system sound off, unless they are listening to music in the CD drive. One design group that this writer had the pleasure of working with incorporated auditory icons in one of their program modules. The particular program event involved controlling a mechanical subsystem to fill up with water or empty it out. The auditory icons transpired over the physical operation time with an ascending or descending tone series to denote the flow of water. The users who would describe themselves as computer enthusiasts found the earcons amusing. The other users found them annoying and asked that they be turned off.

Perhaps the better uses of earcons might include signals indicating that an invisible operation has started or finished, or applications akin to the pop-up and wait window. On the other hand, auditory icons are popular in children's software. They often take the form of musical looney tunes that do not appear to serve an objective other than to hold interest, or perhaps add life to a potentially boring program. Some are mixed with speech displays, which might be necessary for giving children who cannot read much some instruction about what to do in their program.

Speech Interfaces

Speech interfaces might be used to give the computer commands or as part of a text-to-speech display. The essential points on verbal commands appeared in Chapter 6. Van Nes (1992) offered some recommendations for optimizing speech displays.

The computer's speech should be clear and natural, with the intonations and inflections that correspond to normal speech. One important goal is to hold the listener's attention, and monotonic droning does not do any better here than in real-time speeches.

Longer passages of text should be accompanied by pause and repeat controls. Some applications might benefit from a text or visual outline of the material to assist navigation through the material. One way to enhance navigation is to use quite different voices, usually male and female voices, for different parts of the text sequence. The shift in voices should correspond to a shift in the content of the message.

Some research has been done on the use of voice annotation to make edits to a word-processing document. Van Nes (1992) reported that users making the edits preferred a voice system by 2 to 1, but the text annotations were preferred by the people who actually had to make the changes in the document.

Earcons and Spearcons

The diminishing size of the GUI on cell phones and other hand-held devices has prompted the development of new sonification techniques to compensate for the visual restrictions. *Earcons* (Blattner, Sumikowa, & Greenberg, 1989) are brief auditory displays that consist of a few musical notes. They are meant to be as memorable as visual icons, but without the literal connection between the sound and the referent that a true auditory icon would have. On the other hand, they require sound designers to come up with the intuitive quality that makes a memorable earcon. Earcons are meant to be hierarchically structured to convey first- and second-level menu items; different timbres would be used to separate the second-level categories. A limitation of the earcon concept, however, is that a set of sounds cannot be readily augmented if new menu items are introduced later on. There is no formulaic theory behind their design; their evolution is several steps behind that of the visual icon in that regard. Although a few such earcons might serve the purpose of conveying status information to the user (e.g., "Windows has finished booting up"), the viability of earcons as actionable objects is another matter entirely.

Walker et al. (2013) thus introduced a new sonification system that purports to be more functional than earcons. Spearcons are text-to-speech (TTS) conversions that have been sped up so fast that they are not readily recognized as actual speech. Frequency patterns underlying the phonemes are preserved, nonetheless, so there can be a familial resemblance between similar but different menu items such as "save" and "save as." In a series of experiments, Walker et al. compared the user responses to auditory icons plus TTS, earcons plus TTS, spearcons plus TTS, and TTS alone. In the former three conditions, the icon, earcon, or spearcon was presented first and followed by the TTS; if one design was better than another, the user should be able to respond faster and more accurately without having to hear much of the TTS that transpired in real time.

In two experiments where it was possible to develop full sets of auditory icons to represent objects, earcons clearly produced the poorest results for speed and accuracy; the other three modalities were tied. In a third experiment that involved learning representations of 30 nouns or cell phone operations at two levels of hierarchy, participants using spearcons outperformed those using earcons in both cases.

Their fourth experiment involved learning and recognizing a set of outdoor objects or events such as buildings, intersections aids, obstacles, plants, usable objects (e.g., public phone, or garbage can), and landmarks. This time it was possible to create a hybrid set of true auditory icons and earcons and compare the set against spearcons. Spearcons produced better performance.

The final test was a comparison of TTS alone and spearcons alone. The task was to navigate a cell phone menu at different levels of depth. Spearcons produced the shorter set of response times at all levels of navigation depth by a clear margin. All told, the spearcon technique could accommodate the widest range of objects or menu items and were best for speed, accuracy, and learning compared with auditory icons or earcons.

Animation and Hypermedia

By the time animation became available in user interfaces, digital video and audio clips were also available. Systems that combined any combination of these display techniques became known as *hypermedia*. The human factors studies that were reported in recent years often compared combinations of media techniques for their impact on human performance. Many of the applications were centered on educational products, where performance was measured by student learning. Some of the commercial applications, such as those that appear on web sites, were evaluated in terms of the amount of time it took for a visitor to find something, and how much time they spent with the site overall.

According to a review article by Betrancourt and Tversky (2000), the data concerning the effects of animation in a variety of instructional program types did not unequivocally support or refute the merits of including animation as an instructive device. In another analysis of the problem, Dehn and van Mulken (2000) reviewed some evaluations of programs that contained an animated character that was apparently giving a narration, and reported only a small effect favoring the use of the animated character compared to narration alone. Later studies involving the same comparison reported no loss in learning by deleting the animated character (Craig, Gholson, & Driscoll, 2002; Mayer, Dow, & Mayer, 2003).

Three other themes seem to have emerged. The first is that animation would show an advantage when the goal is to portray something that actually moves, such as ballet steps or the operation and repair of a mechanical device. In those cases animation could be better than video, particularly when the sequences are brief (Koller, 1992; Sukel, Catrambone, Essa, & Brostow, 2003). It has shown an advantage in teaching computer algorithms, which might require procedural steps (Catrambone & Seay, 2002).

The second theme is mental capacity, which is familiar from previous chapters. Too many audio or visual effects transpiring at once may capture global attention, but the viewer's attention becomes split between too many channels of incoming information to be able to attend to the important material properly (Mayer et al., 2003; Mayer, Heiser, & Lonn, 2001). This rule applies to extraneous music, or to text that scrolls in addition to the narration.

The third theme is that animated programs and hypermedia generally offer an advantage for students with poor reading skills or for whom English is a second language (Henao, 2002; Liu, 2004). Future research might be more definitive, however, as to whether text continues to have an advantage for the learners' retention of details as opposed to the basic ideas of the content. Students with better reading skills do not appear to share the advantages of animation. The quality of the animation, text, or other hypermedia element may be confounding variables worth considering in future human factors studies on this topic (Schmidt, 2001).

The Internet and the Web

The Internet and World Wide Web (WWW, or what some users refer to as the World Wide Wait, or simply the web) introduced some new themes in the human factors of system operation that were not hitherto explored in human–computer interaction. This section begins with a brief history of the Internet and the web. The web-related topics considered in this chapter are search engines, information foraging theory, navigation, and web page construction and usability. Other distinctive human factors issues related to the Internet are considered in Chapter 13 in the context of collective intelligence.

Origins

The Internet started in 1969 as a U.S. Department of Defense enterprise to keep the critical government functions operating during times of emergency, for example, nuclear disasters when military control centers planned to operate underground. The system was known as ARPANET in those early days. The network around the United States evolved quickly, and it included physicists (e.g., from national laboratories) and other interested parties in the scientific research community (Gilster, 1993).

Access was initially restricted to large computer sites, all of which existed for noncommercial purposes. Some large sites eventually sold services to qualified users such as educational institutions who could connect their personal computers or mainframes. In time, end users could connect to their mainframe from home by dialing a phone number through their computer's modem. By 1983, ARPANET evolved into USENET, while the military operations channel bifurcated away from the main system. Numerous networks of computers were evolving during that time throughout the world. Eventually, electronic protocols and address systems were worked out so that all systems could communicate. The irregular size and organization of networking systems still persists today.

Perhaps the most used communication systems were electronic mail, which was transmitted in plain text that looks like typewriter type (ASCII characters), file transfer protocol (FTP), and search engines such as GOPHER and VERONICA. FTP systems were particularly useful to physicists who wished to disseminate their laboratory reports to the interested community quickly, rather than having to wait the better part of a year for the journal publication system to complete its course. Computer scientists used FTP to pass computer programs around also. The anonymous FTP sites were constructed to allow anyone access to the material. The search engines were developed to search the Internet for any documents on specific topics that might be held on any of those sites.

FTP and search systems were command driven. If a search produced some results worth reading, the user could enter a keystroke and go to the computer containing the file, then use the FTP commands to retrieve the file. Files that were prepared in ASCII were readable on screen. Anything that was word processed (mainframe computers often offered this capability to their users, especially when mathematical characters were involved) or contained graphics was available in postscript format. Postscript is a data-encoding system that makes desktop printers and laser printers apply dots of ink to exactly the correct positions on a printed page. Postscript files could not be viewed on the computer screen, however. The portable document format (PDF) that is in widespread use today is essentially the postscript format with human-readable features added on. Eventually, interface software evolved and insulated the users from having to know the commands, and they could point and click instead. The residual typing only existed to insert the address of a destination computer in a dialog. In fact this transition was just another example of the replacement of command-driven computer control by point-and-click interfaces.

The network providers essentially regulated Internet traffic until it crossed a critical hub that was regulated by the National Science Foundation. The norms for use prior to 1997 were that a user could transmit information that could be of interest to other users, including information about items for sale, but general commercial use was not permitted. Private organizations could acquire access and a dot-com address, but they were subject to the same norms of use. An announcement of an item for sale was considered information when it was first transmitted, but it was considered advertising (and not allowed) on second and subsequent transmissions. It was doubtful that any Internet police would actually crawl out from the telephone jack and spank anybody for advertising, but the user-reader population at that time was so steeped in the norm of noncommerciality, that they would admonish users who broke the norm. The norm was predicated on another norm, which was the unfettered transmission of useful information, and advertising (which later became known as *spam*) sucked up precious transmission bandwidth on the network and storage space in individuals' mailboxes while it was waiting to be discovered and deleted.

In 1997 the United States liberated the use of the Internet for commercial purposes. Commercial purposes meant the sale of services, technology deployment to increase bandwidth, and lots of spam, all of which proliferated. Software became easier to use and friendlier to millions of computer novices. Some users found it to be more fun than shopping through mail-order catalogs. In any case, the user population changed considerably from the culture of users that inhabited the Internet in the 1980s and early 1990s.

Some aspects of the Internet fall under the legal auspices of the Federal Communications Commission (FCC), particularly when security and other issues affecting U.S. citizens are involved. Inasmuch as a great deal of Internet traffic occurs outside the United States and across its boundaries, there is no systematic regulation in place. Security concerns in the United States since 2001 led federal government agencies to monitor e-mail sent across the U.S. border. Reading and sorting the huge quantity of legitimate e-mail to find a suspicious item requires—you guessed it—autonomous search engines that use strategies similar to the ones the humans use.

The web was first launched in 1989 as a system for connecting the file libraries of computers into one huge library of material (Mayhew, 2003). Its technology is an outgrowth of the previous generation of search engines. Color document display and easy access made the web's contents and utilization thereof especially attractive, particularly in the wake of the sudden influx of new commercial users after 1997.

Although the web has the convenience of a huge library with searchable contents, its contents are in no way organized except by means of a user-defined search and the initiatives of content providers to supply links to other sites; the system is self-organized in this regard. Mayhew (2003) noted, and all users should beware, that web contents are unselected, unedited, and in no way imprimatured with any professional credibility simply because they are on the web. Credible and conscientious people and organizations do produce reliable material of course, but the burden of responsibility is on the user to determine what is credible. Perhaps the basic skills of critical thinking that are applied to conventional published material and speech are as applicable here on a wider scale. That said, the remainder of this section of the chapter pertains to the human factors of web operation.

Search Engines

Web users could be looking for something specific or exploring (*surfing*) for entertainment, or a bit of both at a given time. The user's first probable step is to search the web for sites pertaining to a topic of interest. The user's skill for defining a search will often affect the results, especially when topics are complex or keywords have multiple interrelated meanings and uses. The available search engines, nonetheless, will respond to whatever keywords are entered by the user. The naive user may require some trial-and-error experiences that the expert user can sidestep. Having some prior knowledge of one's topic area is also helpful.

Search engines first became available in the late 1970s as a means for searching through technical databases such as *Psychological Abstracts* or *Chemical Abstracts*. This author remembers the Lockheed Dialog® search system vividly; it was an interactive search engine that could be used on virtually any such database of technical abstracts. The data sources containing the abstracts corresponded to periodical print publications produced by professional societies. The databases still exist and thrive, although today access is more convenient in online form compared with a hand search of printed documents. The databases are equipped with their own search engines that work like Dialog®.

An abstract is a carefully written preamble to a technical journal article that captures the essence of the article's contents. The entries in the databases contained the full bibliographic citation of the article and the abstract, and each entry was tagged with an index number. A user searching for articles on a topic would search the keyword index in the publication for index numbers that corresponded to the keyword, then look up the abstract numbers. Each volume or issue of the abstract publication had to be searched separately to see what was published and indexed during a specified period. This strategy became unwieldy with the technical information explosion that occurred during the 1970s, if it was ever wieldy beforehand. The automatic properties of a computerized search that covered all the issues of the database in one fell swoop were highly desirable.

Literature search engines were, and still are, based on set theory. A user would enter a keyword, and the system would cough up the number of database entries that were found and give this set a number. The user could then enter a second keyword, and a different set would be generated with a set number. At this point, the user could specify "and" or "or" to combine the two sets into the union or intersection of the two sets. The result would be designated as a third set with the total quantity of qualifying items designated. The user could proceed at this point to view the abstracts, and then make personal selections of which items should be pursued for the intended purpose. The user was advised to be aware of synonyms for keywords, or for technical words that were close in meaning. For instance, "human factors" and "ergonomics" have different shades of meaning that might be relevant to some people but not for others; someone searching the literature might do well to put an "or" between those two keywords before continuing the search.

Literature search engines did evolve slowly, and were only available to library staff in most universities. Eventually, when all university faculty had desktop computers and the proper connections from their offices, they could all do their literature searches with just a swivel of the desk chair. *Psychological Abstracts* and many other databases are now accessible from a web-based technology that is operated by different software companies, which offer additional controllers for limiting searches, expanding to broader topic areas based on keywords selected and packaging up the selected items for the user.

The contemporary forms of the source databases are often augmented with additional web links that would allow the user to retrieve the actual documents. This feature is limited by contracts and agreements with the original publishers, and whether the library that hosts the search engine also has a paid subscription to the material. The phrase *free access to information* was originally intended to mean "easy to acquire because you can find it and retrieve it," but that meaning was too often conflated with "you do not have to pay for it one way or the other."

The Dialog® concept was built for systems that had direct connections to the database providers' computers, usually through a TELNET function. Searching Internet sources, on the other hand, required a different strategy. The GOPHER system allowed the user to browse the file indexes of computers that were connected, see menus of available materials, and retrieve (download) them. This was more user friendly than requiring that a TELNET address be known, remembered, and typed next to a cursor. (We all kept little cheat sheets in our desk drawers with favorite access coordinates.) GOPHER allowed the start of a navigation process, but with all the menu trees that looked alike on the users' screen the users had to remember where they were located at any point in cyberspace; thus, GOPHER did offer a bookmark capacity. The set theory method of narrowing did not apply, but some capability of shaping a search strategy was made possible by VERONICA (Very Easy Rodent-Oriented Netwide Index to Computerized Archives; Gilster, 1993).

There are some contrasts between the search strategies meant for a fixed data base of published scientific documents and those available on the web. The contents that the engines search are the contents of the web sites. They place a priority on the home page of the web site, and there does not appear to be any identifiable consistency to the depth of processing within a site. Although they are built on the GOPHER and VERONICA strategies, they are limited by the file structures of the web sites themselves. Goodies that are buried too deeply past the home page will be missed by the rodent.

The user enters one or more keywords at once in the dialog box, and the system responds with web addresses and a brief description that comes from the web site's text. The search request automatically connects multiple keywords with an "and" gate. The results are presented in an order selected by the search engine that reflects the probable relevance of the site to the user's stated request, starting with those that meet the "and" criterion. Searches of this type pull up more irrelevant sites than a strictly set-theoretical search. Although this strategy may facilitate browsing by a user who is not pressed for time, it can be daunting to see the note saying that 5,600 sites were collected in response to a query, especially if nothing resembling the goal of the information search shows up on the first two or three screens of material.

Academic search engines produce results that are almost always formatted the same way—title, authors, publication source, abstract—organized in the user's choice of ascending or descending chronological order. Web search engines do it differently. They produce page titles; authorship information is random unless an author's name was a keyword, a line or two of text containing the search terms, and the uniform resource locator (URL). Although the search engines respect the "and" gates to some extent, the algorithm prioritizes results by how many links there are to and from the target site. If one were to search for a site by name perhaps because the URL is not known, the target site might not come up on top of the search. Instead one might obtain many sites with "news about" the target. Many of the intruding sites are there for advertising and marketing purposes. Sites with strong commercial interests often pay for web services that magically enhance their placement within a search, or just position their ads on the search results page.

Information Foraging

The behavior patterns that people use to find information bear a substantial resemblance to the patterns that animals and our prehistoric ancestors use to find food (Pirolli, 2007). The food supplies are self-organized into ecological *patches* that support the growth of the food source. Some environments have more patches, and some patches are more profitable with respect to the amount of food or information it can deliver relative to the amount of work it takes to collect the information (food) and find one's way to the patch itself.

The forager's behavior is shaped by a certain need for efficiency in terms of both effort and time constraints. (Less time spent starving is better than more time starving.) Whenever possible, the forager goes to the most productive-looking patch first, exploits it until the law of diminishing returns sets in, then moves on to another patch. Recalling Tolman's (1932) principles of expectancy, valence, and instrumentality, the rat that has learned the maze knows where the cheese is. The rat goes to the place where it has learned that the largest cheese exists with the greatest likelihood. If he finds none, he moves to the place in the maze (patch) with the next-highest cross-product of valence and expectancy.

A forager who is new to a particular information landscape (maze) might not know where the good patches are, but could stumble on a good one early on. In Pirolli's (2007) theory, the forager sets out on a random walk to find patches and picks one that looks good relative to the information search (dietary) goal. When it is time to move, however, the forager is unlikely to know where to go next, but can rely on *distal stimuli* that suggest, based on experiences with other information mazes, which places to try first. The guess might be accurate or not; one can visualize a contribution of signal detection theory here whereby the forager determines which patches are correct hits, false alarms, and opportunity costs. The profitability of a foraging expedition is a function of the number of patches one can visit which is inversely related to the time spent finding the sites to visit.

Opportunity costs take the form of diverting time and attention to bogus leads when they could be better spent by exploiting an information-productive patch. That brings us to the logistics of junk mail. To paraphrase one of Pirolli's examples, imagine that a business person receives job orders (where the real profits are located) in the mail, but the mail is also overwhelmed by advertising. The advertising itself could contain something of possible small value, but the job orders are clearly large and positive. When the rate of real job orders is high, the time spent on sifting junk mail is not worth the effort. When the rate of real job orders ebbs to a low, however, the sifting junk could be worth the time to process it. Of course this parable assumes that the individual has only two possible tasks, the real job in one possible form, and sorting junk. The business reality often includes maintenance tasks that have to be done eventually and actions that could lead to new business. The consequences of not doing one or the other could very well trump the utility of any junk mail.

Information is hierarchically organized for two reasons. The first is that organization adds value to a morass of relatively raw data by showing relationships among ideas that usually range from broad to specific; this point is developed further in Chapter 12. The second is robustness; changes in information at the lower levels of the hierarchy have little impact on the broader concepts until a critical mass of change is reached. Information is organized around hubs and authorities. Just as airlines target certain cities to be their more central locations through which flights pass in and out to go somewhere else, some information centers serve the purpose of drawing people to a well-groomed collection of links to other related valuable resources so that the forager wants to go there first and then choose a direction for further foraging. In the scientific literature, the review study, which pulls together results from primary research and extracts themes going on, trumps other single articles. It is a likely read for exactly that reason plus it serves as an inventory of all the meaningful studies that were published before it; the forager can then follow particular themes backward in history as necessity seems to dictate at the moment.

The medium for measuring the influence of a journal article is the number of citations to that article found in later articles. The analogous metrics for web sites are the number of hits to the home page and the number of links from the site to other places and the number of links. The distribution of node sizes within a network is distributed as an inverse power law, with a large number of small sites and much smaller numbers of larger sites (Barabási, 2002).

In addition to what has been expressed to this point, information foraging theory (Pirolli, 2007) captures a number of mathematical–statistical phenomena that can be saved for further reading. Importantly, however, the principles were supported by studies of actual web user experience. Of interest, the number of sites a person visits in the context of a search also looks like an inverse power law with many searches ending at one site, but with a long-tailed distribution ending around 30 site visits. If the connection between power laws and self-organizing phenomena has any generalizability, it would appear that not only do sites self-organize in their information content, but networks of sites do also as Barabási, (2002) indicated, and the information forager is also following a nonrandom pathway. Pirolli (2007) characterized the search in progress as "getting the scent" for the information target. Once the foragers get an idea of where the goodies are located or supposed to look like, they follow the trail and ignore distractions as well as possible.

At this point we can isolate some pointers on human cognition that would draw people to a particular web site. First, the user should have a good mental model of the structure of information that exists on a topic in the cyberworld. The site designers do not have much control over this mental model, although they might consider how they should make their sites more responsive to users' mental models. This challenge requires knowing something about the user population and its diversity of experts, novices, and casual travelers. It may be instructive to find out what keywords the searchers might use, and then see how the web site might fare in the listing of search engine results under those keywords. One trick that some designers use is to position a group of plausible keywords in white type in the white space on the home page. The user does not see the scramble of words, but they exist enough for the search engines to find them.

Navigating the Site

Navigation in a terrestrial sense involves finding one's way though a physical space from one point to another. One begins by finding landmarks and orienting toward or away from them. As learning of the space improves, it is possible to specify routes from one point to another and to select among alternate routes to reach the destination. The same principle applies to web site navigation, although the physicality of the analogy is a bit weak.

For sites that are rich in content, options, and things to do, the merits of the site can be overshadowed by the sheer complexity of the site's architecture. A standing recommendation is to reduce complexity whenever possible (Whitaker, 1998). In practice this means designing the site with a simple linear layout of the main sections of material. Two landmark structures are the home-page button and the top-level menu of options. Both should be handy to grab for a user who might either be getting lost or simply choosing to leave an area and look at another one.

So much for getting started and getting lost. Once a user has finished reading the material or is otherwise at the bottom of a file, an availability of a button for "top of file" and "home page" is often good enough. Designers should also consider whether some material that was initially planted in one area of the web site might be useful to someone prowling around a different area. Links across the site should be strategically located. On the other hand, files that are not available, or inadvertently moved during an update, should not be linked. Check the integrity of the links.

Some site designers see an advantage to linking to other sites on related topics, whereas other designers do not. One motivation for providing the links is to position the site as a place to find links and give it an appearance of centrality. Perhaps the goal was in fact to provide a set of links to a specialized group of users. On the other hand, there are two downsides to external links. First, sites that are all linkage and no unique content will not receive many return visits; users will bookmark their favorite links and just go there.

Second, links to outside sites exist at the mercy of other site operators. Sites are erected, torn down, or moved at unpredictable moments, leaving the site with the links in a situation known as *linkrot*. The solution is simple: Check the links regularly and erase those that are no longer operative. The wisdom from architecture applies here: Less is more.

On the other hand, links that are really active should be identified as such. The common designations are to underline the words that comprise the link or to change color and font style. In some sites, there are only mouse-overs, which the users might not notice unless they moused over everything (Flanders, 2004).

Web Pages

The very existence of a web site known as www.webpagesthatsuck.com is a telltale sign that design atrocities happen every day. Its author, Vincent Flanders (2013), contends that the best way to teach good web design is through the example of bad designs and then figuring out what can be done about them. Problems include "accessibility, usability, and regular design problems including general ugliness." The site features bad designs that are reported by web surfers regularly in a column called, "The Daily Sucker." The site also features the top 10 atrocities of web design for a given year. The problem of internal and external linkrot is alternatively known as "Mystery Meat Navigation." Other common flaws are poor graphic design and typography, bad color contrasts and juxtapositions (color might not add information, but it can be a severe detractor), and the use of flash graphics that have no function beyond dizzying the viewer.

Some aspects of good webpage design rely on perceptual psychology and graphic artistry. Bertin (1981) proposed a scheme for evaluating graphic designs that carries over well to webpage design, according to Czerwinski and Larson (2003). The seven aspects of a graphic design are form, color, size, textures, orientation, value, and position. The Gestalt laws of perception (Chapter 4) are good guides for construction and perception of form, whether we are considering the form of a particular image or the formatting of smaller objects within a larger display.

Color has been considered already in the context of human–computer interfaces, and the same sensibilities apply here. *Value* in this context pertains to the attention-getting properties of an image element, either by virtue of color brightness, graphic style, or other techniques use of animation. Similarly, too much flash on an object relative to its actual importance to the site or page undermines the goal of moving viewers to where they want to go because what they see overwhelms what they need to see (Cloninger, 2002).

The appropriate size of objects is relative to the other objects that must appear on the same screen. Larger controls or displays designate relative importance. At the same time, there will be lower limits at which small objects will be missed because they are too small or because they are overshadowed by larger ones.

Orientation refers to the positioning of an object relative to other objects in the display (Bertin, 1981). Is it on an angle? Is it close to or away from another object? Is it positioned in some way as to convey depth? Do we care about those positioning cues or are they conveying pointless information?

Textures refer to the level of detail on the surface of an object. A relatively detailed photograph juxtaposed to an area of large text conveys a distinction in texture. Texture attracts attention to the image as a whole, assuming all components are meaningful, and breaks up the monotony of solid text.

Position refers to the specific location of an object on a page. Apparently, the upper left and lower right corners command the greatest attention from the viewers (Czerwinski & Larson, 2003). Advertisers might want to place their ads in those positions, but the host site might want those spots for their own purposes. On the other hand, a bouncing animated ad in a key position could distract attention from the serious purpose of the other contents of the web site.

Czerwinski and Larson (2003) noted that cognitive variables that may have one form of relevance when they are considered in isolation in an experiment have a different impact when combined with other variables in a web experience. The parameters of graphics are one example. Another is the number of menu options. Although the magical number 7 ± 2 seemed to be a good rule of thumb for determining the number of menu items, using only the magic number of menu items when there are many items that need to be included at the second and third layers only presents confusion to the user as to where to find a particular object. Czerwinski and Larson obtained better results with a layout containing 16 first-level menu items under which the organization of final objects was logical from the viewer's point of view.

Animation was not part of Bertin's (1981) original graphic scheme, but it is part of the web experience. The points raised earlier concerning animation in programs of a broader nature would apply here. The cognitive load issues have emerged in studies of commercial web sites as well. Hong, Thong, and Tarn (2004) reported that site visitors did manage to find flashing items more quickly, but there was no visible impact of flashing on recall. The differential demand associated with flashing items should be reserved for the most important aspect of the site page, if it is to be used at all. Flash diverts attention from other items on the page. Flash and animation are not always in the site's best interests. Commercial sites that take too long to load up due to pointless high-bit-density graphics, or for any other reason, are going to lose customers or viewers (Karayanni & Baltas, 2003).

Interactive Pages

Some of the good advice about interactive web sites dates back to the earlier days of interactive interfaces (Galitz, 1993). Some interactive interfaces are intended to be analogous to paper-and-pencil business forms that are already in use. Indeed, some people might prefer to use the paper method instead of the computer interface. Tax forms would be an example. There is utility to some users to log into the Internal Revenue Service and enter their personal data there. Other people are more comfortable filling in their forms on paper and then mailing them in. The point here is that the computer interface should be structured the same way as the paper form. This plan will make use of the user-based experience and expectations (population stereotypes by any other name), and it will also facilitate the work of people who are going to use the data after they have been entered. Note that there are two sets of users in this set of applications.

When designing a computer-based form from scratch, Galitz (1993) recommended that the form be devised so that it requires very little formatting on the part of the user. If a decimal point is part of a requested number, the decimal point should appear automatically in the correct place so that the user is not required to type it. Drop-down menus for fields for portions of address such as U.S. states are handy. They typically insert the two-character postal abbreviation instead of the full spelling of the state. This is, on the one hand, just an example of a drop-down menu replacing keystrokes; on the other hand, it puts data into a storage field that can be accessed and reformatted to produce mailing labels (as an example). Drop-down menus are usually hidden under a downward arrow, and they do not interfere with the layout of the form when they are no longer needed.

Data-entry fields should be labeled. As with labeling controls on a conventional machine interface, the labels should be consistently located relative to the empty field, for example, above, to the left, or below. Locating the label below the data-entry box is the least popular option, however. Many other points about producing a questionnaire are relatively the same as those associated with paper-and-pencil surveys; thus, the reader is referred to that body of literature. Remember: You want to be able to compute everyone's responses as well as collect them, so ask the right questions in the right manner to get the information that you need in the form you can process later.

An important situation occurs at the point where the actual sale transpires. Vora (2003) introduced the concept of friction-free purchasing. The analogy is that when we go to a store, we spend time driving through traffic, finding a parking space, looking for the items, comparing alternatives, standing in line at the cash register, and driving home. On our less fortunate occasions, we return the item to the store and buy something else instead. These are all types of friction during the purchase experience. A good sales web site should make it easy to find the shopping cart, investigate the selections, monitor what is in the cart, remove things from the cart, and make the final purchase. The action sequence for the final purchase should be insulated from advertisements, customer surveys, and other distractions. Importantly, Vora reported that about 80% of online shoppers shop that way because of time constraints, and about 50% shop online in hopes of finding good deals for less money. A site can gauge the effectiveness of their shopping system, according to Vora, by the number of abandoned shopping carts, which indicates purchases that were initially intended but ultimately not completed.

The shopper should receive a confirmation that the order was received and some information about shipping time and how to get customer service if necessary. Ultimately, the process should instill trust and confidence in the vendor and the vendor's computer system. This writer observes that computer-based business may have evolved with the intent of saving time and money on the part of both the shopper and the seller, but the process is often coupled with degraded customer service compared with what people had come to expect from real-world sales experiences.

Extreme Graphics

To some aficionados of the graphic arts, web site designs are often uninspired, although they would get a free pass on their aesthetics if they were actually functional. There are artist enclaves, however, who recognize the web site as a unique medium of expression and are actively experimenting with graphic design techniques and styles for this medium. Graphic designs can collide mercilessly with usability, however (Cloninger, 2002). Although it is not the most extreme style ever offered, the *Mondrian style* is radical nonetheless and features page layouts that are organized around rectangles of different sizes and colors. Select one, and the screen re-organizes, treating each new page as a pop-up window. The problem here is that anyone trying to link to an internal page would be unable to do so. There is also the problem of having too many open windows, which can overload a viewer's system eventually. There are also styles that are radical in the graphics, but very predictable in their actual functionality. Some of the radical styles find uses in product sales, fashion displays (e.g., apparel), or showcasing artwork.

The *pixilated punk rock style* is perhaps the most extreme style, as most of its distinguishing features are intended to rattle and disorient the visitor. Some of its antics include fake malfunctions, keyword searches that produce nonsequitur results, missing or misleading labels for links, random links, excessive use of animation and audio noise, unusual image cropping, image enlargements that take the viewer to the pixel level, full-screen images with no visible navigation clues, and unlabelled mouseovers where clickable links might ordinarily be found. The pixilated punk rock style does not appear to have found any applications other than as a performance piece that compels us all to question what we are doing when we are building a site:

> [P]ixelated punk rock is a usability expert's worst nightmare. These sites give their visitors a false sense of control, all the while manipulating and coercing them at every turn. But wait a minute: Haven't I just described most modern advertising in a nutshell?… Should user empowerment and usability be the goal of an e-commerce web site? Do you really want to encourage surfers to wander around your online stores willy-nilly, reading this FCC-mandated earnings disclosure or that unmoderated product review? Instead, don't you want to funnel them as quickly as possible to that bright, candy-like "order now" button? (Cloninger, 2002, p. 119)

The quick visit with extreme graphics brings us to another point before moving onward: Context matters. The design of the site, the user population, the intended tasks are all affected by fundamental differences between sites that provide information resources in the sense of the global library, product information leading to sales, carrying out an automated government service such as renewing the registration on one's automobile (Bennett & Flach, 2011). There is also the question of what else the users were trying to do when they jumped on the web to look for something. The earlier example of walking while using a cell phone is perhaps just one example of a task being really part of a dual task scenario.

Virtual Reality

Virtual reality (VR) technology is an outgrowth of numerous artistic and technical developments from prior decades—theater, cinema, perspective drawing, pinball machines and arcade games, stereoscopic photography, 3-D auditory displays, and animation, not to mention all the computer-programming skills and hardware that make it function. Burdea and Coiffet (2003) defined virtual reality as "a high-end user–computer interface that involves real-time simulation and interactions through multiple sensorial channels" (p. 3). They made an additional point of distinguishing VR from intermediate technologies such as those where a virtual component is superimposed over a view of actual reality, or telepresence, in which images from remote cameras are assembled into a portrayal of the actual reality onto a cinematic viewing screen. The former type of system does not qualify as VR because the actual reality still peeks through; the experience is not really simulated. The latter type of system may provide a breathtaking viewing experience, but it does not allow for interaction between the viewer and the program. In other words, it is one thing to view the Martian landscape, and quite another to build a sandcastle on it.

Since its inception, VR has been characterized by two themes: *immersion* and *interaction* (Burdea & Coiffet, 2003; Pimental & Texiera, 1995). Immersion of sorts occurs when we read a fascinating book or watch a good movie, but in VR it means, far more literally, that the system user becomes part of the program. Some science fiction movies have captured that concept well in a hyperbolized form. In fact, the interplay between science and science fiction dates back over a century to the classic works of Jules Verne and H. G. Wells, with many mind-expanding episodes since then (Turkle, 1984). Interaction does mean that the user can introduce actions to what is essentially a movie script, and the script responds. The program behind the script is prepared for the user's presence, and can do things to the user (in virtual form) as well.

The VR concept became popularized with the applications to games and entertainment. After all, how else could we stretch the baby's toes so that she will be easier to grab, ride a buggy through the deep ocean without an environment suit, or play croquet with the boss' head (Kelly, 1995)? After all, anything can happen in a cartoon. The rules of reality can be used, modified, or totally rewritten. Although the design of the virtual worlds themselves falls outside the scope of this book, some of the fundamental principles of program rules are considered in the next chapter.

On the other hand, Burdea and Coiffet (2003) emphasize that VR exists to solve problems in real life. Practical applications include medicine, industrial skill training, working in extreme environments such as outer space, hazardous waste removal, and military training (Pimental & Texiera, 1995; Bergamasco, Bardy & Gopher, 2013). The practical applications require a maximally close parallel between the rules of the virtual system's operation and the rules of real system operation. Some of the more reality-grounded rule groups include the angles of incidence where a human user makes physical contact, or uses weaponry against, an agent (or avatar), or other elements in the program or vice versa, and the duration, force, and implications of a physical contact. The combination of sensory experiences should provide enough sensory input, and input of sufficient quality to give the operator a sense of "being there." A large number of sensors would be needed to permit the types of changes in perceived locations of sights and sounds that occur as the individual moves through the environment. The system should be transparent in its delivery, meaning that the equipment generating the virtual experience introduces as little additional input as possible so that the human experiences the virtual world and not the machine (Bargamasco et al., 2013).

Several human factors issues became apparent in the development of the new controls and displays for VR equipment. Some problems were obviated with better hardware when the designers recognized the problems right away. VR has also opened up two new areas of psychological study—haptic perception and emotional programming.

Helmet and Wall Displays

If the distinction between a display and a control became blurred with GUIs, the two become more intertwined in VR. The user becomes part of the display in many instances. Furthermore, the concepts are not limited to one user and one machine, which has been a convenient assumption in earlier chapters of this book.

The VR helmet has probably become an icon of the industry, although it reflects a technology that is more than a decade old. Nonetheless, the basic VR helmet is an extrapolation of the heads-up display. The design can take the form of an intermediate technology, such as the superimposition of diagrams and reference material over a real visual image of a piece of equipment that an operator wishes to build or repair. It becomes VR, however, when the actual reality is blocked out, and all that one experiences is the content of the simulation. The discoordination between movement and the corresponding visual input is the primary source of simulator sickness.

The graphic content of the helmet display was limited by the graphics-processing capability of the computers to which they were connected. The original equipment used wire frame images. Pimental and Texiera (1995) remarked that if our regular vision were that poor, we would be declared legally blind in most states. The technology evolved to accommodate solid box diagram images, so that a wall could appear solid rather than transparent. In another leap forward, surfaces of walls and other objects could acquire the level of texture that distinguishes a bookshelf from a solid blob. The resolution of 2-D images is expressed as $N \times N$ pixel density; 3-D images are calibrated in polygon density. Higher graphic realism is related to the polygon density.

The VR helmet is equipped with visual tracking devices. The image presented to the helmet display depends on the orientation of the head, just as it would in real life. Figure 11.10 depicts a classic example of the control–display relationship. The pilot turns his head toward the direction of a target, squeezes the control, and the airplane artillery fires directly on the target—more or less ("Oops, there goes the orphanage … Sorry kids, but there's a mosquito in the cockpit … I'd better file a report on that …"). The early equipment had a significant delay between the moment of the head movement and the corresponding display on the viewer. It was difficult to walk or navigate through the virtual world when the visual feedback, which is instantaneous in the real world, was delayed.

VR helmet designs are still in use, but they are being replaced by two other forms of displays. One replacement design resembles a set of eyeglasses with the exception of a small processor that sits on the nosegrip. This type of relatively lightweight equipment is particularly adaptable to situations where the viewer is engaged in a real-world task, such as surgery or biomedical viewing of regions deep inside the human body. The virtual

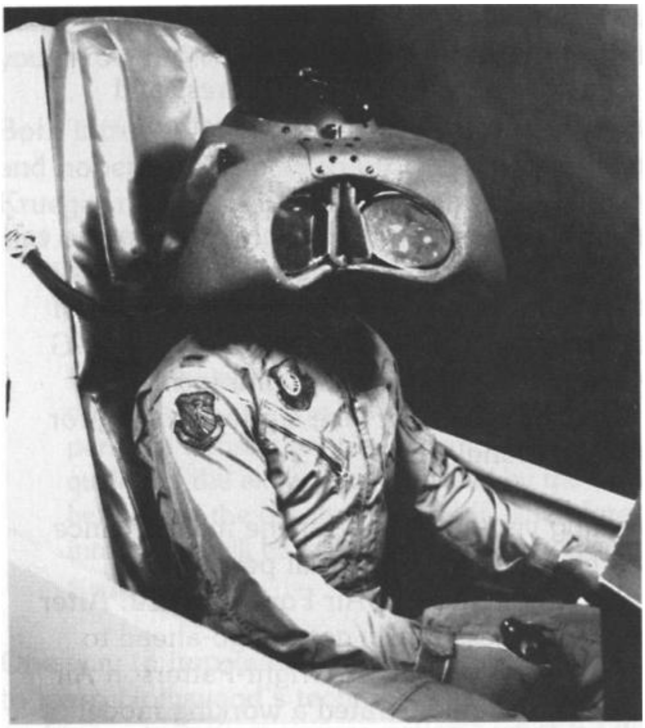

FIGURE 11.10
The helmet-controlled aircraft artillery aiming device. (Photo by Wright-Patterson Air Force Base, in the public domain.)

material comes up to the screen when the user turns the display on. The annoyance should not be much greater than that associated with bifocal glasses, at least in principle.

Another type of display format is the wall-mounted viewing screen. This class of displays can be as small as three standard computer screens—minus the space associated with the frame between them—or as large as a cinema screen. The full-wall screen can be a fairly good substitute for the VR chamber, but the immersion is not as complete, all other things being equal. In this context the user enters a room that is equipped with all the visual, auditory, and tactile display elements and interacts with a program that surrounds the user in a literal sense. The 1996-era systems required a network of projection cameras that cast portions of the image onto the correct locations in the room. Computers can now generate visual displays in convincing formats onto overgrown desktop displays that are mounted on a wall.

Wall-mounted displays still require a visual tracking device if a helmet is not used. The distance of the human to wall needs to be computed in order to display the images within the resulting field of vision. A field of vision of 37° appears to be the desirable target presently (Burdea & Coiffet, 2003). Figure 11.11 shows a wall display that is part of an experimental set-up for studying gait and its correspondence to visual cues (Katsavelis, Mukherjee, Decker, & Stergiou, 2010). The objective of the system is to build better methods for rehabilitating people with impaired mobility caused by stroke or other medical conditions. The visual images are driven by cameras that watch the person on the treadmill, which is programmed to deliver specified routines. The images are stereoptic, and

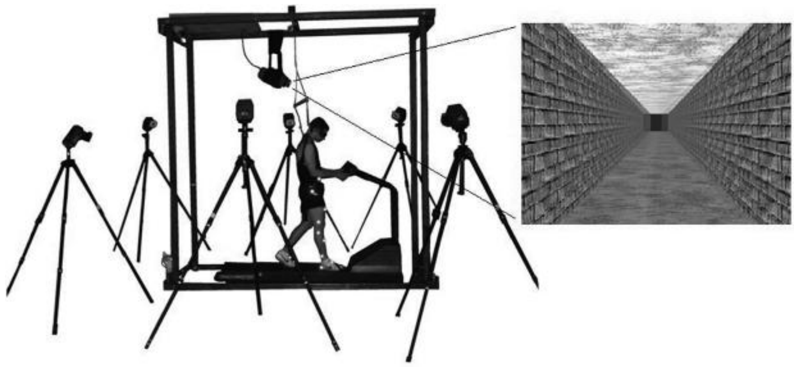

FIGURE 11.11
Setup for wall display–style VR. (Reprinted from Katsavelis, D., Muhkerjee, M., Decker, L., & Stergiou, N., *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 165–178 and 239–256, 2010, with permission from the Society for Chaos Theory in Psychology & Life Sciences.)

the person wears 3-D glasses during the episode. The images that are projected emulate the experience of working through a real environment. A similar setup has been built to train athletes to row in a team (Bergamasco et al., 2013).

Wall-mounted displays can also be used for more conventional purposes such as showing multiple media in windows for groups of viewers simultaneously. A miniature desktop version is currently available. The miniature contains the spatial equivalent of three larger sized VDTs without the frame boundaries between them. Pamoram Technologies, its manufacturer, claims that the new format of display represents a new paradigm in computer equipment. The VDTs of the past have been peripheral devices to the computer, which was the central piece of equipment. The new display system is now the center of attention (and cost), whereas the computer and other input devices are the peripheral attachments. Ideally the multiscreen display systems, such as the one shown in Figure 11.7, could be replaced by a wall-sized version of the same type of display. The power plant control station in Figure 11.9 would still have a way to go before it qualified as VR, however. Nonetheless, VR spin-off display technology has other possible applications. The large display systems also purport to be very useful to VR programmers whose virtual desktops are packed with complex files.

Glove and Body Controllers

The data glove shown in Figure 11.6 is a device from the mid-1990s. The idea behind the glove was that it is used in conjunction with a head-mounted display. The user could view objects within a 3-D space, reach out and grab an object and move it around; the results of those actions would show up on the visual display. With this basic equipment package a user could interact with a program displaying the current medical condition of a human body, take virtual surgery actions, and see the results—assuming the program contained the proper biophysics knowledge base, of course.

The early data glove designs gave rise to the data suits. In one classic illustration, two users could put on the suits, move their arms and hands, and the actions of one person would show up in the display of both of them. To make matters more amusing, the program portrayed the two users as lobsters in an aquatic environment. Although the demand for equipment that would allow us to become lobsters is probably still small, the technology advance was primarily the interaction of two or more users in a system simultaneously. Tactile feedback from the data glove or body suit was eventually added (Pimental & Texiera, 1995).

Currently there are several alternative technologies available for data suits, each with its own assets and limitations for data acquisition and translation. This class of devices is essential, nonetheless, for the immersion of a user in a VR chamber, in which the equipment must track the position of bodies and their detailed movements (Burdea & Coiffet, 2003).

Anthropometric Issues

If the hand is too small relative to the glove, the sensors in the glove start to wrinkle over each other, producing a dilapidated image. The solution, albeit expensive, is to prepare gloves in different sizes.

The manipulation of objects in a virtual environment is limited by the reach of the person's arm or hand. This is probably not a problem when someone is using a finished program; in fact it would be consistent with reality if physical reach did limit the user's actions. It creates a bit of a problem during the programming stages, however, especially when a wall-mounted display is depicting wall-sized objects. The cubic mouse (Figure 11.12) was developed as a response to the situation. The cubic mouse facilitates a combination of pointing, capturing, and rotating in 3-D space, and moving objects in the virtual space. The model shown in Figure 11.12 contains six action buttons.

FIGURE 11.12
The cubic mouse. (From Fröhlich, B., Plate, J., Wind, J., Wesche, G., & Gobel, M., *IEEE Computer Graphics and Applications, X,* 12, 2000. Copyright 2000 by IEEE. Reprinted with permission.)

Haptic Perception

It is one thing to hold an object, and quite another to let it slip out of one's hand or squeeze it to death. Systems without the proper tactile feedback are susceptible to these failures. The human haptic system is sensitive to temperature, pressure, and texture. Temperature is encoded into the tactile delivery system with heat sinks (Wesler, 2004).

The pressure and texture functions are currently running into the limitations of the human sensory system. For instance, two stimuli of 1-ms duration must be separated by 5.5 ms in order to be perceived as two separate stimuli on the same location. Two successive stimuli on different locations must be separated by 20 ms in order for the operator to detect which stimulus arrived first (Jones, 2001).

We can "see" our way around in the dark if we use our hands to follow walls and furniture. On a good night we can find the flashlight in the drawer when the power goes out, also by feeling our way around. The psychological determination of 3-D objects requires a combination of pressure and texture sensations. The latter is generated by vibration. The psychological theory that underlies how texture and vibration sensations are coordinated into perceptions of objects is largely unknown. Complicating matters a bit is that the information that contributes to the perception of an object involves not only the pressure and texture information, but also the movements that the person used to arrive in contact with the object (Smith, 2001).

"He doesn't know his own strength." Real-world clumsiness can be complicated another notch in the virtual world. Haptic feedback systems need to be adjusted for the strength of the user, and clear rules for this aspect of human–machine coordination are still in the making. Systems that provide feedback that is too strong can harm the operator. Operators who are too strong can damage the equipment. These design issues become complicated further if the system is meant to be used under conditions of stress (Smith, 2001; Wesler, 2004).

Training Systems

The fidelity of perceptional phenomena, subjective awareness, and performance or system response are extremely important in training systems. The classic rule for transfer of training is that training is optimal to the extent that the stimuli received, behavior produced, and the rules of reward, cause and effect are exactly the same between the training situation and the real one. Learning that is dependent on a feature of the training system that does not have a counterpart in the real world, or is too different from that of the real world, could transfer negatively (Bergamasco et al., 2013). Sometimes it is better to leave a stimulus feature out of the training environment altogether if it is too dissimilar to the real condition.

In other cases, where the training situation utilizes hints and prompts for the beginner, the support feature can be phased out of the experience in much the same way as training wheels are removed from a child's bicycle. Another support structure for learning involves the use of accelerators, which control the speed of the activity. An example would be juggling very slowly at first so that the learner can grasp the phasing of motions.

Emotions in Human–Computer Interfaces

Emotions have only played a small role in human factors until recently. Historically, the concept entered the scope of human–computer interaction in the mid-1980s under constructs such as computer anxiety. During that epoch, many organizations were introducing new computer systems to offices and as part of automation systems for production. Some employees responded favorably to the changes if they saw them as an opportunity for a career advance or a skill enhancement, while others were less happy about the idea of their skills becoming obsolete. The obtuse computer command systems and unfriendly interfaces in vogue at the time could not have helped matters.

Stress and Anxiety

The roles of stress and anxiety in work performance and system safety were described in Chapters 9 and 10. Adaptive interfaces (Chapters 2 and 6) detect stress on the part of the operator and alter the appearance or configurations of displays and controls in response (Klein, Moon, & Picard, 2002; Hudlick & McNeese, 2002; Khalid, 2004; Schmorrow & Stanney, 2008). Possible responses of the machine system to emotional information would include altering the information load that is facing the operator and engaging additional levels of automatic decision processing. At the present time, adaptive interfaces only respond to stress levels as determined by physiological indicators of cognitive workload; they do not make any particular use of the colorful range of human emotions that usually come to mind when humans interact with each other. Practical applications are limited by the current status of psychological theory that connects cognition with emotion and by measurement techniques for momentary changes in emotion or mood.

A completely different role of emotions shows up in VR applications for curing phobias. To be brief, the premiere approach to curing phobias relies on a process called systematic desensitization. In this process, a person is given increasing exposure to the source of their fear, for example, heights, spiders, or flying. Some of the exposures are mental imagery exercises during therapy, and other exposures are homework assignments such as to go visit one place or another. VR systems have the potential to deliver systematic exposures in the office environment. For instance, a patient can put on the helmet and take a virtual elevator ride instead of driving around town to find suitably incremental elevators. Flight simulators lend themselves to therapies for fear of flying. Again, it is too soon to make pronouncements with great assurance, but evaluation studies on record indicate that a VR-enhanced therapy can be more effective than traditional means alone (Anderson, Rothbaum, & Hodges, 2003; Wiederhold & Wiederhold, 2003). VR systems for social phobias, on the other hand, place a greater demand on the realism of the humanoid avatars within the program and the emotional expressions that they exude (Rizzo, Neumann, Enciso, Fidaleo, & Noh, 2001).

Emotions as Information

The clinical application brings us to some additional roles for emotions in VR. A VR system is more convincing to the user if it gives the user a sense of presence. Presence is currently an active research area, and is thought to be related to the realism of the animation, graphics, and scripts. Part of the realism, however, involves the extent to which the user's emotions are invoked during the virtual experience, and the similarity of that experience to real-life emotional triggers (Riva, Davide, & Ijsselsteijn, 2003; Slater, 2004; Bergamasco et al., 2013). Realistic emotions triggered within the user should be met with compatibly realistic responses from avatars; users appear to be more fond of empathetic avatars than self-centered ones (Brave, 2003; de Rosis, Pelachaud, Poggi, Carofiglio, & De Carolis, 2003).

Although the technology of speech synthesis and speech recognition has improved greatly since they were first introduced, naturalistic interactions between people and machines are limited by the machine's ability to recognize laughter and other subverbal and nonverbal communication elements such as vocal inflections and the facial expressions that usually accompany them. This aspect of human–machine interaction is thought to be critical for the development of household robots or other personal assistance robots.

Some advances in laughter detection have now been reported. Scherer, Schwenker, Campbell, and Palm (2009) started with multimedia data from an online conversation. Burst of laughter were isolated in 200 ms clips along with the visual elements. A neuralnet program (see Chapter 12) was trained to recognize bursts of laughter with 87% accuracy. At present it is unknown how well the machine learning would transfer to the recognition of laughter by people other than those in the training video. Another outstanding challenge is to design the robot to engage interactively with the humans in a naturalistic fashion.

If laughter recognition for the household robot is a challenge, the recognition of other human emotions is no further out from the starting gate. There are six types of emotions that register in human facial expressions that are recognized in virtually all cultures: fear, happiness, sadness, disgust, surprise, and anger. Classification algorithms are starting to become available (Schels, Thiel, Schwenker, & Palm, 2009), but the limits of generalizability and potential for comfortable interaction still apply.

A similar research program is underway to program robots to recognize and make communicative gestures such as a hand wave (Sadeghipour, Yaghoubzadeh, Rüter, & Kopp, 2009). Humans make these gestures in different ways, but with the same meaning. The robot needs to recognize the class of gestures and possibly use them in conjunction with a speech program to facilitate natural communication between the person and the machine.

DISCUSSION QUESTIONS

1. Trace the origins of the device commonly known as computer to machines, tools, or devices that were known prior to 1945. Can the computer be considered a convergence of the earlier ideas? What new elements appear in the computer idea that did not exist before?
2. At least one human factors engineer was known to remark, perhaps sardonically, that every human action and ability can or will be reduced to two motor skills—one is point and the other is click. How often is this statement true, in your estimation? Do situations exist where it is not true at all?
3. Where in a human–computer interface would Fitts' law apply?
4. What other principles of controls, displays, or psychomotor skill from earlier chapters apply equally well to the human–computer interface?
5. If you look at the file directory on your computer, you will see mention of an A drive, a C drive, and perhaps drives D, E, F, or G. What happened to the B drive?
6. A pundit once said, "Every year the cost per bit of information goes down, but the cost per bit of useful information goes up." Explain what this statement means.
7. Many innovations in computer-based technologies occurred for reasons other than human factors issues, but human factors improvements occurred nonetheless. How many such improvements can you find hidden in the evolution of computer systems?

8. Consider the technology innovation of installing identification chips in pets. Does this technology actually solve a problem, or is it excess baggage that only increases the cost of living or doing business? Are there any other technology applications that you would consider excess baggage or unnecessary complications?
9. Consider the use of identification chips in students' ID badges by the school district in Texas. Does tracking students' whereabouts on a continual basis serve any educational purpose? Does it really solve any perceived problems? Are there any downsides to the system, either functional or social?
10. Consider Bahr and Ford's (2011) recommendations for the use of pop-up windows. Are they likely to solve problems with pop-ups as the user usually experiences them? What is the likelihood that advertisers or anyone else will follow them? Why?
11. Go to the U.S. Postal Service site, www.usps.gov, and figure out how to apply for a permit to mail a monthly magazine at a discount rate and what forms need to be filled out. How long did it take you to get this information (assuming you were not interrupted by people and events that are unrelated to this task)?
12. Describe a plan for evaluating the effectiveness of a VR system. Hint: Try answering the question more than once with different VR applications in mind. What aspects of system performance would you consider here that might not have been considered?
13. If a VR or other system is designed to manipulate the user's emotion levels, what would be the implications for the design of voice-operated controls?
14. Participants in human factors studies that involve wearing EEG helmets or sensor gloves may have been willing because of the novelty of the idea, the premise that the results might assist them in their work in some way, or perhaps because they were being paid for the participation. What do you suppose would be their reaction to the idea of developing monitoring systems that were to be worn by employees everywhere so that employers could monitor their levels of alertness throughout the day?
15. Let's consider some critical incidents in Internet sales: (a) Tell us about a purchase transaction that you made that went very smoothly. (b) Tell us about another purchase that felt like a complete abomination. What was different about the two experiences?
16. Is the concept of usability that has been assumed so far throughout this book really adequate? What would be some implications of reframing the concept to "usability for how long?" Can you think of any devices that were *too* useful? (Hint: In *whose* opinion would they be too useful?).

12

Programming, Artificial Intelligence, and Artificial Life

This chapter explores the void between the control and the display in the human–computer interface. The concepts of computer programming, artificial intelligence, and artificial life emerged in close succession, thanks largely to the work of John von Neumann. As we saw in Chapter 11, the first computer required a human to set a myriad of knobs and dials to correspond to a desired computational operation. The basic concept of computer programming allowed the user to specify and implement most of the control operations in a fast and flexible fashion (Heppenheimer, 1990). In doing so, however, there became a disjoint between the control operations required to *set the functions* of the machine, and the control operations required to *operate* the machine. For instance, what I imagine that I am doing now—typing text for this book through a keyboard interface—is substantially different from what a computer programmer did at another place in time. The programmer specified the rules by which my keystrokes converted into text that I see on the screen or on paper if I were to use the control actions for producing a print.

Von Neumann contributed his mathematical skills to the Manhattan Project, which operated out of Los Alamos National Laboratory, to build the first atomic bombs. In those days, there was a job category called the *computer,* which was a human who made computations. Computation requests were submitted from the laboratories that collected data to usually two people who were told only what the computational goals were. They were not told what the experiment was supposed to be about or how it was supposed to turn out. Modern psychologists will recognize this work technique as a double-blind strategy. The disjoint between the laboratory and the people making calculations ensured that errors in judgment would not creep into and affect the results or their interpretation. The two computers, which were required to agree, ensured the desired modicum of reliability.

The Manhattan Project was a historical landmark and the beginning of a modernizing society where numerous high-speed calculations were needed everyday. Computers thus gained early notoriety for their ability to process large quantities of numeric data. The non-numeric information-processing activities in later generations of technical systems were also known as *computations.* The same holds true for the information-processing activities of living systems.

Von Neumann's vision for the programmable computer extended far beyond the ability to make numeric computations. Some of his more poignant mathematical proofs led to his proclamation that all life could be expressed as the flow and processing of information (Levy, 1992). It would thus be possible to express all activities of life forms in this fashion and, in essence, create models of life in a silicon environment. Hence, we arrived at the concept of *artificial life.*

There was a large gap between the goals of artificial life and the technical capabilities of the early 1950s. Thus, progress had to occur in modest incremental forms. *Artificial intelligence* became the bite-sized piece. Can we use computer programs to emulate human thought processes? In doing so, we should be able to speed the process, enhance the reliability, and expand the scope of what humans can accomplish. Thus, computer scientists became very interested in psychological advances in human cognition, and psychologists became very interested in computer-based methods for testing theories about human thought.

The next sections of this chapter explain the major increments in the development of computer programs, artificial intelligence, and artificial life. The overriding principle is that machines have gradually evolved to mimic human thought, and thus the information-processing and control action sequences that humans might employ. Granted, the mimicry remains crude.

Evolution of Programs

Conceptual Levels of Programs and Systems

The dissociation between the programmers' tasks and the users' tasks evolved gradually as programming techniques became more sophisticated. The most elementary form of program was the binary code. Binary code consisted of strings of digits 0 and 1 that specified whether an electron pulse was allowed to pass through a particular location at a particular time. Programming of this nature was obviously tedious and conceptually divorced from the goal of the computational activity.

A major leap forward occurred with the development of programming languages such as Fortran, Cobol, and Basic. These elementary languages bore a great deal of resemblance to human language syntax, although there were major differences. Commands must be specified exactly in a programming language, whereas human language allows us to construct many possible statements that have the same meaning, often with different word choices. In fact, this distinction between programming and human languages gave rise to a new form of program, the natural language processor, many years later.

The elementary programming languages, nonetheless, allowed the programmer to specify numerous binary commands using familiar-enough words and base-10 digits. The 80 keystrokes that were applied to a punch card corresponded to a series of binary operations, thus producing the pattern of holes that were punched in the card. Elementary programs allowed the end user enough flexibility so that the user could input different sets of data, which would always be recognized properly if they were entered correctly. The same operations would always be executed, and the same form of output would always be received. Several forms of program development were inherent in this simple description.

Entering data correctly was a challenge for the typical human who was accustomed to a little tolerance for typographical errors. A keystroke out of place in a computer program could often shut down a program's operation with a fatal error of some sort. The errors were often difficult to find. For instance, numeric 0 was not easily distinguished from capital *O* by the naked eye. Thus, the character Ø came into common use. Similarly, the lowercase *l* was not distinguishable from numeric 1; in fact the two characters were made by the same ley for lowercase *l* on most typewriters in use through the 1970s. If anything, the need for two distinct characters *l* and 1 on a regular typewriter was propelled by the need to distinguish the two characters in the computer environment.

Human factors played an important role concerning the error messages. In order to write in a program language such as Fortran, the programmer interacted with another piece of software called a *compiler*, which recognized the proper specification of Fortran statements, and spit back error messages. It took some time before sets of interpretable error messages were forthcoming from compilers. The lack of awareness of human factors issues was not entirely a result of ignorance or indifference on the part of the system engineers. Computation time and computer memory were costly; although it was often possible for the early technologists to imagine error-checking devices, the computer memory required to execute the software was prohibitive.

People who first acquired their programming skills in a Fortran-like environment were inclined to think of the data and the program as separate entities. Although this distinction has practical value at times, one of the advances in computer science came from the recognition that there was no fundamental difference between the data and the program. This realization led to the subsequent advances in programming techniques.

The next level of advance beyond the program language itself was the entity that might be called a *shell program.* In essence we would have a program that was large and complex such that it could give the user many possible options for data input, analysis, and output formatting. The user would then have to choose the desired options for a given task. The user would enter commands with the exactitude that was required of program language statements.

The statistical analysis programs that were in common use up until the late 1990s were examples of the shell program. A user had to write a little program that told the big program what to do. The end user was essentially writing program rules that affected the use of rules in the main program; the end users' "little program" is described later in this chapter as *metarules.*

Symbolic programs allow the programmer or program user to identify a chunk of program rules, treat it as an intact object, and move it around the program at will. This group of capabilities allows us to compose a bitmap picture, pixel by pixel, compress it into another form of graphic object that is smaller in size, such as a .GIF or .JPG graphic, and move it around the file that we are using to prepare a document. The object-oriented programs and programming languages are essentially what allow us to cut and paste most anything in a document that we are composing; this capability is perhaps taken for granted by modern computer users.

Conceptual Levels of System Design

Computer-programming capabilities evolved from the bottom up, but the typical system designer organizes the design process from the top down. The highest level of conceptualization, that is, that which is most intrinsically human, is to imagine the purpose of the computer program, the range of capabilities that it should include, and the users who will benefit from such a program.

System designers have often relied on metaphors between some real-life activity or situation and the representation of those entities in their programs. Hence, we call the first screen on Microsoft Windows a *desktop.* The verbs *cut* and *paste*, which are common in a variety of word processors, originated with standard operations in the printing industry, where *cut* involved the use of rulers and razor blades and *paste* involved a hot wax machine. (At the time of this writing, Windows 8 has just been released. It appears to be abandoning the desktop metaphor in favor of a layout that looks like an overgrown smartphone.)

The second conceptual level of system development is the *semantic* level. This level of design specifies the operations that can be executed by the system, how they are to be organized, and how they relate to the experience of the user. The third conceptual level of system development is the *syntactical* level of analysis, which compiles the specific control operations that a user must execute in order to make a particular operation occur.

Finally, the fourth conceptual level of system development is the *device-specific* aspects of the program. Ideally, these are not experienced directly by the user, except perhaps during the initial setup. These features of a program recognize particular pieces of equipment in the system and make the necessary transformations to the processed information so that the end result will appear properly on the screen, the printer, or other device. The manufacturers of peripheral devices are in turn prepared to make their devices operable by as many varieties of computer operating systems as possible.

Artificial Intelligence and Expert Systems

Several principles that drove the theory behind artificial intelligence programs and the particular application are known as the expert system. Here we encounter the work of Von Neumann, Gödel, Turing, and Simon. Expert systems exist in several levels of sophistication, which are elaborated on later.

Some Basic Principles

Gödel

Gödel produced two ideas that haunted the development of artificial intelligence (AI) applications. He proved (1962) mathematically that, given any rigorous system of logic, there is at least one legitimate logical statement within the system that cannot be reached from inside that system of logic. Thus, there was no such thing as a perfectly closed logical system. At some point, there is an element of information that must be supplied from the outside, hence the system user always has a role. This logical leak is sometimes known as a *back door* to the program.

The applicability of Gödel's (1962) theorem to practical AI programs rested on an assumption that all knowledge can be expressed as a set of logical *if–then* statements. This assumption carried no intrinsic distinction between knowledge of simple facts, the kinds of mental operations by which conclusions are drawn based on those facts, and a given problem, or for that matter, the kinds of induction that are involved in creative problem solving. If, on the other hand, we could identify forms of knowledge or mental operation that could not be framed as *if–then* statements, then we would have an unprogrammable situation altogether.

Turing

Back doors often have something in common; the missing knowledge link often involves knowledge of the system itself. The Turing test (Turing, 1963) specified a hypothetical procedure by which one could discern whether a system was truly intelligent or just artificially intelligent. Two curtains would be set up in a large room. Behind one curtain would be a computer. Behind the other would be a human. Both entities would receive a question through a mail slot in the curtain. Both would type their responses to the person who asked the question. If the question were sophisticated enough so that neither the human nor artificial source could answer, the human would reply that he or she did not know the answer, whereas the artificial source would respond in gibberish.

Thus, an intelligent system required, at the very least, a modicum of self-awareness. Eventually, an element of self-awareness was introduced by the error diagnosis in the compiler program. Programming finesse eventually integrated the error reports into the bodies of shell programs and latter-day programs. Computer systems eventually acquired diagnostic programs (essentially expert systems) to help repairers make repairs, and help programs for the users. Did these advances bring self-awareness to the system? Not by a long shot, but they were useful steps forward from the vantage point of both theory and usability.

Turing also contributed to the development of a branch of mathematics called *symbolic dynamics.* Suppose we had a sequence of objects or events, and gave each object or event a letter code $a \ldots z$. We might possibly encounter a sequence such as:

$$F_1 = (a, b, c, a, b, c, d, e, d, e). \tag{12.1}$$

The mathematical analysis would allow us to extract sequences from Equation 12.1 so that

$$F_2 = (f, f, g, g). \tag{12.2}$$

We obviously get from F_1 to F_2 by allowing $f = \{a, b, c\}$ and $g = \{d, e\}$.

Real problems are seldom this simple or this obvious. Nonetheless, the math eventually facilitated some more sophisticated operations. For instance, we could identify a sequence of program code, make an operation on it to change it in some way, and put it back into the program where it originated. We can write rules for how these program changes would take place. In doing so, we also arrive at an important insight: There is no fundamental distinction between program and data. The combination of both can be called *program* or *data* interchangeably. All that matters is that symbols that are found somewhere are manipulated by rules found somewhere else, and the rules can be manipulated by other rules in indefinite recursion.

Von Neumann

Von Neumann and Morgenstern (1953) provided some substance to the AI objectives through the development of the economic theory of games. A game consists of a set of strategic options, each of which delivers *utilities* to a player. A utility is a generic form of value. The utilities associated with each option, however, depend on the actions taken by one or more other players. Games may be strictly competitive (also known as *zero-sum,* or noncooperative games), cooperative, or mixed-strategy games that involve competitive and cooperative options.

In a noncooperative game, the values of the possible outcomes for one player are opposite the values of those same outcomes for the competitor. One player's winnings are the other player's losses. Optimal decision making is thus one where one's outcomes are maximized and the other's outcomes are minimized; this is the *maximin principle.* Players do not typically share information in a noncooperative game, and binding agreements are not possible. For competitive games, there is at least one strategic condition that is a stable expression of the maximin outcome (Nash, 1951), known as a *Nash equilibrium*.

Although it is possible to conceptualize a one-person game, where one player works against the environment, or against a slot machine, the more interesting games are interactive and require at least two players. Thus, game theory considers two-person, three-person, and *n*-person games. The simplest games for two players are the two-by-two games, where each player has two options.

Of the many ways of configuring two-by-two games, some games have *trivial* outcomes. Trivial in this case means that the best option for one player is always the same option, no matter what the other player does. For practical applications, it is valuable not to underestimate the importance of a so-called trivial game or to fail to recognize one; they might not be so trivial. The received wisdom of the past few thousand years is not to underestimate one's opponent, which is generally good advice. On the other hand, it is important not to overestimate the opponent's capabilities. In a game of bluff, the power of one player resides in the perception of that power, rather than in the substance of that player's capabilities. When such games are in effect, it helps to simply go through the list of options: If A does this, what is the best option for B? If A does something else, is the best option for B still the same? Is there anything A could do that would change the best option for B?

It is only a small leap in intuition, although a large leap in human effort, to build a program around a complex game such as chess. Chess has numerous options for each player that change with every play. The value of each option depends on what the opponent could do once the play is made. It is strictly competitive. More importantly, it became the prototype computer gaming application in the 1950s. A system user could play against the computer, which acted as a highly skilled player. The evolution of computer games from chess to the graphic-intensive games that we have now is easily imaginable, but beyond the scope of the present discussion. Further topics on cooperation games are developed in Chapter 13 in conjunction with the coordination of work teams.

A cooperative game is structured in such a way as to provide maximum returns to both players if certain joint plays are made. In those cases information sharing is strategic, and binding agreements are possible (Zagare, 1984). The Prisoner's Dilemma game has become particularly dominant in the game literature because of its capability of producing both competitive and cooperative results. It is thus regarded as a mixed-strategy game.

Simon

Rather than pursue a trajectory on the varieties of games here, it would be more germane to refocus on what game theory and related pursuits taught us about human–computer interaction. The first point is that interacting with a suitable computer program would have considerable benefits that would not be obtained by acting without one, all other things being equal. Simon (1957) introduced the concept of *bounded rationality* to express this disadvantage: Although people are capable of rational thought and select the best options for their own interests when faced with a decision, there is a limit to the amount of information that they can process well. Too much information at once or flowing too fast compromises the human's ability to produce optimal decisions. Contemporary cognitive concepts such as channel capacity have built on the basic principle of bounded rationality.

Simon (1957) also introduced a related concept of *satisficing* in decision making: It may not be possible to identify and explore all possible options that are relevant to a decision in time to make the decision. One must examine what is available and one's best shot.

Simon and numerous coworkers soon developed a theory and action plan that involved building computer simulations for complex economic events and organizational behavior (March, 2002). Although the substance of the economics or organizational behavior falls outside the scope of this book, there are human factors issues that can be extracted. First, computer simulations have the potential to augment human computational capacity.

Industrial systems simulations involve the modeling of a combination of technical and human contributions (e.g., Forrester, 1961). A simulation of a human–technical system, however, will produce useful results to the extent that a human factors knowledge base has been incorporated into the program. This caveat leads to a more general problem of how to extract knowledge, code it into a program, and validate the program's output. The topic thus turns to expert system architecture.

Expert System Architecture

An expert system is an interactive computer program that functions like a human expert on a particular set of problems. Expert systems might be built to diagnose failures in a computer system or automobile, select wines to accompany various menus, determine preferable courses of medical or psychological treatment, predict outcomes of military or economic scenarios, or help develop recipes for commercial food products (Binik, Servan-Schreiber, Freiwald, & Hall, 1988; Fox, 1984; Klahr & Waterman, 1986). In addition to diagnosis, these expert systems often suggest remedies or series of tests to be performed in order to continue the query to a final solution.

Expert systems consist of three main components: the database, the inference engine, and the interface. The database consists of conventional spreadsheets of numbers (e.g., means, standard deviations) and facts and interrelationships such as "Condition *A* is correlated *r* with event *X*." The inference engine is the set of logical rules that the human would use to solve a problem. The interface allows the operator to query the expert system and to obtain answers (Denning, 1986). In the more sophisticated interfaces, a natural language processor that interprets the meaning of the question rather than only the array of specific words bolsters the dialoging program, which facilitates the query. Thus, persons may ask questions in their own language style without having to remember exact commands in order to operate the program.

Algorithmic Systems

Inference engines can be sequenced on an evolutionary scale from relatively dumb to relatively intelligent. At the most elementary end of the scale, the rules of inference are simply algorithms. An algorithm is a specific set of computations and logical rules that are always performed in the same order. They can produce limited forms of output and respond to limited arrays of questions. Nonetheless, algorithmic systems can accomplish a lot, especially where they model relatively repetitive human decision processes that have only a few if any variations of procedure.

One illustrative use of an algorithmic expert system is the computer-based psychological test interpretation (CBTI). Algorithm-based CBTIs require five basic components: (a) the reading of responses to questions, (b) the compilation of test scale scores, (c) the use of prediction equations for additional scale scores, (d) the library of interpretive statements, and (e) the invocation of rules to trigger appropriate interpretive statements. The algorithm set would include equations drawn from the publisher's database but also could contain equations supplied by the system user. Many useful algorithms have been derived through statistical bootstrapping of clinicians' interpretations of multiple pieces of data (Carroll, 1987; Meehl, 1956). The most basic interpretive rules would divide a psychological measurement into high versus low, or high, midrange, and low score ranges and trigger one interpretation for each grouping of scores. Another basic type of rule triggers statements based on configurations of high, midrange, or low scores for two or more variables by using "and/or" gates. The algorithmic interpretive program would conclude with print commands to format the interpretive statements chosen by the rule system.

Many of the rules for CBTIs and other decision schemata are produced by statistical regression. This approach to rule-making is called *policy-capturing regression.* We take a set of decisions made by the humans and create an equation (usually linear) that correlates the preconditions with the decision outcomes. Meehl's (1956) important discovery was that once an equation had been identified in this fashion, the computer's continued use of the equation for making decisions produced more accurate results than the human decision makers. This is particularly interesting because the humans initially generated the entire idea of what the decision should be and how it should be made.

Rule-Based Systems

The second stage of inference engine evolution is the rule-based system. A rule-based system is distinguished from an algorithmic system by the use of metarules that govern the execution of other rules. Rules and materials may be deductive, inductive, prohibitive, permissive, obligatory, causal, or probabilistic. Some rule systems allow both forward and backward searches of the logical steps. The following discussion of rule spaces, chaining strategies, and classifications elaborate these points.

A rule space is a set of rules that is structured by a set of logical interrelationships. The rule space would also be characterized by various attributes of rules such as a rule's rigidity versus flexibility, its specificity or generality, and its precision or vagueness (Sridharan, 1985). Flexibility would be governed by metarules, often of the branching type. A rule is more flexible to the extent that its execution depends on a greater number of metarule arguments holding true.

Precision or vagueness would affect the order of rule execution, where precision usually takes priority over vagueness. A precise rule would take the form: "If test score $F < Y_2$ and $F > Y_1$, print statement ~." A vague rule would take the form: "If 6 or more of the preceding 10 rules are false, go to next rule." The former rule specifies exact conditions for using S_1. The latter is defining a default condition that should be executed when any of several possible conditions exist. The vagueness of the vague rule example would be increased to the extent the "preceding 10 rules" are also vague.

Specificity or generality also governs the execution of rules. Here, order is additionally determined by whether the nature of the decision problem requires a depth-first, breadth-first, forward-chaining, or backward-chaining strategy. Chaining strategies involve the specificity–generality concept.

Chaining Strategies

Figure 12.1 describes a set of rules, such that a square or circular node in the diagram represents each rule. The rules are organized at three levels of specificity. A depth-first forward-chaining rule order begins with a general rule, then moves forward toward a specific rule. A depth-first backward-chaining rule order begins with a specific rule or rules, then moves toward a general principle. For breadth-first strategies, the decision strategy begins with an execution of rules at a given level of specificity, then moves toward generality or specificity depending on the nature of the query. In principle, the program is equipped with metarules to execute the basic rules in either a forward, backward, depth-first, or breadth-first direction. The choice of direction would be dependent on the type of question being posed to the expert system.

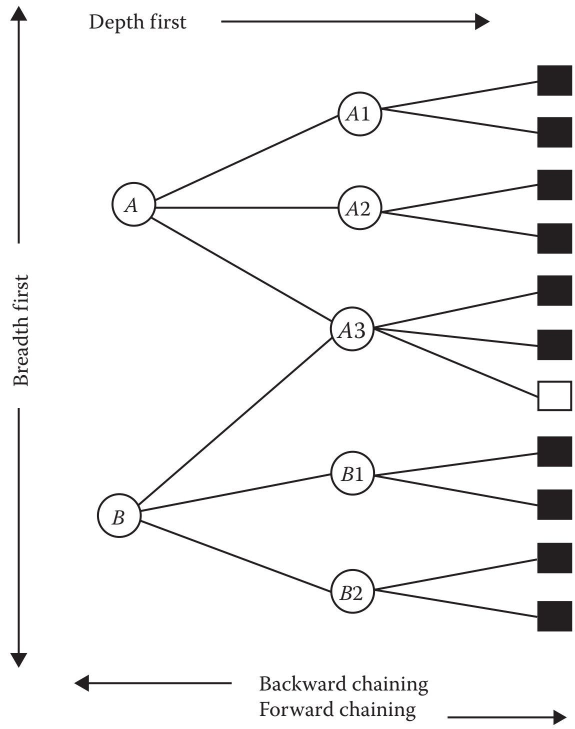

FIGURE 12.1
Forward, backward, depth-first, and breadth-first frame search strategies.

The squares in Figure 12.1 represent objects within a classification structure. Note that an *object* can be a physical object that is known to the program as a value of a variable, such as "Occupational Group = 7." A physical object can be known to the system as a vector of parameter values, in which case it is also a *frame*; the relevance of frames is addressed in a subsequent section of this article. In the information sense, however, an object can be a single value or a frame as just described, but it can also be a block of rules (or program statements) or a graphic that appears on the screen. Informational objects can be moved around the program and utilized multiple ways.

The object squares in Figure 12.1 are all drawn in black except for one to denote a rule search in progress. The white object might light up in a depth-forward search that is executed in response to the query, "Find an object that is a member of *A′* and *B′* sets" (more than one object in Figure 12.1 would satisfy this requirement). Alternatively, the white object may be the subject of a depth-backward search such as, "What are the origins of object *C*?" The white object could also light up in response to a breadth-first query, "Find the object with the most ambiguous classification certainty value."

The normative leadership model (Vroom & Jago, 1988) serves as an illustrative example of a depth-first forward-chaining rule structure. In the model, the manager proceeds through a series of yes–no decisions regarding how to delegate a job decision; the manager's decision rules are ordered in priority. The pattern of *yes* and *no* responses results in the most appropriate delegation strategy for the set of facts surrounding the decision and the workgroup. A backward-chaining order of the same rules begins with the type of delegation strategy that the manager chose, and then proceeds to reconstruct the pattern of yes–no decisions that could have led to a specific delegation strategy.

Classification Structures

Classification structures represent another category of the rule-ordering formats. An interpretive strategy for a test may require that a pattern of test scores define group membership, where the groups are hierarchically organized. Figure 12.1 could also describe an arrangement of 10 objects that are organized into five categories, which are, in turn, organized into two broader categories. One might develop a rule system that classifies a case, C, according to a specific-to-general search strategy, a general-to-specific search strategy, or a commonality strategy. Commonality search strategies seek to determine what C has in common with the categorical classifications in the program; it is essentially a breadth-first strategy. In any classification structure, it is necessary to set a criterion for deciding whether two items are the same or different, or for whether an example is truly a member of a class. For some problems, such as the classification of biological specimens into species, the rules of membership are clear. For psychological or behavioral problems, however, the boundaries of classification are often fuzzy or probabilistic (cf. Zadeh, 1981). Hence, decision makers use hedge phrases such as "it is likely that" or "the certainty level is." A database coupled with rules derived from discriminant analysis, regression, or cluster analysis can increase the precision of some types of classification rules.

A *Horn clause* is a rule that defines a negative instance of group membership, for example, "If C scores high on the anxiety scale C is not a psychopathic deviate." Horn clauses can cause complications with rule structures because they often introduce contradictions. Sharpening the precision and ordering of existing rules within a rule space (Buntine & Stirling, 1988) can minimize complications. An individual rule is overly general if its application causes the selection of a negative instance of a prototype (Buntine, 1987). With psychological data, this type of overgenerality will be a matter of degree, and would depend on the base rates of category membership.

Interface Requirements

Up to this point, no distinction has been made between interactive and noninteractive expert system programs. The distinction is important because rule-based expert systems are equipped with interfaces by which the user may query the database and invoke generic tasks, such as to provide interpretations of scores for various purposes, to classify, or to make recommendations. The need for an interactive mode is especially relevant where the decision goals require a substantial amount of customization or situation specificity from one decision event to another. The interface should allow the user to expand the database, modify rules, or reflect different decision priorities.

Frame-Based Systems

The third stage in the evolution of inference engines is the frame-based system. There are two uses of the word *frame.* A frame can refer to a set of parameters in the database that describe an object in the system. Each parameter in the set would constitute a meaningful piece of information that could be used to place an object into one category or another. A frame can also refer to a prototype in a classification structure; it could reside in the database or in the rule structure. In both usages, a frame is analogous to a chunk of data in human skilled memory (Linde, 1986). For instance, one might write rules that invoke separate characteristics of a prototype or a possible example of a prototype, or one might chunk all characteristics of an example or a prototype into an intact object. That intact object can be manipulated within the program. Some frames represent prototypes (abstract principles) in the system, whereas others represent commonly used deviations from prototypes (examples). Each frame has a number of properties that may be modified or canceled to produce a new and more appropriate result. When frames are added or modified, the rule system is structured to utilize the newly acquired information.

That brings us to a riddle: "What animal is large and weighing over a ton, gray in color, with big floppy ears, a long nose that drags the ground, likes to eat hay and peanuts, and lives in the trees?" Answer: An elephant—I lied about the trees (Brachman, 1985). The conceptual problem is to determine when we can add characteristics to a prototype or deviate from the prototype and still have an example of the prototype.

The use of rules for classification purposes constitutes a rudimentary frame-based system. Frame-based systems build on rule-based system design. Once the profile of scores has been compared with a library of example prototype profiles and the best match has been selected, the program would do more than simply generate a generic interpretation. It could also determine deviations from the prototype and retain the new data for purposes of self-modification of the program's rule structure. Frame-based systems are prone to a frailty in that there is a fuzzy cutoff point beyond which the prototype-deviant profile of scores no longer resembles the prototype frame (Brachman, 1985). The deviant profile might then resemble another prototype or no known prototype at all. Consequently, a frame-based system should contain rules that impose limits on prototype deviation for identifying true prototype matches.

A frame-based system would not only contain a large stock of frames but also would allow users to add or modify frames. When constructing frames, it is important that the rules used to define frames be relevant to the interpretation objectives in question (Linde, 1986). If an interpretive strategy is based on causation data, more frames are required to the extent that the causation mechanisms are not precisely known (Pazzani, 1987b).

In addition to classifying profiles against prototypes, frame-based systems have other potentialities. First, they could describe the probability of a case belonging to a prototype represented by a frame by drawing on knowledge generated through statistical procedures such as discriminant analysis or cluster analysis. Second, they could retain case data for possible building of new frames in a recursive system. Third, they could prescribe a series of operations or activities required to transform an example into a frame prototype by means of a backward-chaining rule complex. Fourth, they could predict or explain new behaviors from preexisting knowledge and rules. The interface for a frame-based system should also allow the following interactions: (a) to enter new rules, (b) to select parameters that govern the execution of various rule groups, (c) to compose a frame, (d) to recall and modify a frame, (e) to compare frames, and (f) to execute the sorting of new cases against prototype frames.

Example Spaces

An *example space*, which is a collection of prototype frames in a frame-based system, needs to be supported by several types of rule groups. These would include rules to assess sameness or difference, rules that set criteria for fuzzy classifications, rules that allow for transformations of prototypes, and rules that set limits on allowable deviations from prototypes.

Another type of rule in frame-based systems is the *open-textured* rule. Unlike other types of rules for frames, where a prototype is defined and examples subsequently provided, open-textured rules define a concept in terms of a set of examples. Prototypes are created by a partial backward search. A new item is accepted as a member of a set if it is similar to any one of the existing members of the set. Open-textured rules are typically fraught with hedge phrases such as "due care," "reasonable length of time," and so on (Sridharan, 1985). Open-textured rules often appear in legal decision making and allow precedent decisions to impact on subsequent decisions. The relationship between precedents and decisions is an example of a recursive process.

Recursive Systems

Automatic revision of the rule set is intrinsic to the fourth stage of evolution (Brachman, 1985; Sridharan, 1985; Thomas & Gould, 1975), the *recursive frame-based system*, whereby the expert system can learn through experience by collecting new frames, and modifying its rule system. The principle of recursion lies at the heart of a self-teaching expert system. A recursive program is one that can call on itself as a subroutine and execute a routine that changes its own programming.

A recursive system has the capability of updating and revising its database and rule structure. In a recursive frame system, a rule is conceptualized as if it were a frame. Rules have properties such as the contents of an *if* clause, content of a *then* clause, the presence and number of and/or gates, the elements of any computations that are involved, the confidence associated with a particular rule, whether a rule is a simple rule or a metarule, and so forth. These frame-based rules would then be read and manipulated by a new set of recursion rules. A schematic of a recursive frame system appears in Figure 12.2. As mentioned earlier, recursive systems capitalize on the conceptual breakthrough in AI that there is no fundamental difference between a program and data (Turkle, 1984).

The left side of Figure 12.2 describes the nonrecursive frame-based system discussed up to this point. Program Sections 4 and 5 on the right side of the diagram describe the new elements. There, algorithmic rules, metarules, and frames for prototypes are read as data. The recursive inference engine then executes rules for rule modification in Step A.

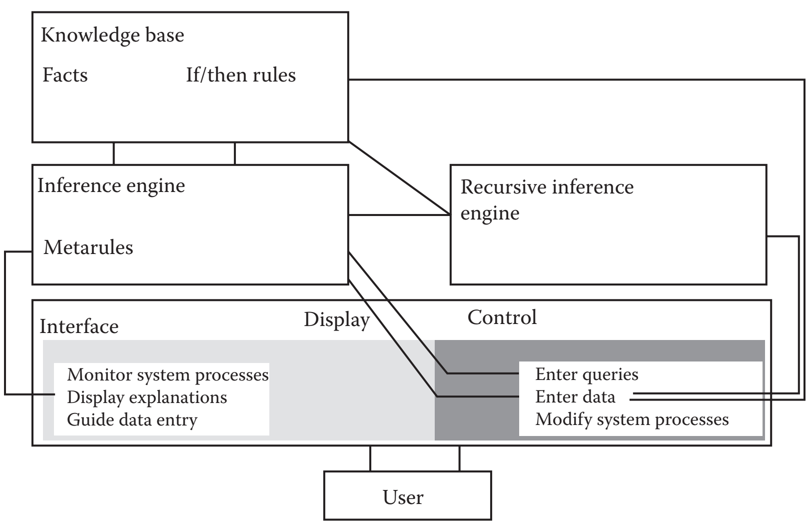

FIGURE 12.2
Expert system architecture with recursion component.

Next, the core program in the left section of the diagram is called a subroutine or symbolic object (Step B). More rules are then executed whereby the new rules are integrated into the original program (Step C). There should be a check for logical inconsistencies at Step C, and flags to the user at Step D to indicate where new library statements are needed. Learning can take place in a rule-based system. The system would contain rules to retain new data and to modify the knowledge base accordingly. A rule-based system with such a capability is structurally more advanced than ones without it, but not as advanced as a recursive frame system. In the latter, both the knowledge base and rule structure can be modified.

Interface Requirements

A new level of interaction occurs in the interface between the recursion rules and the user, in terms of control functions and displays needed to operate the system. The recursion rules would emulate the strategies the user would invoke to modify the program manually. The user may desire the option to invoke or not to invoke the recursion subsystem, or to enter and to modify any criteria for program modification.

Smart Integrated Displays

The smart displays encountered in Chapter 4 originated with the principles of AI. Arguably, the original concept of an expert system has infused so many types of software programs that people seldom recognize a firm difference between expert systems and programs that are not explicitly designated as such. The smart display as we (should) experience it today builds on lessons learned concerning principles of visual perception, recognition primed decision making, ecological psychology, and the fact that we *interact* with displays more so than ever before.

According to Bennett and Flach (2011), our dyadic model of the PMI (Figure 1.1) should be modified into a triadic arrangement as depicted in Figure 12.3. The new element is the system of meanings associated with the task environment. The operator is not a passive viewer, but is instead trying to accomplish a task and is looking for salient visual cues, or affordances, that signify opportunities to take one sort of action or another. Importantly, actions are limited by operators' mental models of possible situations; one needs to understand what one is looking at.

Situations change over time, a point that is addressed by NDS theory specifically. Bennett and Flach argue for including recognition of system states as part of the knowledge base that should in turn be built into smart displays for a better representation of the

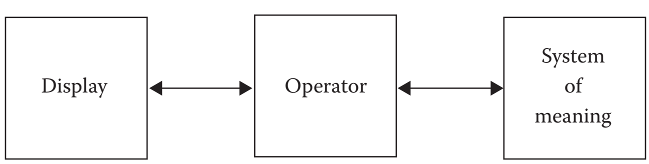

FIGURE 12.3
Triadic person–machine system.

situation in situation awareness. The NDS connection leads to five implications for display design: (a) How many system states are meaningful, and what are they? (b) What variables affect change in the situation? These would be control parameters of some sort. (c) How do control parameters affect the change in the system state? (d) What are the critical values of the control parameters that are associated with the change in situation? (e) What are the boundary conditions that induce limits to problem solving or situation awareness, or might contribute to a reorganization of the situation? In more colloquial terms, when do we have a major game change when the rules of engagement change distinctively?

The systems of meaning that make the PMS triadic are manifest in three levels of abstraction. The most abstract level ("highest" in the jargon of information technology) involves computation of the mental processes that are part of problem solving. AI or expert systems are essentially *computing* mental processes. The second level is *representation* of the results of the computation in a form that conveys maximum information to the viewer. The third level of abstraction occurs at the hardware level, which is deliberately out of the end user's view, while the user focuses on what is being represented.

For an illustrative example, reconsider the discussion question from Chapter 3: The Air Force wants to know whether a new pilot–cockpit interface will help fighter pilots destroy enemy aircraft in air-to-air combat situations. How can signal detection concepts be used to assess pilot performance (in a simulator)? The first step is to frame the performance question as a signal detection problem. Five dimensions are involved: three for space and two for movement. The target aircraft in the dogfight not only occupies a 3-D space while standing still but can move north–south, east–west, and vertically. The predator aircraft can also move in the same space. The goal is move so that the target falls within the firing envelope of the predator's guns. In days gone by, flight motions were calculated by the pilot on the fly (pardon the pun). An intelligent aircraft, however, would be equipped with sensors to detect the location of the target, convey the location information to the predator pilot in a form that immediately results in a strategic movement. When the target is computed to be within the envelope it would represent this status in the form of an audio signal and a flashing light. The predator pilot could then push a button and—boom.

The triadic principle of display design generated the concept of the integrated display (Bennett & Flach, 2011). If there are several pieces of information that contribute to the meaning of situation, the gestalt laws of perceptual grouping would be operational: If the whole were greater than the sum of the parts, the goal is to design a display that would convey the emergent condition in a holistic fashion in one integrated graphic. Figure 12.4

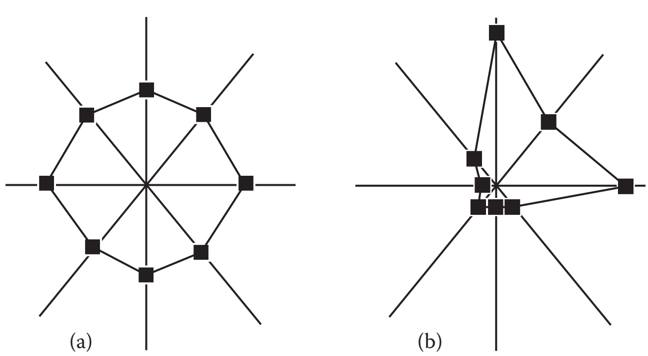

FIGURE 12.4
Eight-parameter integrated display.

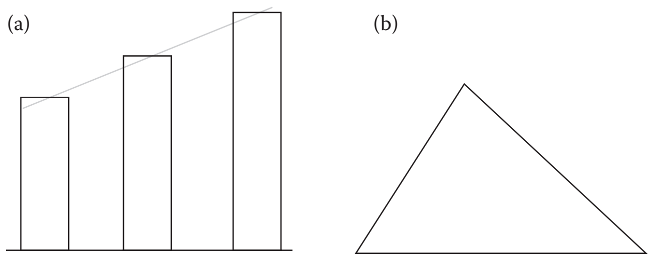

FIGURE 12.5
Three-parameter integrated display.

is an example of an integrated display. The system has eight hypothetical parameters that contribute to a distinct system state. If the values were arranged in a regular hexagon, the state would be interpreted as "normal." If the values were configured as shown in 12.4b, however, the system could be deviant in some way.

Bennett and Flach gave examples of other types of integrated displays that have been tried since the mid-1980s. One experimental design involved using a line drawing of a face where the four parameters of information were mapped to the eyebrows, eyes, nose, and mouth. A comparison of user responses to different face display arrangements and alternative displays of the same information showed that the face display could be the best or the worst option depending on which piece of information was mapped to which part of the face.

Another instructive example involved three parameters that denoted three interrelated aspects of a system. One design (Figure 12.5a) was three bars, and the alternative was a triangle (Figure 12.5b). The critical configuration for a system state looked like "a" with an imaginary line connecting the top of the bars, or like "b" with a right angle at the top vertex of the triangle. The representation with the bars was more readily interpreted by the viewers than the triangle. Both representations invoked the principle of integrated displays, but one was much more usable than the other. Once again, usability studies were very necessary for choosing an optimal display design.

In summary, the primary asset of the spatial displays is that they generate a holistic notion of the system's status that is more salient than the elementary sources of data contributing to the display. Thus, the meaning of the display qualifies as "emergent" (Bennett & Flach, 2013). Their primary limitation is essentially the same: The possible gestalts from the display need to map onto real-world events and thus convey meaning to the operator. Failure to map in a useful manner only permits the display to generate pointless information (or noise) that obscures the interpretation of the elements that are important.

Large-Scale Integrated Databases

What Is Possible?

A hypothetical architecture for a large-scale integrated system appears in Figure 12.6 from the U.S. Department of Defense (2002). The draft includes many possible data sources that can be integrated and cross-linked to identify international terrorist threat risks. Note the combination of pictorial information (e.g., faces, fingerprints), movement data (gait), and

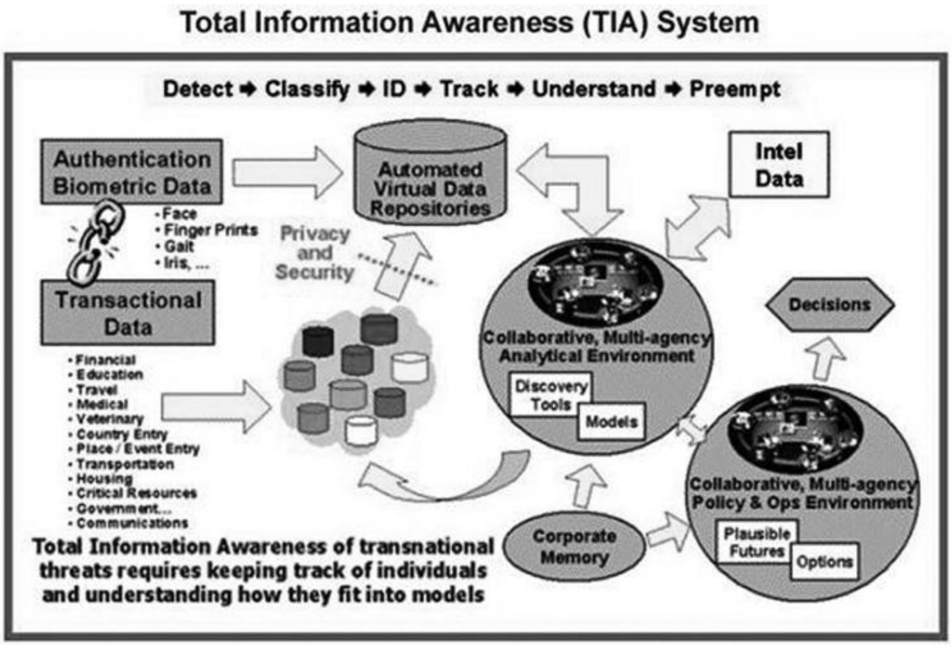

FIGURE 12.6
Overview of a suggested national security information and surveillance system. (From U.S. Department of Defense, DARPA's Total Information Awareness Program Homepage, http://www.darpa.mil/iao/TIASystems.htm, 2002, in the public domain.)

links with other government and private agencies with which the potential terrorist might have interacted. A system like this one would benefit from contemporary computer languages that translate other human and computer languages and information modalities into other modalities. The web page with the Total Information Awareness system is no longer available, and one can only guess why. The model, nonetheless, stands as an illustrative example of what someone could imagine for a massive database.

The arrows and connector graphics indicate information flows among the major components of the system. The little drums to the right of the list of databases shown in small type are icons for a database that date back to the pre-GUI days. It is not clear what the crossbar labeled "privacy and security" really means, but in some semantic networks it could mean that a privacy protection barrier exists and somebody needs to get an "arrow" through it. Similarly, the chain with the broken link does not convey anything specific, but it could mean that if a person could not be identified through physical recognition, then it would not be possible to link to their other data sources.

The question about "privacy and security" leads to another interesting issue concerning medical information that appeared in the list of databases in Figure 12.6. The Total Information Awareness system would require at minimum a massive digitalization of all medical records. There is some impetus for digitizing them already as part of the impending new national health-care system in the United States. In theory, if all records were digitized in a central database, then physicians who were working on a patient could see the results of tests ordered by other physicians and not duplicate efforts by ordering another test. In practice, however, not all tests are ordered to produce the same information for every patient, and the life expectancy of test results could be very short. The efficacy of digitization to eliminate redundancy is thus not as great as it might appear at first blush. Furthermore, the access to patients' medical records is controlled by federal regulations and limited only to the patient, the patient's designated agents (often family members), and the service providers specified by the patient. One political scientist observed, "In today's mostly paper-based health records system, privacy is protected largely by fragmentation, inefficiency, illegibility, and general chaos" (Rothstein, 2007, p. 487). Digitization could be the first arrow to leak through the dotted line.

The interface for the Total Information Awareness System would be very complicated if it were to allow humans to enter a variety of information databases and monitor their coordination. The operation of the inference engine would require additional points of human entry. It is likely that the maintenance of the database would benefit from some automatic devices.

The two pictorial bubbles in Figure 12.6 signify the location of the inference engine, but the substance of the inferences was actually specified at the time the drawing was produced was and is still to be determined. Data mining techniques would be the likely tool for getting it started.

Data Mining

Scientific research involves starting with a research question, organizing a database that would permit answering the question, and performing the pertinent statistical analyses needed to draw the appropriate inferences. One common variation is to build a large database that contains more data fields than necessary to answer a limited set of questions, but is available for later use to answer other questions that might be investigated later on. Another variation is to start with a question and look for a database that could be "out there" that could be used to answer the question. In the first scenario, there is a close alignment between the question and the data that are supposed to generate the answer. The second case is an answer that is looking for a question. There is a third case however, which is a question that is looking for an answer that is looking for a question. In the latter case, the researchers with the questions need to ponder whether the database is constructed closely enough to the way they would have done if they started from scratch. Are the constructs what they are looking for? Are the constructs measured in such a way that does not compromise the interpretation of what the analyses are supposed to mean and what inferences they are supposed to deliver?

In the expert systems discussed to this point, the data could consist of the primary (or raw) observations such as those the research scientist would analyze, or it would contain facts that have been distilled from a primary source into IF/THEN statements. *Data mining*, however, involves applying statistical analyses to large primary databases without a great deal of regard to any a priori research questions or choices of inferential statistics for answering those questions (Hancock, 2012). Mining programs read the fields of a database and execute every statistical analysis they know how to do. Programs differ in the logic they deploy for analyzing data, the types of data for which they are designed to be most pertinent, and the format by which they serve up the results to the human end user. Although their broad-based analytic strategies can be limited or redirected by the human end user, the goal is to set them into automatic operation to produce anything they can find: emerging categories of objects or events, predictive relationships between one thing and another, trends over time and forecasts, and so on.

Data mining has become increasingly common now that huge databases are being generated every day by computer-based systems such as social networking sites, medical record-keeping in hospitals and clinics (Epstein, 2010), travel arrangements, every commercial site where people buy products online (Li, Ogihara, & Tzanetakis, 2012; Ohsawa & Yada, 2009), finance and securities transactions (Kovalerchuk & Vitysav, 2000), and web search results from people who search the web for information. Big Science projects such as the Human Genome Project that commenced at the end of the past century to map the entirety of human genes and chromosomes, got what it wished for: voluminous data that needed some automatic processing for the human investigators to figure out what any of it means for medical practice (Ao, 2008). Furthermore, these databases are updated on a continuous basis, thanks to all the individuals in society who visit physicians, web sites, use online banking facilities, and enjoy the convenience of seeing their friends and posting pictures of their lives on public systems like Facebook instead of remembering and storing their friends' contact information on their own computers or other means of private control.

Although these databases exist for one explicit purpose, nothing prevents their use by entities mining data to extract patterns and produce what are called "models" in Figure 12.6. There is also the cat-and-mouse game of Internet security and malware—viruses, Trojan horses, phishing expeditions, and so forth—that are designed to disrupt the flow of legitimate transactions across web sites. Thus, there are data mining strategies designed to identify the prototypes and flow of malware (Agguwal & Yu, 2008; Masud, Khan, & Thuraisingham, 2012). Data mining programming continues to evolve to answer broader ranges of questions (or make them up) through broader ranges of analytic techniques that fall outside the scope of this book. The more automaticity that is permitted to the programs, however, the greater the potential for surprise to the operator, which is good if novelty is in short supply, and the greater the potential for meaninglessness if some filters are not applied by the human or the machine.

The program F-Finance, however, offers an interesting combination of classic expert system architecture and data mining. The program allows financial traders to insert their favorite algorithms for picking or selling securities into the inference engine (Cao, Wang, & Zhang, 2004; Moemeng, Cao, & Zhang, 2008). The program launches a data collection, applies user-defined algorithms, and reports back on the accuracy or productivity of the algorithms.

In the interim, it is fair to state that the main contribution of data mining to the concept of artificial intelligence is in the production of data and material for inference engines. There has been little in the way of unique advances for controls and displays or related aspects of user interaction, possibly because the work culture surrounding these techniques is still confined to information technology professionals. Nonetheless, there is some potential for new developments:

> [H]uman–computer interaction issues in [distributed data mining] offers some unique challenges. It requires system-level support for group interaction, collaborative problem solving, development of alternative interfaces (particularly for mobile devices), and dealing with security issues (Moemeng, Gorodetsky, Zuo, Yang, & Zhang, 2009, p. 56).

Artificial Life

Artificial life denotes a range of computational techniques that emulate an aspect of living systems. Examples include neural networks, autonomous agents, cellular automata, genetic algorithms and evolutionary computations, and agent-based modeling more generally. The former two items are considered in this chapter. The other items on complex systems are discussed in Chapter 13.

Neural Networks

Whereas the expert system attempts to emulate complex cognitive processes, the neural network is better suited to visual pattern learning and motor response (Chandrasekaran & Goel, 1988) or simple psychophysical judgment tasks (Vickers & Lee, 1998). The neural network programming concepts were initially intended to emulate actual psychological processes. Several possible architectures have been developed over the years, but most have shown to be equivalent with respect to performance accuracy (Greenwood, 1991; Vickers & Lee, 2000). The parallel adaptive generalized accumulator network (Vickers & Lee, 2000), however, now appears to show advantages in response–time modeling, however. Despite the goal of developing programming models that reflect true neuronal behavior literally, the *binding problem* in neuroscience, which is the set of rules by which the firing patterns of neurons correspond to what we experience as a thought or information, has not been solved. See Minelli (2009) for a review of the current state of the science on real neural behavior.

A neural network attempts to emulate four basic actions of self-regulating and adaptive systems: discrimination, recognition of identity, adaptation, and memory (Vickers & Lee, 1998). This system distinguishes one stimulus pattern from another. It must recognize an incoming stimulus as a match for one that is already known. It must produce suitable responses to incoming stimuli as well as change its responses to changes in the stimuli. Memory is needed for all the foregoing functions.

The neural network architecture requires a sensory subsystem and an effector subsystem. Accumulator functions, which underlie both subsystems, are program units that keep track of the number of times a stimulus has been encountered and associated with a response. Accumulator functions represent the learning curves that are typically found in human learning experiments. They may be statistically represented as cumulative normal distributions in some neural network systems, whereas other systems may define learning curves as deterministic functions that resemble the formation of an attractor in the nonlinear dynamics sense (Bar-Yam, 1997; Vickers & Lee, 1998).

Real neuronal firing patterns differ for novel stimuli compared with familiar stimuli. Once a stable recognition-and-response mechanism has been formed, it can then remain stable or shift in response to new stimuli during its virtual life span. Neural network programs thus undergo a training process whereby numerous stimuli-and-response combinations are presented and the program accumulates an association between them. Training may be unsupervised or supervised. In the latter case the human operator is trying to make the program learn a specific connection. In unsupervised learning, there could be some surprises (Greenwood, 1991).

Virtual associative memory functions allow the program to recognize an incoming pattern as an example of a prototype, compare prototypes, and automatically distinguish new prototypes. For these memory processes, each stimulus-and-response pattern is defined as a group of informational elements. A set of cognitive rules is needed within the program to allow the comparisons of stimuli and the execution of modified actions (Vickers & Lee, 2000).

Within the memory banks, the informational elements act like little hooks whereby elementary units of similarity are allowed to accumulate; the memory is now making new associations. Additional rule groups are needed to keep track of the new associations and execute modifications to the stimulus–response patterns. If one were to activate a particular prototype (frame) one could readily call up adjacent frames that had some partial association with the target frame. To prevent possible confusion in the system, the system would need to be equipped with both activation and inhibition mechanisms, like real neurons, to permit the desired extent of activation spread (Kohonen, 1989).

Vickers and Lee (2000) identified a speed–accuracy trade-off in neural networks that bears some resemblance to the speed–accuracy trade-off in human psychomotor processes. They assessed the accuracy of a neural network program for a simple psychophysical judgment as a function of the number of learning exposures. Their experiment also contained a manipulation that required the program to deliver judgments with high or low confidence. Low-confidence judgments actually produced higher mean proportions of correct responses compared with high-confidence judgments. The speed–accuracy function for the low-confidence judgments showed a sharper *S* contour than the function for high-confidence judgments.

Bar-Yam (1997) and Sprott (2013) identified a phenomenon of neural networks that bears some resemblance to stochastic resonance in human psychophysical judgment. It appears that the presence of some noise in the stimuli allows the neural network to make adaptive responses more readily than what would occur if noise-free training stimuli were used. According to Bar-Yam, the absence of noise allows the program to lock into a stable stimulus–response attractor quickly and inhibit the program from leaving the attractor. If noise is introduced during training, the primary stimulus–response attractor may take more trials to acquire, but the program will have also learned to respond to variations in the stimuli.

Autonomous Agents

An *autonomous agent* is a piece of programming that recognizes target stimuli and effects actions when the stimuli are encountered. What makes them autonomous is that they can be loaded into a program system and allowed to do their jobs without direct control of the humans. One might then send an agent out to a group of web sites, look for targets, and report something back to the human who initially sent the agent. The agents are sometimes known as robots.

There are different levels of complexity for the design of autonomous agents. The simplest level would be written as part of a software upgrade package. The agent would look for files with particular names, determine if those files need to be changed, and execute the replacement of one file with another.

The second-level agent can do some web surfing on its own. Web search engines send agents out regularly to patrol web sites regularly with the objective of making a response to a search inquiry that is as complete as possible. The agents might fetch information only when explicitly asked to do so, or they might be sent out on a daily basis to update the master system (and human user) on topics that were requested on a continuing basis; RSS feeds work on this principle. Agents inevitably interface with other users' systems, which they can only do if their presence is unnoticed or recognized as benign. Otherwise, a user either has a security problem or is creating one.

The F-Finance program (Moemeng et al., 2009) mentioned earlier uses agents to gather data from real-world trading sources for testing investment algorithms. Program trading itself has been around since the mid-1980s: Agents look for situations that match a trading target and execute purchases and sales. The agents can create a problem for each other however. The stock market crash of 1987 resulted from trading programs reading each other's trades; a pattern of sales was detected, then the drops in prices following the sales triggered more sales at a loss. There were no fundamental problems with the securities themselves or the market conditions as a whole. The Securities and Exchange Commission responded by imposing an agent of its own: If the Dow-Jones index drops more than 100 points in an hour, all trading stops to give the automated and real-time investors a chance to reset their parameters.

Figure 12.6 makes reference to several databases involving pictures, fingerprints, medical services, animal health, and travel patterns. Each type of database is formatted differently and limited in scope. The process of extracting and collating data from disparate sources can be accomplished by using multiple autonomous agents that mine the individual databases and converse with each other to produce results by combining and integrating their findings (Cao, 2009). The agents work in a *swarm* configuration (see Chapter 13) where agent-to-agent controls and displays keep the agents' information collection coordinated.

Swarming agents may be the means for getting around the dotted line of privacy and security in Figure 12.6. Users supply medical information to physicians because it is essential for the medical service and different information that is essential to travel services to airlines. Both the airline and the clinic state they guarantee privacy over the records so that anyone who is not a legitimate part of the service provision does not have access, and they will not sell the information to outside parties. Marketing and sales operations are one large source of demand for information. An autonomous agent, however, can get into each database and get out quietly, then collate all the information from or for users that do not have a legitimate access to the information. The agents would need to be recognized as friendly.

A similar problem arose in conjunction with the social networking site, Facebook, which reportedly has nearly a billion users worldwide, most of whom upload pictures of themselves and other people to share with their "friends." Facebook, until recently, put a tag on each picture that was uploaded so that it would be possible to deploy face recognition software to every picture to catalog who was in it. According to Sengupta and O'Brien (2012), Facebook received a lot of pressure from investors to sell picture data to increase profits, and from regulators in European telecommunications agencies to stop tagging pictures altogether because of privacy laws there. This type of software for mining and collating data had been in use outside of the United States and the European Union for monitoring political gatherings and possibly preempting protests and insurrections (Fuchs, 2012).

Validation Issues

Having described expert system architecture, it is now possible to assess the probable sources of validity and invalidity. Lanyon (1987) identified three levels of validity assessment. The first is the statement level of analysis whereby each statement in a report is evaluated for correctness. The second is the narrative level of analysis where the validity of the global impression of a person created by the set of interpretive statements is assessed. The third level pertains to the correctness of decisions made by the expert system as interpreted by the end user of the system. Several types of validity are elaborated here. Some validity issues pertain to the knowledge base, and others pertain to rule structures.

Knowledge Base Validity

Authorities on expert systems appear to agree that the cornerstone of validity lies in the extraction of knowledge from subject matter experts (Culbert, Riley, & Savely, 1989; Gruber & Cohen, 1987; Pazzani, 1987a; Vale, Keller, & Bentz, 1987). Even if subject matter experts agree, the information they provide to establish the knowledge base may not be optimal. Research conducted on "illusory correlations" (Chapman & Chapman, 1967, 1969) suggests that there are often systematic errors in experts' (and nonexperts') judgments of the relationships between two events. These phantom relationships might not exist at all, might exist to a lesser degree than was thought, or might be in the opposite direction than what was thought.

In other words, even experts' opinions are frequently based on stereotypes, superstitions, or assumptions that have not been validated. For instance, in a study of eminent clinical psychologists it was found that they performed only slightly above chance (Little & Schneidman, 1959) in their diagnoses of clinical disorders. Moreover, an extensive body of research on clinical versus actuarial prediction has demonstrated that, for many applications, actuarial predictions are superior to clinical predictions (Goldberg, 1968, 1969; Meehl, 1954).

System planners need to work closely with subject matter experts to identify the information that should be included in the expert system, to determine the level of specificity of that information, to ascertain the structure of the information, and to collect the specific facts that need to be included. In the case of psychological knowledge, experts can and do rely on published research for CBTI content, but a substantial amount of psychological research is not in a form amenable to CBTI rule building (Lanyon, 1987). Although the validity of the psychological concepts may have been established in carefully controlled or constricted situations, the validity of new generalizations made by CBTI system developers may be questionable and are often untested (Lanyon, 1984).

Expert Knowledge Space

The first questions to address are how to choose the human experts and how many should be involved. The solutions are embedded in another question: What is the structure of the knowledge base? According to Koppen and Doignan (1990), the first system design task is to identify the range of problems the designer wants to solve and to specify what subproblems, or elements, are involved. Next, experts need to be queried regarding which elements are closely connected to other elements. Specifically, if an expert knows how to solve element A, can he also solve B? Experts would be given a list of element pairs in the form of a survey or interview and asked, "If an expert can solve A correctly, then is solving B correctly likely, based on the specific knowledge required to solve A and B?" Elements are thought to be quasi-ordered such that element B may build on A, but that sort of directionality is not required. Each resulting aggregate of elements is thought to represent a knowledge space. Having identified the knowledge spaces, the next task is to identify human experts for each space.

Koppen and Doignan (1990) elaborated their ideas with proofs concerning the separateness of knowledge sets, and addressed to some extent the logistics of collecting data from experts about the structure of knowledge. They used the structure of mathematical problem solving from the point of view of teaching as their reference example. Note that psychologists commonly solve problems with element similarity through the use of factor analysis, or by multidimensional scaling, its close relative, when perceptions of similarity are at issue. It would thus appear that multidimensional scaling could be used on a matrix of similarities of knowledge elements to produce a small and efficient number of spaces.

Extraction of Knowledge

Having identified the knowledge base and the problems to pose to the experts, the experts are asked to solve the problems and to reconstruct the mental steps required to solve the problems. Ericsson and Simon (1980) cautioned that much of the verbal data obtained from the experts can be easily influenced by the instructions they are given regarding how to complete the tasks and reporting how the problems were solved. The relative primacy of the two tasks (solving the problem and reporting the solution method) and different types of probe requests make different demands on memory or require additional processing for the experts to make a response. The excess cognitive load demand required to verbalize the thought process only appears to slow the verbalization process, but not to affect the accuracy of the report.

Verbalization methods used in experiments have been both concurrent and retrospective in nature (Ericsson & Simon, 1980). In concurrent methodology, experts are asked to think aloud as they solve the problem and are given free rein to structure their verbalizations as they find comfortable. Alternatively, the experts may be asked for specific information at various points in the process. In retrospective methodology, experts may be asked to report their solution strategies after each problem, or after a set of problems. Newell and Card (1985) adopted this approach in their think-aloud technique.

A programming problem arises, however, when experts generate different diagrams, different directions of information flow for otherwise topologically equivalent diagrams, and different odds for each linkage. Rege and Agogino (1988) developed an algorithm for synthesizing conflicting diagrams that could arise in the development of expert systems. It would appear, however, that much of the problem that Rege and Agogino identified could be circumvented by compiling the knowledge base from empirical research, whereby the judgments of many experts are bootstrapped over a wide enough knowledge space.

Militello and Klein (2013) encouraged system designers to identify the most challenging problems that their expert sources have ever tried to solve along with the solutions the experts invoked. The challenging problems tax the underlying expertise more so than the routine problems. The more routine problems and solutions will become more apparent in their relationship to big problems as the big problems are analyzed.

If the most challenging tasks are not already known or easily forthcoming, the critical incident technique (Flanagan, 1954) would be a good place to start. The core procedure is simple: Experts are brought together to swap stories about challenging situations, what they attempted to do, and what happened next. The emphasis in the narratives is on the recounting of events, rather than drawing conclusions about why things happened a particular way. Someone needs to transcribe the stories, as they are recounted; the transcriber should also be aware of the need for objectivity as opposed to editorial comments or inferences.

People with less than maximum expertise should also be included in the discussions because their solutions might *not* be what the best experts would have done. This is important information too. The analysis of narratives is essentially qualitative as the events are sorted into meaningful categories; this phase of the procedure might not be a simple task and the data analysts try to decipher what were the most salient aspects of the situation that contributed to the outcome.

Guion (1998) observed that the extraordinary events are more likely to be remembered and recounted, which make the critical incidents technique ideal for finding the situations where maximum expertise is of interest. Routine events are less memorable, but one moderator of the session can probe for them. The critical incidents technique has been very useful over the years for diagnosing and understanding performance issues in a variety of contexts. The technique has undergone some enhancements over the years as well, in which researchers apply a brief questionnaire or checklist to the incidents; the content of the checklists also varies by context (O'Connor et al., 2008; Renshaw & Wiggins, 2007). In the case of expert systems development, the focus should be placed on the types of decision goals and strategies used by the experts (Militello & Klein, 2013).

Validity of Rule Groups

Levi (1989) presented three methods of assessing the accuracy of a complex rule. The first concerns the estimation of the probability of a dichotomous outcome given source data (cues). The variance of experts' estimation of probability (E_{pr}), relative to actual probability (A_{pr}), and a calculation of MPS:

$$MPS = \sum \left[E_{pr} - A_{pr} \right]^2 / N, \tag{12.3}$$

where N is the number of ratings per expert in the data set. MPS is then compared against the variance of the population base rate (b), which indicates no skill in judgment. Skill in judgment is designated as $b(l - b)$.

Levi's (1989) second method was a split-plot analysis of variance experimental design. In the prototypic experiment, R raters (experts) are presented with n examples. The fixed factor is rater differences; raters might be further aggregated into low- and high-skill groups. The repeated factor is the set of n examples presented to each rater. The dependent measure is the estimated probability of an outcome. In the best of situations, maximum variance would be associated with the examples factor.

Levi's (1989) third method involves the decomposition of the correlation between the true and the forecasted outcome, r, such that

$$r = GR_tR_f + C(1 - R_1^2)^{0.5} * (1 - R_f)^{0.5}, \tag{12.4}$$

where $G = r(Y_t, Y_f)$, Y_t is the predicted value from a linear regression of true outcomes on cues, Y_f is the predicted value from a linear regression of forecasted outcomes on cues, R_t is the multiple regression coefficient for cues on true outcomes, R_f is the multiple regression coefficient for cues on forecasted outcomes, and C is the correlation of residuals corresponding to Y_t and Y_f. The goal is to solve for C, such that large C indicates that the human is contributing judgmental power in excess of linear bootstrap methods (policy-capturing regression).

Interpretation Validity

Errors in the knowledge base or inference engine will result in interpretation errors. Additional sources of error may arise at the interpretation level of analysis in the form of poorly crafted statements, contextual influences that arise when statements are organized into reports, and overgenerality caused by high base rates for the interpretive statements. In light of these potential sources of error, Lanyon (1987) suggested that reports from expert systems that are delivered in text (e.g., CBTIs) be evaluated at both the statement and global levels (Guastello & Rieke, 1994).

The form of overgenerality arising from high base rates for interpretive statements is also known as the Barnum effect among personality theorists (Baillargeon & Danis, 1984; Furnham & Schofield, 1987). People who read the interpretations of their own personality test scores are particularly susceptible to the Barnum effect when they rate their reports for accuracy. The influence of the Barnum effect is greater where the interpretive statements are favorable or flattering to the examinee. Expert clinicians, unfortunately, are not immune to the Barnum effect when rating the accuracy of CBTIs generated for their clients (Prince & Guastello, 1990).

Barnum Effect

In light of the strong potential impact of the Barnum effect on judgments of CBTI validity, Guastello and Rieke (1990) developed an experimental design, the full Barnum design, to assess the validity of a CBTI at the statement level of analysis. The analysis partitions CBTI accuracy judgments into two quantities: variance associated with the Barnum effect, and discrimination validity. Discrimination validity is the relative amount of variance associated with the CBTI's ability to distinguish individual examinees. The design also allows for the assessment of the influence of context on ratings of accuracy for individual statements.

The full Barnum design is based on a two-way analysis of covariance with one repeated (within subjects) factor. For the within subjects factor, examinees are presented with two interpretations of their test scores; one interpretation is a real report, and the other is a bogus report composed of interpretations from mean scores for the sample of examinees. Examinees are asked to rate each statement from each report for accuracy. Ratings may be aggregated over the entire report or sections of the report, and are expressed in terms of percent accuracy. The mean accuracy rating for the bogus report is the amount of accuracy associated with the level of Barnum effect in the typical report. The discrimination validity of the CBTI is the difference between mean accuracy ratings for the real and bogus reports. The between subjects factor may be a blocking effect such as whether the real or bogus report was rated first. The covariate is reserved for a property of the report that could influence ratings of accuracy, such as an index of favorability or length of the report. Variations on the design are possible, such as substituting clinicians for examinees as raters, or other differences among raters as the between subjects effect (Prince & Guastello, 1990).

In the full Barnum design, some interpretive statements can be found in both real and bogus reports. Ideally, these statements should be rated as equally accurate in both contexts. A correlation between ratings for a particular statement appraised in both contexts would determine the degree to which the statement's meaning is influenced by contextual effects. High correlations indicate low influence. Low correlations indicate high contextual susceptibility. Our experience indicates that statements that are most vulnerable to context influences suffer from one of two types of vagueness. One type is simple dearth of content; in such cases, the no-interpretation option would be preferable to a meaningless statement, as Lanyon (1987) suggested. The other source of vulnerability comes from the use of fuzzy words for probability concepts such as *unlikely*, *probably*, *possibly*, *little chance* *of*, *strong possibility*, and so on. Reagan, Mostellar, and Youtz (1989) have embarked on a research program centered on the conversion of numerical probabilities into probability words and vice versa.

One controversial aspect of the full Barnum design is the choice of the reference group from which the bogus reports are derived. To design the experiment to be as favorable as possible to the CBTI, one would choose as heterogeneous a reference sample as possible so that the base rates for interpretive categories would be low. On the other hand, the principle of experimental realism would suggest that reference samples should closely resemble the sample of target examinees. In other words, a clinician who needs to make treatment decisions regarding psychiatric outpatients in an urban area wants the CBTI to tell more about the patient than what is already known by just knowing that the examinee is a psychiatric outpatient in a particular urban area.

Meta-Interpretive Reliability

Much of the error in an expert system report may be traceable not only to the rules, but also to the process by which data and rules are transformed into words, words are compiled into reports, and reports are read and understood by users. Meta-interpretive reliability describes the reliability of that process, and is based on the Shannon–Weaver communication model, along with classical and generalizability theories of test reliability (Endres, Guastello, & Rieke, 1992). Meta-interpretive reliability is the degree to which a human expert who is reading a product from an expert system can determine the original data that the program relied on to produce the report. Of additional interest, Endres et al. (1992) found that the humans' accuracy in determining original data pertinent to one part of the report was influenced by nearby text that was generated from different knowledge–inference combinations.

Decision Validity

In the general case, decision validity studies would compare the accuracy of decisions made by the expert system with those that are made by the computer. The expert systems that are most amenable to decision validity analysis are those that make recommendations for treatment as a result of a diagnosis; the effectiveness of the proposed treatment could then be verified by a third party. Not all CBTIs lend themselves to this type of analysis. Perhaps expert systems for medical diagnosis, such as MYCIN (Shortliffe, 1976), serve as prototype candidates for decision validity analysis. SEXPERT, a psychological tool for the diagnosis and treatment of sexual dysfunction (Binik et al., 1988), would also be amenable to decision validity analysis.

In the case of MYCIN, the program collected initial data pertaining to a class of medical ailments, recommended a series of medical tests, incorporated the results of those tests, and produced a diagnosis and prescription for an appropriate antibiotic. The patient either recovered or failed to recover. Decision validity analysis then determined that MYCIN was correct in its decision making for 65% of patients tested, whereas the most experienced physicians in the study were correct in 55.5% of the cases (Carroll, 1987). Meehl (1954) reported long ago that statistical prediction was usually superior to clinicians' or experts' predictions. Present-day CBTIs seem to support this theme.

Signal Detection Technique

Lehner (1989) presented a method based on signal detection theory for assessing the incremental accuracy of an expert system without having to use human experts in the evaluative process. For each node, the accuracy of the rule is compared against the baseline accuracy rate; the difference between rates is *d′*. There is no rule, however, for aggregating the *d′* coefficients over nodes.

The signal detection method is relatively simple but does not address the question of whether the computer program is more accurate than the human. It would be useful, however, to conduct experiments where the program is compared relative to itself before and after its revision. Indeed, rapid prototyping is an interactive design process whereby a model is developed, tested with experts or users, and revised until users and experts are satisfied with the results. The rapid prototyping technique is equally applicable to the design of human–computer interfaces, which could incorporate a modicum of intelligence (Cohen & Howe, 1989; Newell & Card, 1985).

DISCUSSION QUESTIONS

1. How did the concept of computer programming revolutionize the concept of a *machine?*
2. In what basic ways did AI add new value or subsystems to the human–computer interface or human–machine interface?
3. Suppose you were to pick out a good wine to go with dinner. What questions would a wine expert ask the diner? In what order would the questions be asked? If you do not have this knowledge already, ask someone who knows.
4. Look at Figure 12.6 again. What value do veterinary records have for tracking terrorists? Where would the information come from?
5. Face recognition software is now available to individuals on their smartphones. Can this technology facilitate any undesirable social behavior?

13

Complex Systems

A system can be complicated but not necessarily complex. Complicated systems have a lot of parts that all work together in some way. It is the interrelatedness of the parts that makes the system complex. A system where the parts interconnect as A → B → C → D is less complex than a system with the same four parts that are connected by arrows or links that run backward from D to C, or D to A, or in other combinations, as shown in Figure 2.7.

This chapter is concerned with the nature of complex systems and the particular new principles that they bring that affect the activities of person–machine systems and the understanding thereof. The opening section of this chapter describes the central ideas from NDS and complex systems theory that build on ideas presented in Chapter 20, they are presented here as a set of organizing constructs for interpreting what follows, especially when real-world systems are involved. Next, we consider the places where complexity shows its potential for positive and negative effects. From there the chapter expands on human collective intelligence, group coordination, person–robot operation, group cognitive workload, and safety in complex systems.

NDS and Complex Systems

It would be helpful to recall the basic principles of nonlinear dynamics from Chapter 2: attractors, bifurcations, chaos, self-organization, and catastrophes. Some metrics of complexity were also covered earlier: the fractal dimension, Lyapunov exponent, entropy, and the Shannon information function (Equation 2.3). The latter has reappeared in recent works on the complexity levels of system design. The concept of degrees of freedom that was introduced in Chapter 6 is also relevant here.

A complex system is characterized as a group of subsystems that are interconnected by information flows. Information flows can be steady, cyclic, or chaotic. Information may flow in one direction more so than in the opposite direction, but complex systems typically contain feedback loops. Positive feedback loops increase the probability of an event occurring or increase the quantity of a result. Negative feedback loops induce dampening effects, and they are often found in situations that benefit from the maintenance of steady states.

"Complexity theory" got its name from problems involving many, perhaps thousands, of interacting agents (Holland, 1995). Although there could be distinct rules of interaction between any two of them with specifiable outcomes, it was not possible to calculate the final outcomes by calculating all possible pairwise interactions. Thus, simulation tools were developed in the form of cellular automata, agent-based models, and genetic algorithms. Two other distancing concepts in complex systems are the complex adaptive system (CAS) and emergence.

Classical System Simulations

It is often difficult to envision the behavior of a system with many people–machine systems that operate both simultaneously and in concert with each other. The classical approach to system simulation begins with a flow chart of events. The flow charts are not dissimilar to the elementary information-flow diagrams that were used in Chapter 6 to describe serial, parallel, and hybrid organizations of cognitive processes. The difference this time is that many of the elements can be roped together, and the little boxes in Figure 6.5 can be complex processes involving people and machines.

The flow charts, fault trees, and Petri nets that we considered in Chapter 10 are the next closest approximations of a model that is viable for simulations. At the earliest stage in the thinking, Forrester (1961) introduced a system simulation strategy called *system dynamics*, not to be confused with NDS, which featured the concepts of *stocks* and *flows*. Stocks are objects that accumulate at different points in a process, such as quantities of beer in a warehouse, which are released for use in other parts of the system at later times. Flows are movements of information or objects over time, and they are represented by the arrows that run between the boxes in a typical diagram of a system. It is sometimes debatable as to whether an entity, such as money or the accumulation of trash in one's house, is represented better as a stock or a flow. The decision to characterize the entity one way or the other may be related to the questions that one is trying to answer.

The representations of flows and movements require some mathematical modeling. If the laws of physics are applicable in a given situation, the equations are probably known, and it would be relatively easy to plug them into the system. At other times, however, the models do require some first-hand data collection and analysis to determine which mathematical models are optimal for flow, and what statistical odds are associated with the different possible combinations of events. In essence there is not a great deal of difference between the mathematical rules that are applied in a system simulation and the rules that are used in an expert system. The challenges to the validity of the system are essentially the same as well. An apparently authoritative computer simulation can produce some grossly misleading results if the underlying mathematical and statistical rules are plucked out of thin air with no assessment of their validity.

Artificial Life Simulations

One aspect of the system dynamics strategy that gives it a classical character is that the overall structural representation of the system does not change over time. The thinking behind complex adaptive systems is that the network of boxes and arrows forms, reforms, and regroups spontaneously when the system is in a high-enough state of entropy. This phenomenon of self-organization has been called *order for free*, meaning that the system takes on its own order without the intentional help of management, city planners, or similar others (Kauffman, 1993). From this perspective, we can conceptualize each machine, machine operator, and each other decision maker in the work unit as an *agent*. We can then conceptualize any work system as the end result of a dynamic process of self-organization that was induced by the myriad bilateral interactions among the agents. With so many possible agents, information flows, and material stocks, it becomes difficult to predict the outcome, or *fitness*, of a system by simply looking at it. Thus, it becomes necessary to rely on system simulations for information that accommodates the potential for self-organization.

There are three basic genres of artificial life simulation: agent-based models, cellular automata, and genetic algorithms. For all human factors intents and purposes, they are expert systems. The data consist of rules for agent behavior that the user can lock and load into the program through an interface. The inference engine allows the user to tweak parameters on the model rules and then watch what happens. Displays may be numeric, but often they are pictorial displays showing patterns of color that correspond to patterns of agent behavior. In all three cases, it takes some skill to translate a research question into the program's capabilities and interpret the visual patterns.

Cellular Automata

According to Levy (1992), the concept of *cellular automata* originated with von Neumann's vision of characterizing all life as the flow of information, which in turn can be represented as a logical program. The core applications emulated biological processes (e.g., Ulam & Schrandt, 1986). Cellular automata are iterative calculations that are defined by Boolean algebra rules by which the information related to one cell, or pixel in a computer image, affects the behavior of each adjacent cell at the next discrete step in time. The simplest concept of a cellular automaton begins with the coloring of a single square on a sheet of graph paper, for instance (Figure 13.1). The second step is to color the adjacent squares in the 3 x 3 matrix that surrounds the initial square according to a rule of some sort. Patterns emerge as the process continues. Cellular automata are *spatially local*, meaning that the action of one cell on another is limited to the adjacent cells only. The system is also assumed to have no memory beyond the immediately preceding discrete step (Wolfram, 1986). Importantly, small differences in the initial values of rule parameters can produce some very different end results. Thus, the concept of sensitivity to initial conditions, which is a hallmark of chaos, can be observed in cellular automata.

Cellular automata have acquired numerous applications that go beyond biological phenomena. The image shown in Wolfram (2002, p. 984) illustrated the eventual distribution of sand on a vibrating plate where the vibrations are coming in from five different directions. Another noteworthy example comes from studies in cryptography. The goal of encryption is to obscure a message in such a way as to allow only an intended recipient to decode the message. Encoding systems from the middle of the last century relied on mechanical devices that encoded and decoded the message. The encoding systems evolved to include sequences of encoding logic; for instance, the first time a message is passed the encoding would use a hypothetical System A, but the second iteration of the conversation would use a variation, call it System Al. It now appears, however, that cellular automata technology is, in principle, capable of cracking any encryption system (Wolfram, 2002).

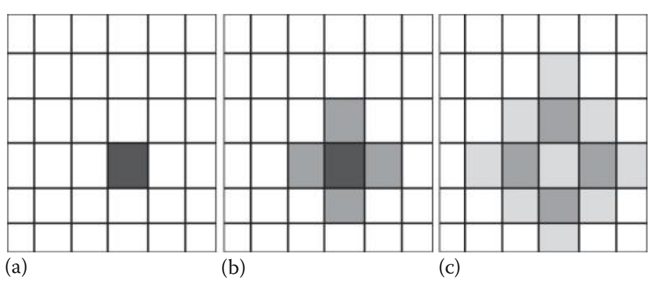

FIGURE 13.1
An elementary cellular automaton in three steps of its evolution.

Agent-Based Models

If one imagines an autonomous agent as described in the previous two chapters, an *agent-based model* will simulate the activities of many such agents in a virtual environment. Although they are also based on the principle of adjacent pixels, agent-based models are not limited to rules that depict literal location of objects (or conditions) in physical space, as the original cellular automata once did.

The agents can be loaded with rules for spatial mobility, rules that reflect economic interactions, and rules that reflect other forms of social connection such as genealogy and cultural similarity. This combination of rule systems, which are intended to emulate real-world behavior, gives rise to *artificial societies* (Epstein & Axtell, 1996) and studies of growing patterns of real urban development and industrial networks (White, Engelen, & Uljee, 1997). Among the more intriguing results from Epstein and Axtell are that if one puts many agents into a confined space and allows them to "trade" many commodities according to the same rules of trade, there will be a distribution of wealth that follows the inverse power law. If one were to compare terrains with high agent density, which resemble an urban area, with low-density terrains, the high density terrains support many more agents and many more products, but the distribution of wealth is more disparate. The disparities in wealth distribution arise simply from the large number of agents and commodities and are not based on any assumptions about capitalism or any other economic system; no such assumptions were programmed into the model.

Again the system user would be looking for dot patterns that indicate self-organized behavior of some sort that is the end result of local interactions between agents. The example in Figure 13.2 arose from a problem in economics (Bankes & Lampert, 2004). The end

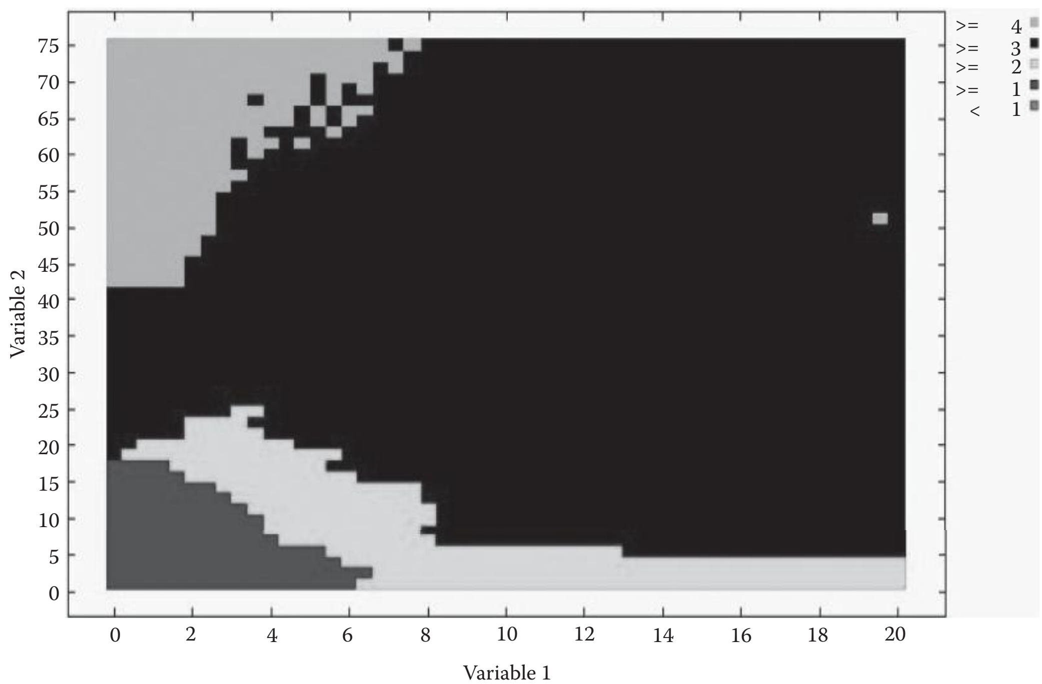

FIGURE 13.2
Example output from an agent-based model. (Adapted and reprinted from Bankes, S., & Lempert, R., *Nonlinear Dynamics, Psychology, and Life Sciences*, *8*, 231–258, 2004, with permission of the Society for Chaos Theory in Psychology & Life Sciences.)

result consisted of four qualitatively different outcomes that were distributed along the ranges of two other variables. Sometimes one can display the key results as multimodal frequency distributions or line graphs and obtain an equivalent level of meaning (Elliott & Kiel, 2004).

Sometimes there is only a small difference between an agent-based model and an agent-based program that goes live in the real world. In the previous chapter, we considered the case of an agent-based mining program for investment decisions wherein the program incorporates the investor's trading strategy, collects appropriate trading data from the real world, and evaluates the efficacy of those trading decisions. Consider next that the trader is happy with his results and decides to implement the strategy. Now imagine thousands of other investors doing the same thing with similar or different trading strategies, and we have high-frequency trading programs. The agents are reading what all other agents are buying and selling on a moment-to-moment basis, faster than any human can decipher the data and apply the decision rule.

Unfortunately, all the trading agents can self-organize in ways that could not be predicted from the analysis of a single trader's algorithm and a data-mining strategy, precisely because the myriad agents were not taken into consideration. For instance a flash-crash was reported on May 10, 2010, when the Dow Jones Industrial Average plummeted 900 points (almost the largest daily drop on record) within a few minutes, then regained itself 15 min later (Wilkins & Dragos, 2013). Prices of specific stocks went haywire in the interim. The financial damage incurred by the automated traders has not been ascertained.

Genetic Algorithms

Genetic algorithms are models of machine learning that are heavily inspired by biological processes. To stick close to the biological metaphor, characteristics of a population are ascribed to *genes*, which are recombined within the population according to a rule base that allows or constrains possible patterns of recombination. Mutation rates are introduced, and the fitness of resulting organisms can be assessed. Some algorithms involve compromises between genetic rules and problem-specific constraints. Hybridized models are known as *evolutionary programs*. Pure genetic algorithms have the advantage of theoretical support and wide applicability but have been known to produce inappropriate results in some applications. Evolutionary programs, on the other hand, tend to be more viable for specific programs, but are narrower in generalizability (Michalewicz, 1993).

For the applications to nonliving systems, one might conceptualize the design of a device, such as an automobile, as an organism with many features that are represented by virtual genes. The program can generate design mutations and cross-breed one design with another to devise a radically new vehicle design or perhaps just a design that represents a gentler transition between an existing popular body type and that of a popular competitor's product. Designers would have to determine what constitutes fitness under these circumstances. The cost of machine retooling for the new design would be one concern. Consumers' views of *gorgeous* and *ugly* would be another. The vehicle would also have to operate properly.

Nolfi and Floreano (2000) expressed a preference for genetic algorithms over neural networks as models of robot learning. If the device were going to emulate a complex adaptive system, it would need adaptive capability. Neural networks are good for programming system learning and fixedness, but the attractor strength of a learned behavior simultaneously introduces an adaptive cost, making the system unlikely to identify and select a novel response to a novel stimulus. The issue of adaptability brings us to the topic of a

CAS. Further issues in system learning are considered later in the sections on collective intelligence and human group coordination.

Complex Adaptive Systems

The concept of a CAS has been used to understand the distinction between healthy and unhealthy organisms, social units, and business organizations. It is only a small stretch to consider healthy and unhealthy sociotechnical systems (large-scale human–machine systems) in the same fashion. The evolution of such systems is characterized by punctuated equilibria—periods of stability that eventually end when new technologies and processes are introduced. Similarly the sunk costs in one technical system will inhibit change to new systems unless the expected advantages are great and worth the effort of retraining the personnel and changing the equipment.

The CAS has three hallmark features, according to Dooley (1997, p. 83): Its order is emergent rather than determined by the designers, its history is irreversible, and its future is often unpredictable, at least in the conventional sense. Of course, the concepts of chaos and its control, cellular automata, agent-based modeling, and genetic algorithms have all evolved in response to the need to forecast possible futures.

Underlying the three hallmark features are numerous counterpoints between the realities of a complex adaptive system and the implicit assumptions of conventional business managers and systems designers. In conventional thinking, which typically assumes linear relationships among system variables, outcome levels are proportional to input levels, and events can be explained by simple cause-and-effect explanations. In the CAS, little things can have big consequences, and vice versa. Control variables within systems behave differently, although two or more variables may behave the same way (Guastello, 2002). Chapter 10 illustrated the limitations of simple cause models of accidents and the appropriateness of a nonlinear model. Furthermore, phenomena emerge as I have stated several times in this chapter already.

Conventional management thinking tends to ignore what it regards as random blips in system behavior and to maintain equilibrium as its objective of control. A CAS, however, would regard many of the blips as not the result of randomness at all, but possible signs of a larger pattern about to happen. The control objectives of a CAS would focus on identifying the possible patterns that are in motion, and navigating them with a repertoire of nonlinear change processes—even if stability remains the desired objective.

In conventional thinking, creativity was reserved for research and development departments or marketing activities, whereas standard operating procedures are otherwise required. In a CAS, standard operating procedures still have their place, but creative thinking is nonetheless considered central to survival and therefore should be central to organizational life. Kjellen's (1984b) simulations indicated that small deviations in a procedure or work specification could be compounded into a large event, recognizable as an industrial accident. A CAS would recognize the deviations before they got too far out of hand and correct the situation.

Self-organization occurs from the bottom up, whereas hierarchies operate from the top down. This class of phenomena gives rise to new concepts of networks and their relevance to the creation of emergent business activity. The conventional frame of reference might regard hierarchies as inevitable and prefer hierarchical control as a means of dominating the agents at the lower order of the hierarchy. The trend in CAS research, however, is that bottom-up self-organized behavior is more efficient. Hierarchies do emerge nonetheless, thus indicating that systems gravitate toward inefficiency, which is a condition that should be deterred. Studies with hierarchical workgroups show that the upper levels of the hierarchy are often overwhelmed by the activities from below, although some managers are skilled at maintaining efficiency in the face of great entropy (Guastello, 2002).

Emergence

Emergent phenomena should be anticipated whenever multiple PMSs are interacting. They cannot be readily decomposed into more elementary precursors or causes. They often occur suddenly, hence the word "emergency," but suddenness is not a necessary feature of emergence. The earliest concept of emergence dates back to a philosophical work by Lewes in 1875 (Goldstein, 2011). It crossed into social science in the early 20th century, when Durkheim wanted to identify sociological phenomena that could not be reduced to the psychology of individuals (Sawyer, 2005). The existence of groups, organizations, and social institutions are examples; bilateral interactions among individuals eventually give rise to norms, deeper patterns, and other forms of superordinate structure. In the famous dictum of the Gestalt psychologists, "The whole is greater than the sum of its parts." Thus, some of the ideas described here are more than a century old, but it was not until the 1980s that social scientists began to acquire the analytic tools to exploit them fully.

When a whole becomes greater than the sum of its parts, it is exhibiting a bottom-up strategy, whereby the elementary parts are in place first, they interact by some means, and produce a whole that reflects the interactions that took place. The artificial life simulations captured some scenarios by which bottom-up emergence could occur. Whereas self-organization is a process, emergence is a result of the process. McKelvey and Lichtenstein (2007) outlined four different types of emergence. The simplest was the *avalanche*, which produces power law distributions of objects of different sizes. The second was the phase shift. The internal organization that would be required for a species to differentiate into more specific types is more complex than breaking into little pieces; phase shifts also occur when the organisms hop from one niche to another suddenly. More generally, a phase shift is a reorganization of the internal structure of the system. Still it is not necessary for a hierarchical internal structure to occur.

The third level of emergence is the formation of a hierarchical relationship among the parts of the system. Driver–slave relationships are examples. A different type of example occurs when a person or group collects information about its environment and forms a mental model of the situation. The mental model does not exhibit the top-down supervenient force until people start acting on the model and the model persists after some of the original group members have been replaced. The presence of an active top-down component reflects the fourth level of complexity for emergent events. Arguably, the dynamics of bottom-up and top-down influences are matters of degree and relative balance.

Goldstein (2011) indicated that emergent phenomena could be still more complicated. First, the automatic internal organization that characterizes Kauffman's (1993) "order for free" might have been overemphasized by some writers. Boundary conditions also shape emergent behavior. Second, in a complex hierarchical system, there can be numerous subnetworks of agents that are drawn from different hierarchical levels. The subnetworks can be connected horizontally and interact in all sorts of combinations.

Phase Shifts

Phase shifts are essentially cusp catastrophe processes (Chapter 2). Organizational change, or change in a complex human–machine system, can be slow and gradual, and thus *evolutionary*. On occasions, it might be abrupt with widespread implications, and thus *revolutionary*. A cusp catastrophe model for this dynamic appears in Figure 13.3. The two control parameters are pressure to change and resistance to change. The sources of resistance, such as sunk costs in an existing system, would induce the changes to become more abrupt. Two tales of technology depict the roles of evolutionary and revolutionary dynamics.

Staudenmeyer and Lawless (1999) described a telephone system where the organization's rate of innovation outpaced the decline of older technologies. As new features were added to the system at the users' requests, the technology became "brittle," meaning that a small change required extensive relinking of the many parts of the system so that the system could stay internally coordinated. The technology eventually reached a crisis point where it became more sensible to rewrite the big program from scratch. The demands on the business for doing so were extensive. As the pressure to change the system mounted, the organization's knowledge about how the program was initially assembled was dwindling. Fortunately for the telephone company, there were still a few people left in the organization who still remembered how to manipulate the original program code, all 15 million lines of it, after 20 years. There still remained the problem of how to shut down the existing system while the new system was being installed and tested.

The transition job required not only people and programming machines, but a massive reorganization of personnel into different workgroups, compared with their usual work patterns, to coordinate the job. The dynamics of workgroup coordination are considered later in this chapter.

The second story of radical technological change involved two competing technologies for a particular type of computer program. Both technologies enjoyed evolutionary growth until one producer got the idea to make his program more compatible with other programs that the users might already be buying and using. The system with the greater number of compatible linkages to other programs suddenly made great market strides, to the detriment of the competitor. The users who had initially selected the technology that increased the compatible linkages could easily enjoy the ride. The other users were required to seriously reconsider their equipment choices and many switched their equipment (Lange et al., 2000).

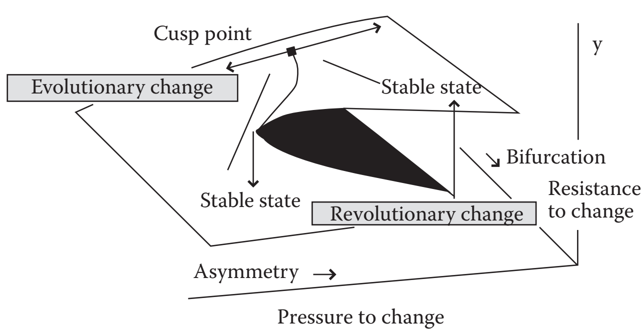

FIGURE 13.3
Cusp catastrophe model for evolutionary and revolutionary change in organizations. (From *Managing Emergent Phenomena: Nonlinear Dynamics in Work Organizations*, p. 84, by S. J. Guastello, 2002, Mahwah, NJ: Lawrence Erlbaum Associates. Copyright 2002 by Lawrence Erlbaum Associates. Reprinted with permission.)

Complexity Catastrophes

The foregoing example of a telephone system concluded successfully. There were some positive contributions from group dynamics and team coordination that helped the process. Coordination issues are considered further on in this chapter. The point for present purposes, however, is that there is no guarantee that self-organization will arrive at a successful conclusion. Kauffman (1995) introduced the concept of a *complexity catastrophe*, where the system was too complex to reorganize effectively. A real-world example would be an organization that designs and implements one new product within a fixed number of years and succeeds. The same organization might be able to introduce two or three such products in the same period successfully. But when it tries to introduce, say, eight of them, they all turn into a total mess of design errors, system failures, and financial embarrassments.

The introduction of too many new products could be alternatively characterized as a group level problem in cognitive workload. That has already been shown to be a literal case of catastrophe theory dynamics for work performance (Chapter 9). Although it would appear that the problem could be alleviated by expanding the workforce to include more designers, engineers, and pertinent others, there still remains the problem of getting all the products out the same corporate door and managing their implementation in the field, especially when the products are interrelated in some way. There is a qualitative difference between eight free-standing products that are equivalent in scope to each other, and one big product that is eight times the size of one of the smaller ones with additional interrelationships among the eight components.

Synchronicity

Synchronicity is a special case of self-organization where the agents in a system act independently at first, but eventually start doing the same thing at the same time (Strogatz, 2003). One of the more dramatic illustrations was reported a few hundred years ago regarding a Southeast Asian firefly. In the early part of the evening, the fireflies would flash on and off randomly. After a few hours, however, they would all start flashing together so that the entire forest would flash on and off. *Synchronicity* can be produced even in nonliving systems with only minimum requirements—two coupled oscillators, a feedback channel between them, and a control parameter that speeds up the oscillations. The oscillators synchronize when they speed up fast enough. The principle also has been demonstrated with electrical circuits and mechanical clocks.

Real-World Complexity

This section expands on the sources of complexity in real-world systems and some of the problems it could present. Problems are considered at the level of the individual operator's experience, the system design perspective, and decisions that fail from a lack of awareness of complex systems principles.

Individual Operators

Although psychology has given up trying to quantify the operator's mental workload capacity in terms of bits of information a long time ago, the workload imposed on the operator can often be measured effectively in terms of the number of states and probabilities (p) of those states. Shannon's entropy function is repeated here for convenient reference:

$$H_s = \sum_{i=1}^{r} p_i \ln(1/p_i). \tag{2.3}$$

According to Endsley et al. (2003), complexity shows up in several forms: (a) the number of objects, states, symbols, and properties of the system components; (b) the number of linkages or relationships between them; (c) the breadth and depth of hierarchical layering present in the linkages (hierarchical clustering); and (d) the degree of branching present in the linkages (logic propositions) leading to conditional operations or modes of operation (p. 138). They go on to say that unnecessary complexity can be eliminated at the design stage by stifling "feature creep," which is the tendency for designers to add new features (often called "upgrades") to keep their products looking "new and improved." One needs to set priorities for which features are most likely to be used the most often and leave behind the ones that the operators would use only rarely.

Hierarchical layering involves the use of modes for displays and controls; the problems with mode error were discussed earlier. The recommendations here would be to keep the use of modes to a minimum. Similarly, the levels of automation versus manual control should be kept to a minimum as well.

Simplifying Designs

The trick to simplifying a design and still keeping it effective for its intended purpose is to look for places where the system demands too much information to operate. The hunt begins by specifying three types of functional requirements: the goals of the user for the design, the physical outcomes that are intended, and the process that gets from the former to the latter (Suh, 2007). The next step is to identify modules of connectivity that are uncoupled, decoupled, and coupled. In an uncoupled module, the control–result sequence is independent of other control–result sequences; uncoupled modules produce the least information drain.

Coupled systems often arise when there is a sequence of steps involved, such that if the operator needed to change one element, the other elements would need to be readjusted. A repetition of steps would result. Coupled operations tend to be less resilient against random fluctuations; a perturbance that affects one subsystem directly affects other subsystems indirectly. One never really knows the total information requirement of a system; one can only work with a particular representation of the system. Nonetheless, the quantity of information specified in Equation 2.3 becomes magnified in a coupled system such that the summation sign become a cross-multiplication, Π.

Coupled systems are typically less transparent than uncoupled systems, which presents another level of information demand that Suh calls "imaginary complexity." Imaginary complexity can start at the design stage where the understanding of the user's task is unclear, and provisions are made for options and features that are not necessary. From the operator's viewpoint, the control of coupled subsystems is not well understood leading to extraneous operations.

Suh suggests two solutions to the design dilemma, once the sources of complexity have been identified. One is to decouple the system as much as possible. Another is to introduce *periodic complexity*, which in principle makes the operations problems go away after a period. The illustrative example was the airport and airline confusion that is caused by bad weather conditions—cancelling or rerouting flights and passengers and take-off delays. Once set in motion the airport confusion will continue until the end of the 24-hr cycle, after which time the flight schedule starts over again and the problems of the previous day are no longer in operation.

Endsley et al. (2003) observed that a similar problem arises from the use of automation, which in principle keeps the human input and error rate to a minimum, but often does not. They noted that the design trend at the time was to program systems to stack jobs sequentially and leave the operator out of the loop once the stack was set in motion. The problem, however, occurs when something in the automated system does not work right, the operator is thus slow to notice and intervene, and concomitant errors pile up.

Helander (2007) recommended reducing the number of feedback loops in the system as a means of decoupling. The rationale is that feedback is valuable if the system needs to be corrected by the operator. When there is a low need for correction, having fewer loops presents less complexity. The iterative product design sequence, now familiar, does its best work when couplings are involved, and more couplings would predict that more iterations would be needed to get the system worked out properly. Fewer couplings would reduce the number of iterations needed.

Revenge Effects

The failure to recognize the complex nature of a complex system can produce some untoward consequences. *Revenge effects* occur when our attempts to control a problem by means of a technological intervention actually end up making the problem worse (Tenner, 1996). Here are some examples: An animal species is introduced to control a nuisance insect. The insect population is indeed controlled, but the predator animal overpopulates because its natural predators were not available to curb its population. Similar dynamics have occurred with plant species that were transported out of their natural environments to new environments. The eucalyptus tree that was a hardy source of good wood in Australia produced worthless splinters elsewhere in the world while it overtook ecological niches that were occupied by viable crop trees.

In medicine, the prevention of some common ways of dying only serves to increase the odds of death by other diseases that are less tractable medically. In recent years, medicine has curbed the odds of death from heart disease or heart failure, only to meet with an increase in the incidence of diabetes, pneumonia, or Alzheimer's disease among the elderly.

In computer technology, the joy that is shared by all workers who have a computer on their desks results in less time spent on expert work, with more time spent on lower level word-processing tasks that were once done by nonexperts. There are other examples of information technology appearing to save work, but instead only saving work for some people and making more work for others. Similarly, the "paperless office" was forecasted since the early 1980s, but a decade later, more paper was being consumed not less. Instead of making more information products, more *versions* were produced on the way to getting to the final version.

Sometimes, fortunately, revenge effects happen in reverse. In the case of war, we often find new ways of saving lives that carry over to postwar life. Other inventions such as refrigeration had the same fortuitous effect; the original objective was to transport food supplies for the military across the Pacific Ocean without spoiling. Civilians eventually benefited from a radical improvement over the old icebox.

Revenge effects are not the same as side effects or trade-offs. They are the result of the complex nature of the system. When a limited intervention is taken to change the system in some obvious-looking way, another aspect of the system reacts in ways that were not anticipated. Tenner (1996) reminded us of the fabled Murphy's Law: Anything that could possibly go wrong will go wrong at the worst possible moment.

Side effects and trade-offs, on the other hand, are properties of the intervention technology, rather than of the system itself. A medical side effect would occur when a medicine that would relieve one problem could induce another problem, usually in rare circumstances or with prolonged use. A trade-off is similar, but we call it a trade-off if the secondary effect is more probable: The intervention might indeed stop a problem but it gives you a different one instead, and you now have a choice as to which problem you prefer.

Revenge effects should not be confused with any of three other similar nasty habits of complex systems, according to Tenner (1996). A *repeating effect* is where we end up doing the same thing more often rather than gaining time to do other things. In word processing, it appears that people make more versions of a document, but not necessarily any more information products. People who are on the receiving end of these closely similar documents must take care to discard any old versions of the same documents so as not to rely on errors in the earlier versions. Cutting and pasting does indeed save retyping of potentially long passages.

A *recomplicating effect* occurs when a technology that was first introduced to simplify some tasks ends up introducing new and different complications. For example, an organization prints and mails a document to a large but defined group of people. Then the organization decides that it wants to restrict distribution of the document to the defined group and not make the document easily available to the world. So it sets up a password system, which takes a lot of work, and then finds out that the password system requires a lot of maintenance because of high turnover in the group membership and because members changed their Internet service providers (IPs) frequently. IP numbers were one component of the user–identification system.

A *recongesting effect* occurs when an idea that was very attractive at one point in time becomes so popular that the system becomes useless because of the sheer quantity of people trying to use it. Tenner (1996) gave some examples: The electromagnetic spectrum was once sufficient for broadcasting. Now it is overloaded with media and telecommunications channels, such that users must revert to phone lines for a medium. Meanwhile, the telephone companies are running out of phone numbers in some regions of the country. At this time of this writing, landline phone traffic, short e-mails, and a lot of web browsing are now being offloaded to cell phone bands.

As another example of a recongesting effect, shortcuts that are introduced to an urban highway system to relieve congestion might have the effect of moving greater congestion to another location because everyone is looking for a shortcut. Indeed, there are occasions where driving through the city streets is faster than taking the highway route.

New Complex Systems

Several large-scale systems are currently in development, notably NextGen air traffic control, the smart power grid, multirobot systems, and new automation capabilities for private automobiles. The former two are described next. Multirobot operation is considered later in this chapter after introducing a few more principles of human operation. The automobile applications are considered in the next chapter in conjunction with other roadway issues.

NextGen Air Traffic Control

The workload issues facing air traffic controllers have been well recognized along with growing volumes of air traffic, even though solutions have been slow to materialize. A mega-solution seems to be on the horizon carried under the slogan "NextGen." The core idea involves new forms of automation, one for aircraft, and one for the control tower.

The take-off and landing portions of the flight put heavy workload demands on the pilot, and it is what air traffic controllers grapple with most of the time. The last portion of a landing usually requires the pilot to deautomate and negotiate the landing to the target location based on cues received from the direction perception of the interface, possibly some heads-up displays, and the control tower. The new initiatives involve replacing the human vision and control operations with synthetic vision. Synthetic vision involves a system of cameras outside the aircraft that produce a digital video feed to the pilot; at present it is not clear how much information should be presented to the console and how much to an HUD (Ellis, Kramer, Shelton, Arthur, & Prinzel, 2011). In principle, it will improve the visual display and automatically connect to the automated landing operation at the tower.

The other portion of the innovation is to build a digital communication link between the aircraft and the control tower. An automatic control system is now under development that will land properly equipped aircraft automatically. One challenge at present is that controllers in different parts of the country now work with different interfaces and symbol systems, and a standard would need to be developed that is at the moment not widely shared by anyone (Billinghurst et al., 2012). A recent survey of controllers, however, showed a split in opinions regarding the relative merits of NextGen (Cho & Histon, 2012). On the one hand, it would shift workload associated with routine conditions to automation thus freeing up human channel capacity to respond to unusual conditions. On the other hand, it could place more load on the human for having to monitor the automation for emergency conditions.

The phase-in period for even the best system design is not going to be immediate. The lifespan of a commercial aircraft is approximately 30 years. It would be unreasonable to trash aircraft that are still in their "adolescent" years for not having the latest and greatest communication links. The result is a complexity issue that is not completely different from what was observed in the power plants in previous chapters—cascades of technologies from different generations of equipment design all in simultaneous operation. The NextGen systems, once implemented, will need to contend with a cascade of technologies and aircraft. The issue currently is how the mix of equipage makes additional complexity for pilots and controllers who might have to decide which configurations of automation should be applied to an aircraft (Gregg et al., 2012; Lee & Prevor, 2012).

The Smart Power Grid

The electric power grids that operate in North America and the rest of the industrialized world provide a continuous source of power to industries and households through a system of generating facilities and distribution facilities. Irregularities in the cost of production, sources of production, and demand are neatly flattened through the sourcing and delivery systems. It is essentially a chaotic controller that has been overlooked as such for a good many years. This is not to say that micro-outages have not occurred in recent years when the volatility of supply and demand components exceeded the tolerance of the switching systems, which are mediated by humans to some extent.

The so-called *smart* power grid involves two aspects of technological evolution. One is the source of power. The primary forms of power currently come from the processing of fossil fuels, nuclear power, and water turbines. Solar, wind, and geothermal energy are new sources that have become very attractive because of the rising costs and limited supplies of fossil fuels and the unpopularity of nuclear sources in the wake of the incidents at Three Mile Island and Chernobyl. The challenge for large-scale solar power is that the power needs to be created when and where the sun shines, stored, and transferred to requesting power plants at disparate locations. The challenge for wind power is similar.

The switching concept behind the smart grid is essentially the same idea behind the relatively local grids that exist now except that the scale of the imagined system increases several orders of magnitude to encompass entire continents (Bushby, 2011; Holmberg, 2011; Ivanoff, Yakimov, & Djaniykov, 2011; Sarfi, Tao, & Gemoets, 2011). The implementation is easier said than done as it requires all participating plants to have state of the art equipment and compatible technologies; many installations are a long way from that standard. The smart grid system is expected to run through the Internet.

Compatibility of software and equipment is a recognized problem in grid computing, especially when data mining is involved (Dubitzky, 2008). The implications for usability and the new roles for human operators have not been disclosed by industry sources at the present time. Data-mining architecture for "decision making" have been considered, however, along with ways for making the results transparent to the user.

Modularity

Hollnagel (2012) introduced a useful method for analyzing complex systems into their component parts and linkages. The functional resonance analysis method (FRAM) characterizes each component as having six parameters: its inputs, outputs, preconditions for action, resources that the component uses, means of control, and time horizon for operation. The relationship among components is depicted as a flowchart of hexagonal boxes and multiple links from each parameter to any parameters of the next component. A small-scale example appears in Figure 13.4, where outputs from two components are inputs to the subsequent components, and a resource provided by the first component is used by the third. All components, parameters, and links should be specified with the understanding that some parameters and links could be more important than others.

A component could fail for a number of reasons related to its parameter values: speed of action was too fast or too slow; a distance error where something fell too short or long, a sequence error including repetition, wrong actions were taken on the wrong objects, duration of an act was too long or short, something went in the wrong direction, and timing errors (p. 72). The next step, which makes good use of complex systems thinking, is

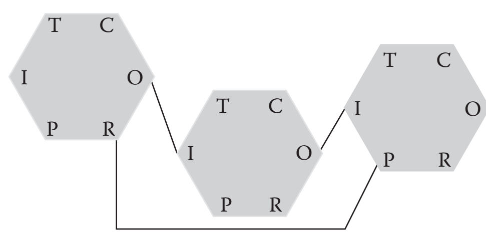

FIGURE 13.4
Example of FRAM modules.

to analyze variability in each of the components or sources of error. Here one determines which ranges of tolerance, and the extent to which variability compounds as one process links to another. The last step is to determine the consequences of errors.

Multiple PMSs

DeGreene (1991) made three points that are relevant to the present exposition. First, the early person–machine systems were conceptualized at the interface level where bilateral interactions between one person and one machine occur. Information that was considered relevant was predominately atheoretical in nature and mostly geared to defining critical numeric values for one function or another. The newer technologies have enabled a greater number of functions to be controlled by multiple person–machine systems. As depicted in Figure 13.5, information flows between the person and machine pretty much as usual, but there are also flows between machines and between humans. The machines are linked into networks. People can communicate either in real space–time or through networks via machines. Information from one PMS affects another, and delays in communication can result in various forms of uncoordinated action. The information that transfers is often far more complex in nature.

Second, he observed that although the concept of the system had been introduced at an earlier time (Meister, 1977), the full implications of the system concept had not been fully explored. The concepts of self-organization, the complex adaptive system, and others considered in this chapter have only been possible and practicable in the last 15 years or so.

Third, chaotic sequences of events do occur. Sources of chaos are potentially inherent in two places. One is in the information flow that is conveyed by the machine's displays. The other is in the output of the PMS. Stimuli arrive over time in many cases. Humans are highly variable in their ability to interpret patterns arising from chaotic or simpler nonlinear sources (Chapter 7). Thus, historical and predictive displays have been used for decades with growing levels of sophistication. Chaotic controllers can take two basic forms. One form might be designed to regularize the flow of information to the operator. The other would recognize and interpret patterns and identify options to the user. The specific means of control were detailed in Chapter 7. Self-organization and synchronization are other outcomes of a chaotic system, which is to say that the system might not remain in its chaotic state for an extended period of time.

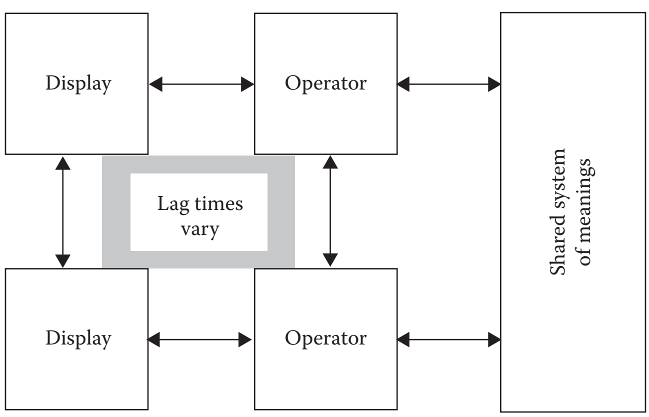

FIGURE 13.5
Triadic representation of multiple person–machine system interactions.

Networks

Networks have become a favorite topic of study within the NDS and information technology communities, although the idea originated in mathematical social psychology in the early 1950s. The applicability of the ideas for interpreting the structures of brain circuits is also gaining momentum. The uptake within the HFE community has not been great to date, but with the rapid growth of massively interconnected PMSs, data mining, autonomous agents, and new opportunities for system failure, it would probably be a good idea for the HFE community to consider at least the basic principles and some particular research findings that appear relevant at present. Those wishing to pursue the mathematical theory of network analysis should consult Newman, Barabási, and Watts (2006), Boccaletti, Latora, Moreno, Chavez, and Hwang (2006), and Rubinov and Sporns (2010).

Social Networks

A social network is a set of individuals, or possibly groups or organizations, that communicate with each other in some fashion. The limits of some networks may range in firmness from rigid boundaries to permeable and diffuse boundaries. The concept of social networks was first introduced by Bevales (1948), and the study of networks, once known as *sociometric analysis*, was greatly augmented by the introduction of graph theory (Harary, Norman & Cartwright, 1965). Networks are represented as geometric configurations, or geodesics, that show the placement of each communicator relative to others in the network with connections drawn among those members of the network who communicate directly.

Two classic configurations appear in Figure 13.6a and 13.6b. The star configuration shows one central person with others radially distributed. The pentagram shows the same number of people who are capable of N-way communication. These and other configurations not shown represent the formal and informal communication patterns in organizations and any hierarchies that are involved. A formal communication network is typically whatever is drawn on an organizational chart of authority structures or a flowchart of a system operation. Informal communications are those that occur but are not drawn on the chart; the patterns of interaction are probably influenced by dyadic task interactions, friendship ties, or family ties.

Information about a network can be valuable for understanding sources and destinations of messages, influences that produce distortions of the message en route, and the social or workgroup processes taking place. They might also serve as the basis for engineering telecommunications equipment or the location of transportation hubs (Freeman, 1979; Kantowitz & Sorkin, 1983). They have had some value for uncovering "organized crime" activities, understanding communities, markets, social change (Jacobsen & Guastello, 2011; Wellman & Berkowitz, 1983), job mobility (Levine & Spedaro, 1998), and the transactions within and between discussion groups on the Internet (Wellman, 1997) as well.

The content of communication could be as important for understanding emerging networks as the raw quantity of communication. For instance, Bales, and Cohen (1979) classified communications within a 3-D taxonomy of dominant versus submissive, friendly versus unfriendly, and emotionally expressive versus controlled. These parameters of communication in combination could lead to cohesive (and probably cooperative) or polarized (and probably competitive) groups (Axelrod & Hamilton, 1981; Flache & Macy, 1997).

At one level of analysis, communication (in the conventional use of the word) and social exchange are not appreciably different, but the content of exchanges or communications can

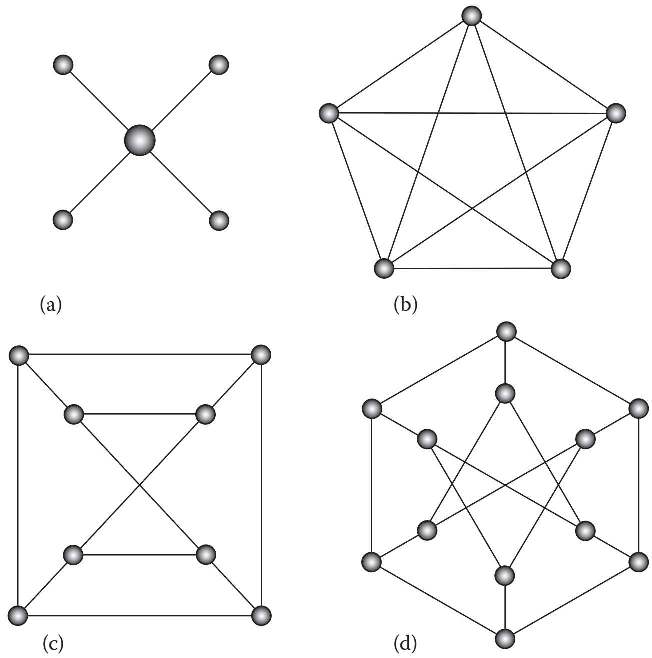

FIGURE 13.6
Examples of network structures.

promote some different dynamics. In a task group, the exchange might consist of approval for approval, or perhaps approval for task compliance (Flache & Macy, 1997). In friendship ties, however, a bit of time is required for a link to fully establish. A link will form if the friendship initiation attempt of one actor is reciprocated by the other (Zeggelink, Stokman, & van de Bunt, 1997).

Smith and Stevens (1997) introduced a motivational component to network formation among individuals. According to the prevailing theory in social psychology, motivation to affiliate takes the form of arousal. When two people affiliate there is an arousal reduction, which is mediated by the endorphin mechanism. Once an affiliation link has been established, a set of four feedback loops form for arousal and arousal reduction within and across the two participants.

In a high-entropy social milieu, the formation of links begins with interactions among people, which are analogous to the interactions among atoms, at least in part (Galam, 1996). In this line of thinking, something akin to temperature occurs in the social environment that causes the people within to bounce around faster as molecules would do within a container. Random contacts occur until a drop in entropy occurs when the right combinations of people or atoms are found, which is when a link occurs. If the process is allowed to continue further, social structures evolve that support the emergence of primary leaders and technical leaders in either real-time or network contexts (Guastello, 2007; Hazy, 2008).

Smith and Stevens observed, furthermore, that unlike the molecule analogy, additional people can join the aggregate until a group is formed. The extent to which the aggregate grows might depend on context variables such as the complexity of the task or activities involved, the type of relationship between individuals in the groups, and whether competitive exclusion is part of the group rules of the relationship. Additionally, individual differences in the desired number of friends can affect the growth of networks in terms of size or density (Zeggelink et al., 1997).

Nonhierarchical Structures

There is a conventional tendency to equate hierarchical systems with the formal lines of authority associated with organizational charts. Authority is not a necessary assumption, however. Rather in the NDS context, a hierarchical relationship between two system components is unidirectional in its influence. When more than two components are involved, there is a one-to-many relationship as well. The latter feature is also known as *span of control* in organizational behavior and *fan-out* in human–robot interactions. Fan-out is explored later in this chapter. Nonhierarchical structures are considered next because they have a strong yet unexplored potential for making multiple PMSs more efficient. Similarly, these structures have been known for quite some time.

The wheel-shaped network in Figure 13.7a is hierarchical because of the asymmetry in communication flows among group members. The transition from the wheel to the pentagram is only the first of many possible nonhierarchical arrangements. O'Neill (1984) developed a set of nonhierarchical structures that contained some important properties: homogeneous information flow throughout the network, a minimum number of meetings to convey the information, with a minimum number of people in each meeting. Figure 13.7c and 13.7d show two examples from his set of [*n*, 3, 2] networks. In [*n*, 3, 2], an information packet flows across the network consisting of *n* people with a minimum of three meetings, each of which contains two people. This series can accommodate 4 to 20 people and allow for incremental growth. If more people are added to the network, another configuration can be formed that still maintains the [*n*, 3, 2] characteristic.

There is also a series of nonhierarchical models that require meetings of three people (not shown). The set of [*n*, *m*, 3] networks can accommodate 7 to 30 people. The number of meetings differs for different network members, but each meeting consists of three people. The structures are modular, meaning that two or more structures can be combined to form another nonhierarchical network. For instance, two [5, 3, 2] networks become a [10, 3, 2] network. Eventually, there is an end to these net structures. Nonetheless, the series is flex-

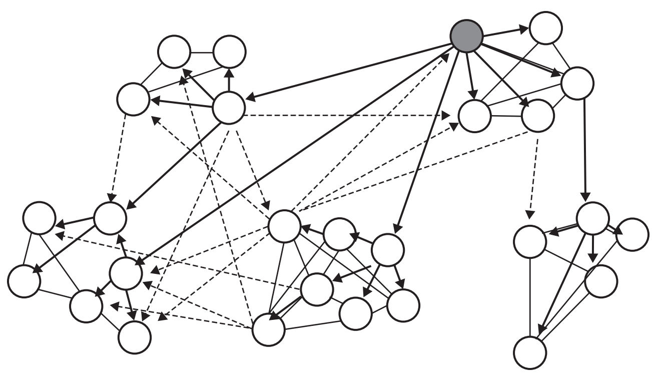

FIGURE 13.7
Clustering in a small world network. (Reprinted from Hazy, J. K., *Nonlinear Dynamics, Psychology, and Life Sciences, 12*, 294, 2008, with permission of the Society for Chaos Theory in Psychology & Life Sciences.)

ible with regard to the number of people involved, preferred meeting size, and modular organization. Thus, hierarchical structures can be minimized.

Centrality

Freeman (1979) noted that many mathematical models for the centrality of a point within a network had been offered over the previous two decades. After eliminating redundant and overly complex varieties, three concepts of centrality survived that were based on principles of degree, betweenness, and closeness.

Degree is the depth to which a point is interwoven with other points in the network. It is a function of the number of adjacent points to any particular point, or $a(p_i, p_k)$, which is relative to the total number of possible links that a network could have given the size of the network, n. Thus, centrality degree, CD, for a point (p_k) is (Freeman, 1979, p. 221):

$$CD = \left[\sum_{i=1} a(p_i, p_k)\right] \Big/ (n-1), \tag{13.1}$$

where n is the total number of nodes in the network. Note that in many discussions of degree, degree is characterized simply by the number of links; the denominator is ignored either because it is equal for all nodes by definition, or because the population size is not known.

Betweenness is the extent to which a point gets in between any two other communicating points. A betweenness indicator, CB, for a point p_k is (Freeman, 1979, p. 223):

$$CB = \sum_{i<j}\sum (g_{ij}(p_k)/g_{ij}), \tag{13.2}$$

where $g_{ij}(p_k)$ is the number of geodesics linking p_i and p_j that contain p_k, and g_{ij} is the total number of geodesics in the network.

Closeness is the extent to which a point enjoys the minimum number of points between itself and each other point in the network. It is actually an inverse function of *decentrality*, such that $d(p_i, p_k)$ is the number of edges between a point p_k and all other points. The closeness indicator, CC, for p_k is (Freeman, 1979, p. 225):

$$CC = (n-1) \Big/ \sum_{i=1} d(p_i, p_k). \tag{13.3}$$

Small Worlds

In a *random* graph, n nodes have at least one connection with another node to start. The connections occur on a random basis, meaning that there is no assumption that any node should have a preference for any other node. If the number of links between nodes is allowed to increase, again on a random basis, the network will eventually reach a phase shift in which there is suddenly a large cluster of interconnected nodes within the broader network (Newman et al., 2006). Clusters with higher betweenness centrality would be known as *hubs*, particularly if the overall network size is large, in which case there is greater differentiation between the cluster and the noncluster nodes.

In a regular graph, all nodes have equal degree. Examples of regular networks might organize the nodes around a circle. Each node can connect to the node to its left and right and to the second node down in either direction. Thus, in this example, each node would have a degree of 4, and the path between any node to any other node is equally long or short.

Next, take a regular graph and inject a couple random links that do not exist already and the result is a *small world* (Watts & Strogatz, 1998). Now the paths between some nodes are shorter than other paths. Interactions between nodes will follow the shortest path in the real world; thus, some nodes will become part of more interactions than other nodes. Furthermore, the frequency distribution of degree for members of the network follows a power law distribution; a small number of nodes have many connections, while a large number have fewer connections. The presence of a power law distribution makes another connection to fractal structure, which can in principle be compared across networks or network configurations.

The presence of small world structure has been observed in a variety of contexts such as small organisms where the neural network is completely known, high-voltage power transmissions, and movie actors. In each case, one needs to settle on a definition of what it means to be connected. In one context, it might mean having made a movie with another actor. In scientific environments, it would mean whether a journal article cited another journal article, or whether people coauthored something together. In another context connectedness would mean whether someone knew someone else well enough to have started a conversation or asked a small favor.

The small world phenomenon enchanted story writers and social scientists long before the formalities of small world networks were developed. The psychological studies that began with Milgram (1967; Travers & Milgram, 1969) in combination with later developments by network theorists led to the principle of *six degrees*: It is possible for anyone to reach anyone else by a median of 6°. The trick is to find the six degrees that comprise the shortest path. The idea led to a parlor game of "Six Degrees: How Can I Trace My Connections to Somebody Famous?"

Small world networks have some interesting properties that are relevant to system science. If there is a random attack on a network (e.g., a computer virus), small worlds are more resilient than random networks by virtue of their ability to find alternative pathways to maintaining all the necessary communications. However, diseases (or computer viruses) travel faster in small worlds if they hit a hub. It is also possible to "rewire" networks to balance the merits of small worlds and random configurations.

The power law distribution of links can be observed as chunkiness in the network layout. Figure 13.7 shows an example from Hazy (2008). Watts and Strogatz (1998) developed a clustering coefficient, which was later refined as (Newman et al., 2006, p. 288):

$$C = 6T/k, \tag{13.4}$$

where T is the number of triangles on a graph and k is the number of paths of length 2. A triangle consists of any possible triple of nodes A, B, C. A path of length 2 exists if, for any triple A, B, C, the edges A → B and B → C are in the network.

Hazy (2008) observed, "Increased density [indicated by the dotted lines] carries risk of interaction catastrophe; leadership [can manage] interaction risk with agents through partitioning" (p. 294). Interaction catastrophe is similar to the complexity catastrophe and work overload related to all the communications. As a result, "The system is challenged in decision making, control, structuring, with respect to centralization versus decentralization, [etc.]" (p. 294). Some of the interactions arise from the agents receiving different stimuli from the environment, many of which could be ambiguous in nature. The disambiguation process that occurs during communications should result in a unified and unambiguous response to the situation. The development of mental models and situation awareness is implied here, and these points are expanded further below in the context of collective intelligence and team coordination. Newman et al. (2006) also noted the potential for nodes to synchronize, lock up, or phase shift as a result of oscillating input through weak ties.

Collective Intelligence

The cover image on Kelley's (1994) book *Out of Control* depicted a stylized office building designed as a grid with large windows and huge bees flying in and out. The implicit message was that work processes resemble a swarm of insects more closely than a machine that is designed for exact reproductions of objects. The queen bee does not give orders to the worker bees. The worker bees figure out their own work routines, which we observe as a swarm, based on one-on-one interactions with each other.

The concept of collective intelligence originated with studies of social insects, such as ants. An ant colony self-organizes into jobs such as foraging for food and nest maintenance. If there is a labor shortage in one of those areas, ant personnel are diverted from another task. Food foraging follows a pattern whereby the foraging patrol heads out in one direction on the first foray, then systematically adjusts its course by a few degrees for the next successive forays. No one ant actually knows the entire plan, but the collective executes it well (Sulis, 1997).

The concept crossed over to human cognitive phenomena when it became apparent that decentralized networks of people produce ideas, plans, and coordinated actions without being present in the same physical location simultaneously. The interaction among people is greatly facilitated by computer-based systems such as standard e-mail, listservers, and web-based technologies (Bockelman, Morrow, & Fiore; Guastello & Philippe, 1997). The growth of virtual communities gravitates to an attractor that represents a stable population. The study of collective communication patterns in technology-driven systems, which often facilitate easy tracking specific statements and their temporal sequencing, has led to a rethinking of human interactions in real time as well (Gureckis & Goldstone, 2006). The same phenomena are sometimes known as *distributed cognition.*

One should bear in mind that the boundaries usually associated with "organization" are semipermeable, meaning that a great deal of information flows across organizational boundaries and might not be centralized within an organization at all. This phenomenon together with analogies to insect colonies was the underlying theme in Kelly (1994). With decentralized or network-based communication and decision patterns, the notion of controlling a human system necessarily changes. Consistent with the idea behind ant colonies, the top-down control that is usually inherent in organizational structures simply does not operate well any longer: Events self-organize from the bottom up. The next section of this chapter considers some selected themes: basic principles of collective intelligence, creative problem solving, team coordination, and the learning organization.

An ant colony displays some highly organized activities such as foraging, nest building, maintenance, and travel. At the same time, each ant does not have a big plan in its little head. Rather, each ant is equipped with elementary schemata that synchronize with those of other ants when larger scale events occur. Sulis (2009) identified several principles of ant collective intelligence from which it is possible to extrapolate analogous functions in human systems. The first two are *interactive determinism* and *self-organization*, which were described in general systems form already: The interaction among individuals gives rise to the final collective result. The final result stabilizes in interesting ways and without external intervention; in other words, the queen ant or bee is not barking (buzzing) orders to the others. The stabilization of a collective action pattern is a *phase transition*.

Recalling the counterpoint made earlier about embedded and embodied cognition, the embodied portion operates automatically, assimilating nuances in the environment. The embedded portion is aware of the nuances in the environment and permits adaptations or accommodations to be made. Environmental nuances, nonetheless, have an impact on the final collective result; the phenomenon is known as *stochastic determinism*.

Probability structures that underlie collective outcomes remain stable over time, however, and are regarded as *nondispersive*. They remain as such until a substantial adaptation is needed such as some regions of the environment become favored or disfavored. This *broken ergodicity* occurs at the collective level. Similar disruptions occur at the individual level as an experience of one agent impacts the behavior of another, thereby amplifying the feedback to a third. With enough uncontrolled fluctuation, the foraging path can shift suddenly. Hence, *broken symmetry* is possible as well.

Some further principles are likely to take different forms in humans in contrast to insect contexts. One is *salience*: The environmental cues that could induce broken symmetry are likely to work to the extent that they are sufficiently visible compared with other cues. In human contexts salience is complicated by *meaning*, which can be operationally defined as the connection to other ideas, personal intentions, and system goals (Kohonen, 1989). Ants do not appear to express much variety in personal intention, but humans do so regularly. Humans and human systems often have several competing goals. That brings us to the last element of Figure 13.5, the shared system of meaning, which is essential for communication.

Also important is that computational experiments assume that individual agents are destined to interact. Sometimes that is the case, of course, but people also reserve the choice to interact or not. Whatever rules they invoke to decide whether to interact are likely to play out in emergent patterns of social organization eventually (Sulis, 2008, 2009; Trofimova, 2002).

Asynchronous Problem Solving in E-Communication

The comparative study of real-time and electronic problem-solving media has uncovered four new group phenomena. One is the *unblocking effect*: In real-time brainstorming groups, some participants have difficulty getting a word in edgewise, particularly if the flow of ideas is heading in another direction. Electronic media, in which communication is asynchronous, give individuals the time they need to consolidate and phrase their thoughts and post them to the group.

Another special effect is the *critical mass*. A review of group productivity studies showed that group brainstorming sessions produce more ideas than are produced by the most competent individuals in about half the occasions where such comparisons were reported. At other times, the group is not more productive than the most productive individual. Some of the disappointing behavior of groups can be explained by whether tasks and rewards were defined for groups or for individuals; social loafing was another explanation (Shepperd, 1993). Dennis and Valacich (1993) found, however, that a critical number of group members is needed to gain an advantage for groups using computer media.

The third major phenomenon pertains to channels of communication and filtering. The moderating effect of channels on the resource–performance relationship was noted earlier. Walther (1996) noted, however, that the filtering process that occurs in a virtual group (real people whose communication and work products occur solely through a computer network) has an advantage. According to Walther, the filtering of social content was overstated in several earlier studies because friendships and romances do form in virtual media. Rather, the anticipation of continued interaction predisposes individuals to act in a friendly and cooperative manner. Unlike what often occurs in contrived experiments, participants can prepare outgoing messages without the stress of interpersonal interaction (Walther, 1996, p. 24), and because of the time asynchrony, task and interpersonal messages can become disentrained. It is also noteworthy, however, that the social remarks increase in cases where participants are not close to their maximum available communication time (Walther, Anderson, & Park, 1994).

Sensemaking and Situation Awareness

Sensemaking in emergency response and other contexts has also been studied as *metacognition*, which focuses on the development of mental models by groups and teams. Mental models play a significant role in group coordination, as described later in this chapter, and the experimental paradigms are different with regard to whether the experimenters assumed that a fixed mental model should be shared at the outset of the task sequence, or whether the mental model should be allowed to develop on the fly, which it appears to do in many cases in the real world. Whereas coordinated groups interact and perform tasks on a rule-based system, metacognition is the process of finding those rules for the first time, which involves utilizing acquired knowledge and responding to the demands of new situations (Fiore et al., 2010).

Sensemaking is closely tied to situational awareness; the group has a better awareness, especially when the rules for future actions are concerned, if it can make sense of the situation. In response, some new software designs that contain multimedia person–machine interfaces for multiple users have been designed for different contexts. Figure 13.8 is an

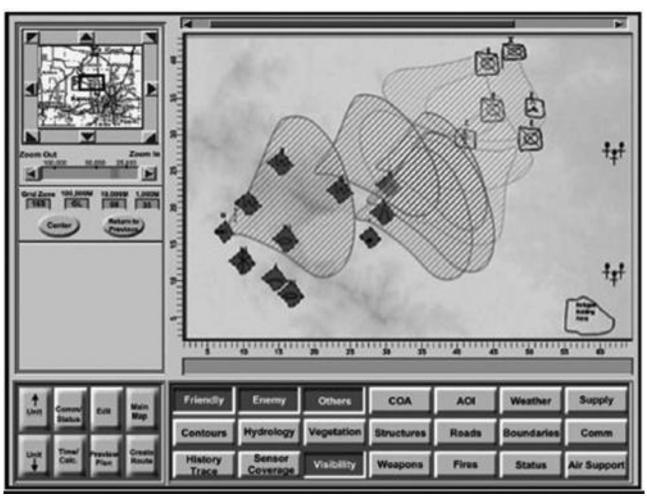

FIGURE 13.8
Example of multimodal interface for situation awareness. (Reprinted from Riley, J. M., Endsley, M. R., Boldstad, C. A., & Cuevas, H. M., *Ergonomics, 49*, 1139–1153, 2006. With permission of Taylor & Francis.)

example of an interface (Riley, Endsley, Boldstad, & Cuevas, 2006) that was designed for multiple military operatives to encode, store, and share information about the movement of troops and equipment. The system also allows the operators to plan tactics from dispersed locations. Agents' real positions are represented by symbols that are neatly drawn by machine. Hypothetical new agent positions that the team is discussing can be entered by using symbols that look as though they were hand drawn with a crayon.

Network Growth

The fourth special effect of e-communications pertains to the natural formation of large discussion groups on the Internet. The formal analysis of network growth is known as *percolation*. The following phenomena are relatively simplistic but pertinent examples. Networks can consist of interacting individuals, but also of larger social units such as organizations wherein specific people are usually responsible for maintaining the connection between one business organization and another. Thus, a node in a network might not be a single agent, but a member of other networks that are implicitly combined with the network under analysis.

Extensive networks of weak linkages are often the best sources of news and technical information (Constant, Sproull, & Kiesler, 1997). They can expand more easily than dense networks because the threshold between being in and out of the network is relatively low. Networks with weak ties have another advantage in that they can provide new information more readily than dense networks; dense networks share information quickly, but the store of content can be expended quickly as well and needs to be refueled by input from other contacts.

Guastello and Philippe (1997) argued, furthermore, that the dynamics for joining a network could be expressed as a *bandwagon* game. Bandwagon is one of several theoretical coordination scenarios that are considered later in this chapter. As a general rule, a bandwagon occurs when the cost of joining an activity decreases as more participants join the activity. In the case of network growth, the two participants would be a specific person who is contemplating joining the group (or network), and the rest of the group would be conceptualized as the other unitary player. Participation by either the individual or collective may be low, moderate, or intense. The extent to which the information flow is positive for an individual, relative to the cost of involvement, predicts increases in network members. Information flow, assuming that it is relevant to individuals' needs, becomes greater and more varied when networks are larger (Guastello, 2002).

The bifurcation Equation 13.5 describes the dynamics of how some participants could enter a network and not grow long-standing links to other network members, while other participants could drop out of the network eventually:

$$z_2 = \theta_1 z_1 \exp[\theta_2 z_1] + \theta_3. \tag{13.5}$$

The strength of a bifurcation effect is denoted by the size of θ_1. Equation 13.5 is essentially a May–Oster population model, where θ_1 is comparable to a birth rate for a population of organisms. Equation 13.5 was shown to be a close fit for the development of a virtual organization composed of individual members (Guastello & Philippe, 1997). Within this structure, it is possible for a network population to reach an asymptotically stable size in which the linkages would saturate or dissipate and die out.

A third model of potential relevance to the growth of networks is the cusp catastrophe model (Guastello, 2002). In this situation, the linkages consisted of joint patents among organizations in the telecommunications industries and the dependent measure was the number of links between an organization and all other organizations in the set. The two stable states would be whether the organization stabilizes within the network or drops out of it. The two control variables were whether the organizations that were contributing the data points came from the telecommunications, semiconductor, or computer–manufacturing industries. Different dynamics were apparently occurring in those industries that promoted faster network growth, but the specifics cannot be reconstructed at this time if they were ever really known.

A new phenomenon has come to the foreground, however, that could affect the broader landscape of person–machine interactions and system design. *Coopertition* is the behavior pattern whereby firms compete in some markets but cooperate to build hybrid products for another. Pathak, Pokharel, and Mahadevan (2013) found in a simulation study that coopertition is likely to be a stable and productive relationship if the business environment favors building on strength, and unlikely to sustain if the environment favors exploring new ideas and strategies. It would follow, therefore, that some people in the research and development fields would need to adopt different strategies for communicating with the same nodes in their networks. Task switching phenomena thus move into a different level of scale.

Dynamics of Feedback Loops

Several researchers have suggested that electronic media are catalyzing the self-organization of a collective intelligence (Guastello, 1998, 2002; Guastello & Philippe, 1997; Mayer-Kress, 1994; Smith, 1994). If that were true, it should be possible to locate chaos dynamics prior to such self-organization. Indeed, Contractor and Siebold (1993) used simulation techniques to determine "boundary conditions under which groups reproduce, sustain, and change existing structures" (p. 535). They noted further that such structures apparently change in a punctuated equilibrium fashion. Structures would be produced in part by the group's expertise in the use of the decision support system, and the group's perception and awareness of norms. *Structures* implied conversational behaviors of very general varieties. Some common examples of norms were quality of the thought product, quantity of production or apparent effort, civility of social interaction, and use of gatekeeping initiatives when participants' behaviors exceed a desired norm.

Their simulations were based on the quadratic logistic map function such that change in the time spent contributing (order parameter) was a function of the expertise contributed by the other participants and the norms. The quadratic function was capable of producing constant, cyclic, and chaotic activity clusters. The simulation results showed different phase portraits for participation time depending on prior expertise with the decision support system, presence of prior norms, and intensity of training between the sessions of activity.

Contractor and Siebold's (1993) findings were corroborated to a great extent by a separate line of research that showed that the quantity of individuals' output did change over time throughout a discussion according to a modified logistic map function (Guastello, 1995). In the latter example, the dependent measure was the length of a person's response in a conversation on a listserver. There were two control parameters. One control parameter was cumulative elapsed response, which was the quantity of contribution by other participants in the conversation that occurred in between consecutive responses by a single person. The second control parameter was personal response style, which was essentially a garbage can variable containing individual differences in output levels that were not otherwise accounted for. Both control variables were responsible for the bifurcation effect.

There was, furthermore, a global linear trend whereby the cumulative elapsed response at Time 1 predicted an increased output at Time 2.

In another effort, conversations among three small groups of problem solvers (with real, rather than simulated, problem objectives) were studied over a period of approximately 100 days (Guastello, 1998). The medium was electronic with communication in ASCII text through specialized conferencing software. The amount of total output from each group per day was recorded, and fluctuations over time were analyzed for dynamic content. (Note the shift in perspective here from an independent measure that is based on individual behavior to one that is based on group output.) Density, the dependent measure, was defined as the number of responses generated by the participants in a 4-day period. Group 1 contained three discussants, Group 2 contained seven, and Group 3 contained eight.

It was determined that the groups' production over time was chaotic, and simultaneously driven by a second periodic dynamic. Equation 13.6 was obtained for density over time:

$$Z_z = e^{(0.25_z)} + 0.43A - 0.26C - 0.34, \tag{13.6}$$

where z_t was density at two consecutive observations in time, A was the number of active threads during the time interval of z_1, and C represented differences among the conferences. The Lyapunov exponent (regression weight in the exponent to e in Equation 13.2) was positive, indicating chaotic behavior and a dimensionality of 1.28. Conference differences were measured as a three-value effect-coded variable indicating the overall quantity of response from each conference; some discussion forums, which are combinations of topics and people, generate greater quantities of discussion than others.

The periodic function was the number of discussion threads, A, that were operating on a given day. The greater the number of threads, the greater the total group output was. Equation 13.7 was obtained for the change in the number of subtopics or threads over time:

$$A_2 = 0.75A_1e^{(-0.36A1)} + 0.33. \tag{13.7}$$

The Lyapunov exponent was negative, indicating that the function was not chaotic; the dimensionality of the time series was 1.70.

The conclusion from the study was that there were two important dynamics operating. One dynamic was a periodic driver, and the other was a chaotic slave. The driver was the unpacking of the conversation into separate threads. The chaotic output over time was observed when the discussion volume reached a critical level of intensity.

Other Temporal Dynamics

After a gap in time, the interest in temporal dynamics of problem solving interactions has resurfaced as a research topic in human factors, both with (Cooke et al., 2012; Gorman, Cooke, Amazeen, & Fousse, 2012; Karwowski, 2012; Russell, Funke, Knott, & Strang, 2012; Strang et al., 2012) and without (Fu & Pirolli, 2013; McComb, Kennedy, Perryman, Warner, & Letsky, 2010) explicit use of nonlinear dynamics. Information search strategies are often shaped by other agents in the network environment (Fu & Pirolli, 2013). Small world phenomena have some interesting implications for information searches. Ideas spread through networks, but not every idea "goes viral." Most network schemata follow "narrow tree" diffusion structures that end the diffusion much sooner than what would occur through a viral epidemic.

The research strategy for the problem solving process requires recording a conversation, then coding each utterance for features of interest. McComb et al. (2010) compared teams' processes in a planning task and how they arrived at their shared mental model. Those who worked face-to-face tended to explore more possibilities and options before arriving at their target model compared with those who worked in distributed fashion through conferencing software. The software-mediated groups took somewhat longer to arrive at their endpoint, but there was no actual difference between the two groups in quality of solution. The explanation for the differences in process basically boils down to talking being easier and faster than typing, all other things being equal.

The nonlinear approach (Gorman et al., 2012) involved coding conversations during a drone aircraft mission and analyzing the content for signs of adaptability versus rigidity and exploring conditions that could affect the recurrence of conversation patterns over time. Pattern formation, entropy levels, and what constitutes a repetition are predicated on the codes that one applies to the data, however (Guastello, 2000); it is possible to code one conversation according to two different protocols and arrive at two different pictures of the nonlinear dynamics involved.

Learning Organizations

Allen (2009) reported a simulation study that examined four learning strategies and the relative effectiveness of each: (a) Darwinian learning, where the organizations start with a random strategy, organizations with good strategies survive, and organizations with poor strategies go bankrupt and are replaced by new organizations with random strategies; (b) imitate the winner, which means that organizations copy others in the environment that have apparently functional strategies; (c) trial and error, where organizations explore possible strategies, try some, observe results, reevaluate and perhaps try something else while continuing to consider new options; (d) mixed strategies, where all of the previous three exist in the organizational environment. Results showed that Darwinian learning produced the worst results for the industry as a whole with the largest proportion of bankruptcies. Imitating the winner worked much better, although it was subject to large fluctuations in profitability levels; it involved imitating the winner's limitations too. Overall the greatest success was recorded for the trial and error strategy where agents learn from their mistakes and continually seek out possible improvements by exploring what he characterized as their landscape of opportunities.

Next, consider the situation where the organization acts a whole unit. Any teams or workgroups that are involved could be coordinated within a larger hierarchy of activities. The notion of a learning organization became a fashionable view of an organization shortly before the notion of the organization as a complex adaptive system took hold (Seo, Putnam & Bartunek, 2004). In its earlier manifestations, learning organizations were those that had evolved processes or structures analogous to individual perception, cognition, memory, and adaptation processes. In later studies, learning processes in organizations are seen to promote self-organization of dominant strategies or schemata from the bottom up. Individuals and teams adopt processes that produce ideas, schemata, mental models, and meanings that are eventually shared with other teams until some become dominant enough in the organization to shape new schemata for newcomers or new responses to new challenges (Van de Ven, 2004).

The perception, situation awareness, or sensemaking processes in organizational contexts require information exchange networks that extend outside the organization to other organizations in the same industry, other organizations in different industries, and of course customers. Van de Ven gave an example of a successful use of bottom-up development of wind turbine technology. Danish industries started with relatively simple technology, and through close interaction with customers and their needs, shaped a premier technology that is financially successful for the organizations involved. Would-be competitors from the United States, however, took an isolationist strategy and attempted to leapfrog the stages of development by developing an advanced technology quickly. They maintained little communication with customers and were generally unsuccessful in their efforts.

Group Coordination

Systems composed of multiple person–machine systems do require coordination to function optimally. Although it is tempting to consider the full range of phenomena in group dynamics here, most of what is known in psychology can be considered generic in the sense that the principles have not been specifically implicated in person–machine system dynamics. The interested reader should see Zander (1994) and Hackman (1990) for general principles and specific examples in work organizations. Coordination is considered here because of its relevance to person–machine systems and because it has some known implications for complex adaptive systems.

Coordination occurs when two or more people do the same task or complimentary tasks simultaneously. Numerous examples can be found in manufacturing work teams, hospital emergency rooms, military operations, sports teams, and musical or theatrical performances. Coordination has been measured in empirical studies as the time delay between a group member's action and another member's contingent action (Brannick, Roach, & Salas, 1993), the quality of communication between members of a group (Daily, 1980; Gittell, 2000), or the performance of a group in a task wherein members must take correct actions in a correct sequence (Guastello, 2002; Guastello, Bock, Caldwell, & Bond, 2005; Guastello & Guastello, 1998). There are several theoretical aspects to coordination dynamics that emanate from implicit learning theory, the theory of shared mental models, game theory, and nonlinear dynamics.

Implicit Learning

Groups acquire coordination as an implicit learning process, which is a form of learning that takes place while the group members are explicitly learning to do something else (Frensch & Runger, 2003; Seger, 1994). The group members might be less attentive to the implicit learning process than they are to explicit learning goals. In this case, the implicit learning objective is to learn how to respond correctly to actions and nonverbal signals given by other group members. A group performance experiment demonstrated, however, that coordination that was acquired while learning one task transferred to subsequent tasks where the performance goal that required coordination was different (Guastello & Guastello, 1998).

Although implicit learning is usually studied as an individual process, it can take the form of a group process also. The group process involves coadaptation and mutual entrainment among the participants, and thus qualifies as a self-organizing process. There is now sufficient evidence from neurological, individual behavioral, and group behavior levels of analysis that learning processes are self-organizing processes. The initial phases of behavior are chaotic in performance or output, with asymptotic learning curves taking over as the process becomes complete (Guastello et al., 2005).

Shared Mental Models

Studies indicate that the coordination among workgroup members is greatly enhanced if the members have a shared mental model of their tasks, procedures, and group processes (Banks & Millward, 2000; Cannon-Bowers, Salas, & Converse, 1990, 1993; Druskat & Pescosolido, 2002; Stout, Cannon-Bowers, Salas, & Milanovich, 1999). Shared mental models may be induced by cross-training the group members in each others' roles, or by discussions and presentations of groups' task models (Marks, Sabella, Burke, & Zaccaro, 2002; Matthieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). The sharing of mental models also requires consistent organizational support if the group is not a freestanding group (Druskat & Pescosolido, 2002).

Despite the advantages of shared mental models, however, NDS studies on coordination that involved general systems and animal models did not require discussion groups, cross-training, verbal communication, or leadership for coordination to occur. This does not mean that something akin to a shared mental model did not exist. It would mean, however, that group members had to arrive at the shared mental model through their own individual observations. Game theory would suggest that any mental models that are relevant to coordinated action can be reduced to rules, options, utilities, and sometimes trust in the rationality of the other players.

Role of Verbalization

Clear, precise, and prompt verbal communication produces a positive impact on coordination. In one example situation where the task participants were military aircrews working on simulated flights, a panel of raters evaluated the crewmembers on communication and other variables (Brannick, Prince, Prince, & Salas, 1995). Ratings of overall task expertise were correlated with the ratings of communication effectiveness.

Xiao et al. (1995) found that breakdowns in coordination among hospital emergency room staff could be traced to two groups of faults. One group of explanations consisted of interruptions of task flow resulting from conflicting plans, inadequate delegation or diffusion of responsibility, or inadequate support. The other group of explanations was comprised of verbal communication faults. The apparently independent contribution of verbal communication motivated a fresh look at whether verbal communication is necessary or sufficient to produce good team coordination. This question was addressed in the studies on game theory, nonlinear dynamics, and coordination, which are considered next.

Game Theory

The central premise of game theory (von Neumann & Morgenstern, 1953) is that people make (economic) decisions that will optimize their own outcomes. The utility associated with an option usually depends on which options are selected by one or more of the other economic agents. Games can be classified as strictly competitive, strictly cooperative, and mixed competitive and cooperative. Prisoners' Dilemma is perhaps the best known game because of its mixed competitive and cooperative properties. Prisoners' Dilemma, Intersection, and Stag Hunt are all classifiable as coordination games because optimal solutions require that participants select the same options as the other participants at the same time. Intersection and Stag Hunt are two strictly cooperative games that have found relevance in complex workgroup dynamics. Bandwagon is another strictly cooperative game, and it was used as part of an explanation for the development of virtual communities. An important point here is that not all coordination dynamics in work units have the same dynamic.

Intersection Games

The intersection game is analogous to the decision process that occurs at a four-way stop intersection. Each car and driver must figure out which rule is in play for determining which car takes a turn proceeding through the intersection at a given time. The rules in the state's driver's manual might be in effect, or a different variation might be occurring. The drivers, nonetheless, must figure out the rule and apply it to their own behavior. Each driver must decide correctly or else negative utility occurs. Car wrecks are too expensive and dangerous for experimental purposes, so the basic principles of Intersection were captured in a card game (Guastello & Guastello, 1998; Guastello et al., 2005). The card game required four participants to play a sequence of cards in a particular order in order for the group to garner game points. Eight hands of cards constituted a round of the game. Different rounds of the game involved rules of equal or greater difficulty.

Game theory investigations often manipulate, or at least make use of, the utilities associated with the game's strategies. The utilities for the Intersection card game experiments remained constant as shown in Table 13.1. If a game is played in an iterative tournament, a strategy that eventually dominates (or evolutionarily stable state, ESS) will be close to the Nash equilibrium for a simple one-shot exchange (Maynard Smith, 1982). If we add some complexity to the possible game options and utilities and give up the assumption of strict competitiveness, however, the ESS cannot be guessed from any knowledge of the Nash equilibrium. Rather, the ESS becomes highly dependent on initial conditions as they pertain to options and utilities (Samuelson, 1997). Thus, ESS experiments are needed to identify real behavior patterns from games that emulate real-world problems and decisions.

The card game experiments (Guastello & Guastello, 1998; Guastello et al., 2005) showed the following results. Coordination is a nonlinear dynamic process. Coordinated behavior can be observed as chaos in its early stages of formation, and followed soon afterward as a system of self-organization. Coordination tasks of greater difficulty might reflect only the

TABLE 13.1
Matrix of Utilities for the Intersection Coordination Game

		Group of Three Players		
		All Coordinated	**Partially Coordinated**	**Not Coordinated**
Target	Coordinated	4,4	1,1	0,0
Player 4	Not	1,1	0,0	0,0

Source: "Coordination Learning in Stag Hunt Games with Application to Emergency Management," by Guastello, S. J., & Bond, R. W., Jr., 2004, *Nonlinear Dynamics, Psychology, and Life Sciences, 8*, p. 350. Copyright 2004 by the Society for Chaos Theory in Psychology & Life Sciences. Reprinted with permission.

chaos portion of the dynamic for some groups, but groups that are more adept may reflect both dynamics. The transition of players' game strategy choices from the first exchange to the ESS is a learning phenomenon. Learning phenomena themselves are nonlinear dynamic phenomena, as previously mentioned. The outcome of a learning set is typically observed as a fixed-point attractor in the learned behavior.

Nonverbal Communication

Human coordination, like animal coordination (Reynolds, 1987; Semovski, 2001), is fundamentally a nonverbal process. The nonverbal nature of coordination has been observed in the mainstay of game theory studies in economics, in which the experimenters are concerned that discussion will alter the perceived utilities of strategic choices. Coordination without verbalization was observed for Intersection card games as well. Verbalization can augment human coordination, however. This is not a trivial point; verbalization is neither necessary nor sufficient to produce coordination. In Intersection games it was shown that verbalization promoted sharper learning curves and better total performance. Verbalization did not compensate for the effects of changing personnel within a coordinated group, however (Guastello et al., 2005).

Shah and Brazael (2010) examined nonverbal communication in terms further with the objective of being able to isolate key behaviors that could be transferred to human–robot interaction; robots should interact with humans in a manner that humans feel comfortable. They considered team performance resulting from a mixture of verbal and nonverbal (usually gestural) behavior implicit and explicit commands or requests, and conditions of uncertainty, ambiguity, and time pressure, all of which contribute to stress. Overall 90% of the teams' communications were coordination messages; this point did not vary with the stress conditions. The teams produced *more* coordination messages, however, under time stress.

Minimum Entropy Principle

Not all teams experience a drop in performance or coordination when they have to work under greater time stress or uncertainty (Serfaty, Entin, & Deckert, 1993). Shah and Brazael (2010) remarked, "It appeared during these experiments that teams were changing their coordination and information-seeking strategies" (p. 236). This observation is consistent with the minimum entropy principle discussed in Chapter 6 in conjunction with individual workload.

Coordination itself reflects a reduction in degrees of freedom in group mental workload compared with relatively random parallel action. Experiments from other areas of the human resources literature show that greater task complexity (load) produces greater levels of coordination, and coordination efforts play a larger role in overall performance (Cheng, 1983; Shah & Brazael, 2010; Stevens & Campion, 1994).

Changes in Team Membership

Group coordination, when sufficiently established, can withstand small changes in group membership without affecting the level of coordination in the remaining group. Studies in laboratory cultures illustrated how a group norm can persist even when every member of the group of 10 members was eventually changed in a gradual sequence (Weick & Gilfillan, 1971). For groups of four players in an Intersection game, however, a sharp decline in coordination was observed when three players were replaced, leaving one original player. If only one or two players were changed, however, the group's performance was unaffected (Guastello et al., 2005).

Gorman, Amazeen, and Cooke (2010) found a different result when comparing groups that worked together in intact groups or those that involved swapping personnel from other groups that were already trained to do the same task. Their groups were three-person military reconnaissance teams. Groups that exchanged personnel actually performed better in terms of analyzing information about targets that was produced by unmanned aerial vehicles and finding targets. The margin of improvement for the mixed groups was traced to stability in their coordination dynamics. Intact teams tended to restrict their interpretations of information or search strategies, using a narrower repertoire of sensemaking strategies. They tended to react to unpredictable events as corrections to their schemata already in motion. Mixed groups, on the other hand, mutually adjusted their cognitive schemata and searched more thoroughly.

Group Size

Group size acts as a control parameter in ant collective intelligence (Sulis, 2009). A critical mass of ants is required to produce a sufficient momentum of interactions to get a nest-building project going. The same principle appears to resonate in research with human creative problem solving groups. Groups outperform the best-qualified individuals if the groups are large enough to produce a critical mass of ideas (Dennis & Valacich, 1993). Groups also have the potential for outperforming individuals because they can review the available information more thoroughly, retrieve errors more reliably, and rely on the skills and knowledge bases of more people to formulate a solution to a problem (Laughlin, 1996); here the critical mass of people probably varies with the complexity of the information-processing task.

Campion, Papper, and Medsker (1996) observed that groups need to be large enough to contain enough resources to get their work accomplished, but not be so large as to induce coordination difficulties. Social loafing is more likely in larger groups, however. Loafers or free riders would find utility in joining the group with the expectation that someone in the group would get the job done, and all members would benefit. Hierarchical group structures can introduce more opportunities for inefficiency (Guastello, 2002).

Using a research hypothesis concerning group size, one can assess the potential trade-off between critical mass and deficits in coordination (Guastello, 2010). If there is a group emergence effect at all, there would be an optimal group size associated with the teams' performance. If larger groups perform better, a group dynamic is operating that would be consistent with the critical mass principle. If midsize groups perform better, critical mass would be associated with the midsize groups and coordination loss with the larger groups. If smaller groups perform better, the group dynamics would reflect widespread loafing. If there were no effect for group size, then the teams' task was carried out by the most competent individuals; it would then be debatable whether the others were loafing or just not competent at the task.

Stag Hunt and Emergency Response

Response teams for emergencies, such as earthquakes, require coordinated and self-organized efforts for rescue, medical services, damage containment, and evacuation (Comfort, 1999a; Guastello, 2002; Koehler, 1999). The unfolding of events depends on initial conditions of specific time, location, and preparedness of the response teams. In a real emergency, groups of rescuers, health service providers, and food and shelter providers seem to materialize out of nowhere, according to the classic stories. Official response teams are typically overtaxed by the demand to fight fires, carry the injured to medical facilities, manage evacuation in an orderly fashion, work around self-organized evacuation attempts, stifle criminal behavior, and contain damage. Failure to make the right response at the right time could compromise any of the objectives.

Forest fires, earthquakes, and floods are certainly challenging. Although there are many unpredictable events, such as wind shifts, aftershocks, and the specific geography of the affected area, all those events behave just like natural disasters—with predictable indifference to the humans involved. The situation becomes more complex when the source of the disaster is a sentient being—a terrorist of sorts—whose goal is to wreak as much havoc and fear as possible and elude captors.

Comfort (1999b) analyzed the details of the emergency response efforts in 11 then-recent earthquake disasters. Up to 400 organizations could be involved between the efforts of international government, national government, local government, and civilian nongovernmental organizations. Coordination of efforts could be assessed in terms of information search strategies, information exchange, and organizational learning during the emergency response process, and the effectiveness of actions taken. It was noted that organizational learning enhances lower level action, thus making reliance on hierarchies less critical.

The relative success of emergency response efforts depended on the articulation of common meaning, trust between leaders and citizens, resonance between the responding systems and the physical environment, and the sufficiency of resources. The least effective adaptive responses were severely hampered by communication outages with no backup systems such as satellite-based technical systems. Other critical sources of dysfunction were traced to political problems such as an ongoing civil war, which placed political objectives at crosscurrents with the common good.

Comfort (1999b) characterized any form of adaptive response as an edge-of-chaos phenomenon. There are rapid swings between periods of instability and stability, even in the best of conditions. Most adaptive final outcomes depended greatly on good sociotechnical infrastructures. Preparedness to reduce risk before the disaster allowed the community to minimize damages and bounce back effectively after the disaster.

The coordination requirements for emergency response strongly resemble the structure of a Stag Hunt game (Guastello & Bond, 2004). In the classic Stag Hunt story, all hunters within a group must decide whether they want to join the group to hunt a stag or run off to hunt rabbits alone. If enough people join the collective and exert enough effort, the stag will be caught in theory. The rewards are then distributed among the stag-hunting agents.

In classic disasters such as earthquakes, some people evacuate immediately (hunt rabbits) and take care of their own families and property first (hunt rabbits). Others join collectives to locate and help survivors (hunt stag). There are also people whose form of employment is to engage in the emergency response (hunt stag). In a response to an attack, the relevant military and governmental agencies may pursue an orchestrated agenda (hunt stag) or follow disjointed and uncoordinated plans (hunt rabbits). Table 13.2 lists some differences, however, between a theoretical game of stag hunt, a real stag hunt, emergency response to a natural disaster, and emergency response to a sentient attacker. In a theoretical version of the game, the two basic options are to hunt stag or rabbits; these options are the same for the hunters in real stag hunts, natural disaster management, and attack management as well.

TABLE 13.2

Comparison Between the Stag Hunt Game and Some of Its Real-World Applications

	Stag Hunt Game	Real Stag Hunt	Natural Disaster	Sentient Attacker
Hunters' options	Stag vs. rabbits	Stag vs. rabbits	Stag vs. rabbits	Stag vs. rabbits
Stag's options	Not part of process	Evasive actions	Trivial strategy	Evade or attack
Utility structure	Enough effort leads to reward	Enough effort leads to reward unless stag is smarter	Enough effort leads to reward unless disaster is stronger	Enough effort may lead to reward but attacker often competes effectively
Evolutionary properties	Subgame perfect, or manipulated by experimenter	Stags may get smarter over time, lowering hunters' odds	Hunters' odds may decay over time; hunters' coordination may deteriorate if uncertainty increases	Attacker rewrites the game and affects odds; hunters' coordination may deteriorate if uncertainty increases

Source: "Coordination Learning in Stag Hunt Games with Application to Emergency Management," by Guastello, S. J., & Bond, R. W., Jr., 2004, *Nonlinear Dynamics, Psychology, and Life Sciences, 8*, p. 350.

Some differences among the situations can be observed when we consider the role of the stag. In a theoretical game, the stag does not contribute to the options or outcomes for the hunter. Although a real stag does not attack, it can take evasive action or stand and fight. If it survives repeated exchanges with the hunters, it might get smarter, thus decreasing the odds of success for subsequent hunters who do not also get smarter. Herd animals learn to perceive threats and learn evasive maneuvers after their herdmates have fallen prey to a new predator.

In the natural disaster, the disaster plays an active role according to what is known in game theory as a *trivial strategy*: The disaster does exactly what it was going to do—in accordance with the pertinent laws of physics and geography—without any regard for the actions of the hunters. For the hunters, success is a matter of degree and a matter of their strength and cooperation against the strength of natural forces. A trivial game is not necessarily trivial in terms of the importance of outcomes to the gaming parties. Importantly, the presence of a trivial strategy characterizes the game as competitive, with respect to the interactions between the humans and the disaster, even though the interactions among the humans may be strictly cooperative.

In a confrontation between the humans and a sentient attacker, the attacker by definition is utilizing offensive tactics in addition to evasion and defense. The competitive interaction is clearly nontrivial. Unlike the stag hunt, this type of stag occasionally eats the hunters. As part of their overall strategies, the competing parties try to expand their own options, and limit the opponent's options and utilities.

The latter two situations are clearly not subgame perfect with respect to the humans' options and utilities. Subgame perfection means that it is possible to predict the outcome of an iterated game from the outcome of a single exchange; that would only be the case if the options and utilities facing the players were the same on each exchange. In a disaster situation, however, a turn of events on the part of the natural phenomenon or the attacker may shift the utilities associated with the maintenance of a group effort versus self-interested responses. As a result, it is not possible to predict the collective outcome on the basis of any equilibria (saddle points) inherent in a particular play of a game. *Oligarchic reaction* *functions* (Friedman & Samuelson, 1994), such as regulatory agencies or the United Nations in the case of international conflicts, can drastically alter the utilities and options for the competitors.

Some interesting evolutionary effects have been recorded for iterative Stag Hunt games. Players often gravitate to suboptimal choices (Carlsson & van Damme, 1993; Rankin, Van Huyck, & Battalio, 2000). The suboptimal strategies are usually tied to the perceived uncertainty of attaining the win–win outcome. A large amount of the uncertainty arises from a player's estimate of whether the other players will decide in favor of a particular option. In the case of a team trying to coordinate against a hostile competitive agent, the full participation of the group can be seen to fluctuate over time, with downturns occurring when the competitor gains some advantage against the group (Guastello & Bond, 2004). Verbal communication does help the coordination efforts, but a trained group can anticipate the proper responses of teammates when a communication outage has occurred.

The ER group's results will be optimal if everyone in the group pulls together on each part of the job. A recent study (Guastello, 2010) examined the impact of team size and performance feedback on adaptation levels and performance of ER teams was examined. Performance was measured in an experimental dynamic decision task where ER teams of different sizes worked against an attacker who was trying to destroy a city. The complexity of teams' and attackers' adaptation strategies and the role of the opponents' performance were assessed by nonlinear regression analysis; the analysis featured the use of the Lyapunov exponent (a measure of turbulence in a time series) associated with the performance trends. The results showed that teams were more readily influenced by the attackers' performance than vice versa. The teams with 9 or 12 participants were more likely to prevail against the attacker compared with teams of 4 or 6 participants; only teams of 12 people were effective at dampening the adaptive responses of the attacker, however. In all cases, when the attackers scored points, the teams' performance on the next move declined. The attackers' performance patterns showed greater levels of turbulence, which was interpreted as adaptability, than the teams' overall.

Human–Robot Interaction

A robot is a semiautonomous machine that is capable of perception–action sequences and a modicum of autonomy of movement. Unlike its disembodied cousin, the autonomous agent, which is sometimes called a "bot," the robots under consideration here are capable of moving around the physical environment and taking action on it. Although humanoid embodiments are being developed gradually, other design paradigms might emulate the physical mobility of dogs, fish, or insects. Unmanned aircraft are another class of designs.

Three types of PMS configurations are shown in Figure 13.10. The simplest system (Figure 13.9a) contains a single person, a single robot, and a machine controller in the middle, with control and display links at both connections in the process. It is essentially the plan behind the operation of industrial robots, although large industrial systems can have many manufacturing robots connected to the machine controller. Importantly, there have been several reports of robot arms that fatally crushed workers against building structures when they swung into place suddenly (Chapter 10). The general solution was to build safe operation areas for the humans who could then control the robots from a suitably safe distance. Although some accidents could be eliminated in this fashion, the remote control solution does not address the potential for injury to other humans that might be in the work area and not visible to the controller (Gethman, 2001). Thus, an evaluation and

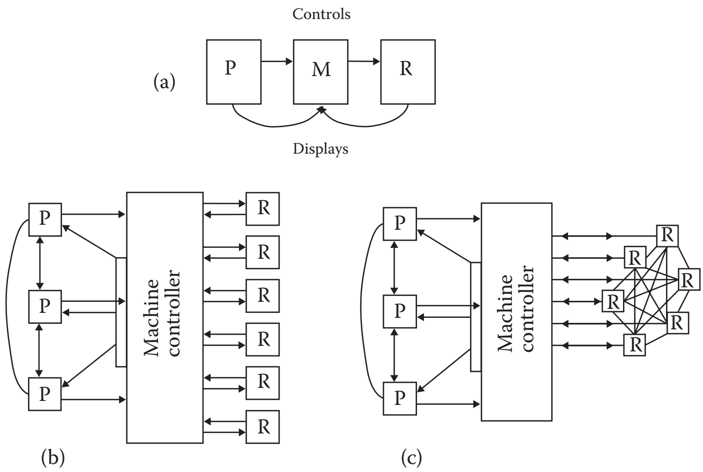

FIGURE 13.9
Configurations for human–robot interaction.

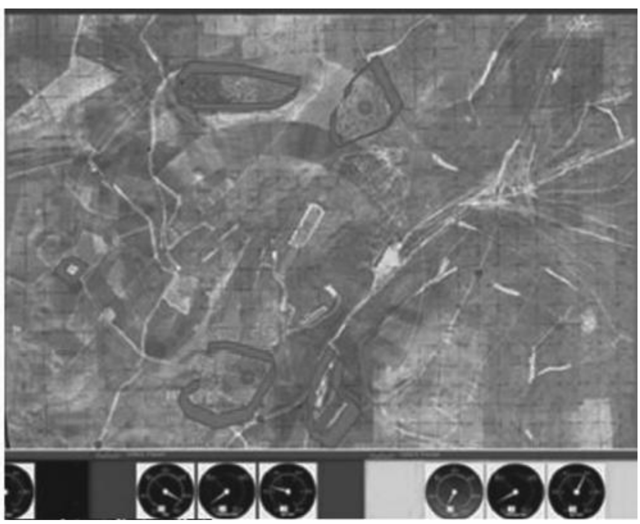

FIGURE 13.10
Software-enhanced video feed from drone aircraft. (Reprinted from Abedin, S., Lewis, M., Brooks, N., Owens, S., Scerri, P., & Sycara, K., *Proceedings of the Human Factors Society, 55*, 91–95, 2011, in the public domain.)

control of the workspace around the robot is essential for safe operation, and should be implied by the boxes in Figure 13.10 that represents the robots.

The connection between multiple robots to an operation controlled by one human is a system feature known as *fan-out*, or *span of control* in other contexts. The ideal level of fan-out is a standard system question in any system where automation is involved (Goodrich, 2013). With contemporary experimental systems, the fan-out ratio is more favorable for drone aircraft, where one person might control two or three drones, than it is for land-roving robots, where the fan-out ratio tends to run negative requiring two or more people to service one robot.

Multirobot systems require attention to each robot, so some attention switching is involved. As the number of robots increases management strategy shifts from management by consensus to management by exception. Instead of the human determining whether they agree with each move made by the robot, they shift attention toward simply fixing what is wrong (Goodrich, 2013). Finding the right level of autonomy is a design question; giving the robots higher levels of autonomy (more automatic processes) does not always make the job easier. Robots get into trouble, get stuck in loops, and need corrective interventions from the humans in order to continue operation. The performance of ground-based robots does not deteriorate gradually; problems tend to arise suddenly. Other robots should be able to function autonomously, or "tread water" until the human can finish one task before servicing another.

As a means of responding to the shocks to the human workload, the configuration in Figure 13.9b might be adopted. Several humans would observe the full display, and when problems arise the human operators would coordinate their actions to provide a complex response to a complex situation. Meanwhile, intelligent display designs continue to be developed. Figure 13.10 shows a processed video feed from a drone aircraft (Abedin et al., 2011, p. 92). The drone was equipped with synthetic vision, which it coupled with a map and GIS program. The human operator searched the display for targets. The artificially intelligent version of the display contained a preprocessor that automatically identified the targets sought by the humans and marked them accordingly. The evaluation results were that the humans missed fewer targets, which was the goal, but at the expense of situation awareness. The additional time required to process the image disrupted the sense of time between the collection of the image and the current location of the drone, which in turn compromised the operator's ability to act on the targets.

The humans depicted in Figure 13.9 are not always the only stakeholders in the video feed. One team of humans keeps the drones in operation, and the real system users are the dismounted ground troops. For instance, ground troops in the Israeli army regularly conduct operations searching for ground targets with the aid of a feed from a drone aircraft (Ophir-Arbelle, Oron-Gilad, Borowsky, & Parmet, 2013). The open questions are whether the search would be enhanced by adding an unmanned ground vehicle that provides another video feed from another angle, and how much of an increment of improvement can be expected when the search targets are people versus vehicles, or open-air environments around the target versus dense urban environments. The additional workload associated with the additional feed is substantial, so the size of the payoff is important. One visitor to a technology conference sponsored by the U.S. Department of Defense was quoted as saying:

> [E]veryone tacitly acknowledge[s] that sensemaking is a problem—cameras are sprouting like mushrooms from every conceivable mounting space in the hopes of giving the operator that magic missing viewpoint. However, it is all ad hoc; there is no understanding of what viewpoints are needed for which tasks or how to transition between them. So-called perceptual interfaces remain Neolithic (Ophir-Arbelle, 2013, p. 27).

A third type of configuration is the *swarm*, which is modeled after the swarm of insects, flock of birds, or school of fish. In this configuration (Figure 13.9c) each robot is an autonomous agent, which is itself complex with respect to cognitive and psychomotor components (Trianni, 2008). Self-organizing properties require information loops between the units, and the current challenge is to develop sensors and response structures that keep the cluster of robots functioning even if one of them should become impaired. A human controller is still involved, especially where the goals of the system's actions need to be defined for a given purpose, but the design questions are the following: How much control could be, or should be, allocated to the human "executive"? How much of the behavior of the system is going to be self-organized? How much broken symmetry can be tolerated? Several science fiction movies have centered on this theme, with nothing good happening to the humans.

Counterintuitively relative to humans, it is probably better not to have all information available to all robots. Instead, the system should be balkanized so that information is filtered into the subnetworks on a need-to-know basis. The alternative outcome would be one in which excessive information induces random reactions, or lock up, similar to other situations involving small world networks. "Structural holes" or elements of disconnectedness could be an asset in this context (Burt, 2004; Fu & Pirolli, 2013; Goodrich, 2013; Hazy, 2008).

Group Cognitive Workload

The measurement of group cognitive workload is an open question at this time. In the extant literature on group workload, researchers have attempted to integrate behavioral, subjective, and biometric data from individuals to ascertain a group-level measure of workload, but without a great deal of theoretical clarity or generalizability of results. Funke et al. (2012) suggested that NDS, with its constructs of synchronicity and self-organization (Strogatz, 2003), offers strong potential for solving the problem. Stevens et al. (2013) reported different patterns of group EEGs associated with different phases of a complex training task for submariners by first applying a set of symbolic forms to the EEGs, and then used recursion plots to compare actual versus randomized data series. Workload questions have not been answered yet from this research paradigm, however. It would appear that a viable rubric for measuring group workload should rely on principles of individual cognitive workload, coordination issues, and task switching between team members (Cooke et al., 2012).

Safety in Complex Systems

As technologies evolve, they involve more complex human interactions and higher degrees of potential entropy. The complexity of human interaction engenders multimodal controls and mode errors (Chapter 7) and the numerous dynamics discussed in this chapter thus far. According to Gethmann (2001), the proper approach for accident analysis and prevention is to ask the question, "How did it happen?" rather than "How could you?" The human may have made an error that appeared to be the cause of the accident, but combinations of system design and policies of management made the error possible. Although psychology has played an important role in system safety for decades, it must now play a more aggressive role in the system design. Some problems in transportation, information technology, and medicine are described next.

Transportation

Kirwan (2001) observed that the landscape of safety and risks in the transportation industry was changing due to increased volume, increased competition, and changing relationships between various forms of air, marine, rail, and highway transportation systems. Engineers must imagine more complex accident scenarios that transcend the more traditional localized boundaries among transportation systems and subsystems. For instance, the safety analysis of a flight should begin with flight planning, and not simply when the airplane pushes off from the flight gate. Delays in one part of the system, especially following an accident, may produce unplanned difficulties for passengers, pilots, and air traffic controllers:

> A fire at an airport causing a shutdown at the airport will divert many aircraft and cause chaos in the skies. The effects of a major airport closure will be felt all over Europe, and the ATM system will take hours to stabilize … The whole system will come under severe strain. Furthermore, many passengers will have to find alternative transportation…, placing strain on other transportation sections and increasing risk on those sectors. This can be called *risk export*, and will probably become more important as a concept as transportation systems become more interconnected and integrated. (Kirwan, p. 329, footnotes deleted)

Another class of problems is related to the privatization of transportation organizations in Europe. Large transportation units have been split up among several private owners, each with their own myopic safety concerns. In the case of marine transportation, Hormann (2001) remarked, "Procedures relevant to safety are not limited to the onboard activities. The performance of the operator's shore base and possibly his offices in outports is likewise decisive" (p. 364). Additionally, it has become increasingly difficult to attract young people to seafaring occupations. The pool of available employees thus consists of increasing numbers of marginalized workers. There is an increased need for education, training, and the definition of new careers in marine transportation.

Information Technology

Computer users have experienced increasing numbers of attacks by waves of viruses, worms, and Trojan horses in the wake of the attack on the World Trade Center in New York on September 11, 2001. Analyses of virus wave patterns indicate a power law distribution, according to Liebovitch (2004). The power law distribution offers some clues to the origin and operation of networks that are producing the attacks.

Brunnstein (2001) explained a fundamental problem: The Internet, IP protocol, and hypertext markup language "were developed for free exchange of information without *any* reference to requirements of security (confidentiality, integrity, availability) and safety (functional behavior, timeliness, persistency)" (p. 585). As a result, the contemporary users for e-commerce and other applications have had to plug numerous and potentially severe leaks.

Furthermore, what you see is not what you get. The fundamental sources of dysfunction originate and persist because they are not visible to the user. Wilpert (2001) reported several growing problems: (a) Fatal crashes of planes and automobiles can be perpetrated with automatic navigation systems that produce incorrect interpretations of data. (b) Autonomous agents in market packages might be infiltrated and mimicked by criminal agents. (c) Similarly, criminal agents have developed mobile code that can crack the security of checking accounts and other banking activities. Firewalls might not provide intended security. Espionageurs have been occasionally successful in reversing critical elements of firewall systems. (d) Limitations might be placed on the use of cookies. Cookies are bits of code inserted to a user's system when the user visits certain commercial web sites. Cookies provide information to the system about the user's past visits and may be used to tailor subsequent transactions with that user.

One might question how computer viruses came into being. According to Levy (1992), they were an unfortunate by-product of the artificial life vision. Well-meaning computer scientists wanted to fabricate a silicon life form, and the logical place to start was with the simplest life form, the virus. The first report of a malicious virus was traced to a computer dealer in Pakistan during the early 1980s. The dealer apparently had local customers as well as an enclave of customers from the United States. The dealer loaded the virus into computers that were intended for the U.S. customers, but sold clean computers to the local customers.

Medicine

Biohazard safety programs are largely borrowed from other safety areas. These need to be redeveloped for biotechnology applications. Critical situations include but are not limited to the following issues. According to a 1997 study cited by Zinn (2001), 3.5% of patients in German hospitals contracted infections that were traceable to the overprescription of antibiotics. Then-current prescription practices have led to several strains of resistant bacteria, and patients can contract more than one pathogen in this fashion. The estimated health-care costs required to remedy these secondary illnesses is roughly $9 billion per year, and the problem is thought to be growing.

Electronic storage of medical documents requires a complex process of handling, labeling, and encoding. The benefit of electronic patient records is that the original documents are not subject to loss in transit between one physician and another, and it is possible for multiple diagnoses to be made at a distance, especially where specialized expertise is thinly distributed. On the other hand, a study of incorrect handling and labeling of data records showed that 90% of records contained errors. Perhaps the most critical error resulted in the report of an amputation performed to the wrong leg of the patient.

The third major area for biotechnological hazards involves genetic products. Even before we consider whether the production of particular genetic products will have positive or deleterious effects on the ecosystem, there are potential problems in the laboratories and production plants themselves with the handling of materials (Czermak, 2001). New materials may produce specialized risks, and it does not help matters that the current regulations regarding the award for patents on genetic materials do not interface with safety standards (Baram, 2001).

Butterfly Effects

Can a shortage of rainfall in the central U.S. produce a political revolution in Syria? This question is essentially the question posed by the "butterfly effect," more formally known as sensitivity to initial conditions (Chapter 2), which has become an iconic representation of chaotic processes (Dooley, 2009). Karwowski (2012) observed that sensitivity to initial conditions is a serious concern with complex systems that has yet to be reconciled.

When the complexities of the medical system are combined with the complexities of information technologies, one obtains more complexity. Frenny (2007) reported a situation in which a major hospital in California received an e-mail from a woman in Pakistan saying that if the hospital didn't pay her "her $500" she would distribute the attached medical files all over the Internet. What happened? Come to find out, hospitals that are attempting to reorganize all their medical data into digitized files in anticipation of the new national health-care system were resorting to offshore contractors to convert the original data into digital files. Somebody forgot to pay the contractor. Fortunately, the problem was caught and fixed soon enough that time, but the story plants one of many seeds of doubt regarding where the next surprise from complexity might arise.

A clinical psychologist at a different site in California records its therapy sessions into digital video files and stores them on a cloud system, which was management's idea, not the clinician's (anonymous communication, 2013). Access to the cloud service was interrupted for 3 weeks because of an exceptional storm that was centered over the northeastern U.S. on the other side of the continent. Eventually, service was restored, but the control over the privacy of the files was a serious concern. The clinician concluded, "The so-called cloud is located in somebody's basement in New Jersey."

The financial crisis of 2009, according to some analysts, was triggered by the collapse of a market for mortgage securities that were being traded for more than they were worth. One could trace the proximal event to a progressive combination of financial deregulations (Dore & Singh, 2009). The problem became worse as the risks travelled through the financial networks very quickly, but any risk management innovations to be had did not travel as fast; the risks had enough of a head start to reach the crisis point before risk management could possibly catch up (Gallegati, Greenwald, Richiadi, & Stiglitz, 2008).

To return to the question about rainfall and revolution, the following story captures what could happen when unrelated systems interact. According to Lagi, Ba-Yam, Bertrand, and Bar-Yam (2011), a shortage of rainfall in the Midwestern U.S. where most of the corn and grain products are grown resulted in a shortage in the harvest in 2009, which resulted in higher prices for corn according to the usual dynamics of supply and demand. Two other factors contributed to the further rise in corn prices. Commodities traders, whose numbers now increased because the mortgage traders had to find new jobs, noted the price increases and were speculating on future prices of grain products. Meanwhile, the use of corn as a source of ethanol for fuel was increasing, adding to the demand. The higher prices for grain products travelled through the market networks until they reached actual food sales in the Middle East. Several countries in early 2010 were experiencing very difficult political times that were apparently getting close to revolutionary proportions. The high food prices added one more stressor—the proverbial straw that broke the camel's back—and set off the events that became known as Arab Spring.

Given that there are other politically precarious countries in the world, Lagi et al. (2011) recommended that policies be adopted to limit the use of corn for fuel so as to increase food supplies and keep revolutionary tensions subcritical. Is this a good solution? Consider what could be a better one, in light of two problems with the corn solution. (a) Not all corn, however and wherever it is grown, is fit for human consumption. Lower quality goes to animals, and the leftovers go into fuel; diverting fuel corn into human corn is not going to work well. (b) The speed and connectedness of the networks is at least as responsible for the outcome as the price of corn. The same dynamic could just as readily happen to a completely different commodity of worldwide importance. The problem was that the commodities markets were too interconnected and overpopulated. The solution, in light of what is currently known about small worlds, would be to disrupt the interconnections among the market trading programs. This solution would not guarantee the prevention of political strife, but it would eliminate spikes in food prices as a contributing variable.

DISCUSSION QUESTIONS

1. Three mathematical models were introduced in the course of explaining collective intelligence phenomena. What are they, and how are they different? Do they have any common properties?
2. Describe a complex person–machine system where coordination is an important component of the system's performance.
3. How might the Bandwagon game explain trends in technology adoption?
4. Reconsider your answer to Question 3. Can you imagine a scenario where Bandwagon dynamics might override good human factors or ergonomic design?
5. What are some limitations to conducting business through e-mail, compared with face-to-face transaction?
6. Describe some situations where an innovation in information technology appeared to save work, but instead it only saved work for some people and made more work for others.
7. Describe some other types of revenge effects in systems with which you are familiar.
8. Consider the innovations that are inherent in NextGen air traffic control. Will they assist or inhibit situation awareness? Make a case in support of either the pro or con position.
9. Consider the connection between iterations in system design and coupling as expressed by Helander (2007) and the cascades of technologies that are currently expected in airport manufacturing and uptake. How many iterations will it take to bring NextGen close to its ideal goal? What would be required to reach the goal?
10. Consider again Suh's (2007) and Helander's (2007) advisements about coupling. Is a large-scale Internet platform a good idea for the smart power grid?
11. Are there any visible similarities between programs for high frequency trading of commodities and the brokering of electric power resources?
12. How much additional complexity in NextGen air traffic control is expected to be caused by differences in equipage? Would it offset any advantages associated with the new automation feature? How would you conduct a study to answer these questions?
13. Consider the interface shown in Figure 13.8 for a military situation awareness application. What are the assets and limitations of this design?
14. Despite any lessons learned from the analysis of small world networks, person–machine systems and information sharing systems are growing toward greater centralization. Who benefits from all the centralization? At what costs to whom?

14

Environmental Design

This chapter considers the human factors principles associated with built environments such as homes, offices, larger building complexes, roadways, and the extreme environment of outer space. Topics such as lighting, noise, and extreme temperatures are fair topics to include in this category, but they were covered already in Chapters 4, 5, and 8. Chapter 8 considered the smallest unit of workspace, which would be the individual's workstation.

Microenvironments

A microenvironment is larger than an immediate workstation, and closer in scale to a room or an office. A private home might also be considered a microenvironment because the occupants usually have easy access to all parts of the home (Kantowitz & Sorkin, 1983). A macroenvironment is a much larger facility, and closer in scale to a hospital, a shopping mall, an airport, an apartment complex, or an industrial site.

Offices

There is actually little research on the best ways to organize offices other than the remarks in Chapter 9 concerning workspace densities. One logical point, however, that proceeds from known human factors principles is to organize the office by function. Communications equipment should be aggregated into one location, large displays that are meant to be shared by several occupants to another, and information storage to another (Kantowitz & Sorkin, 1983). The small office shown in Figure 14.1 is actually organized by function. It has other problems, such as a very small space for so much activity, and overall clutter.

Offices should be constructed around the user patterns. For instance, psychotherapists tend to organize their office space and furniture around a more businesslike theme when counseling individuals, but they build an atmosphere more similar to a living room when counseling couples or families (Anthony & Watkins, 2002). An easily visible clock is important in this type of office as it is in classrooms and other microenvironments where appointments need to begin and end on time. The need for extra seating and good visibility among people talking to each other probably influences choices of furniture and décor in addition to comfort and familiarity. Discrete entrances and exits might also be a concern in this type of environment.

Office configurations should reflect the autonomy and interactions among the occupants (Duffy, 1997). A hive would involve little of either; people work in small cubicles doing prescribed tasks. A cell would accommodate a talented worker who generally works alone and who benefits from being left alone (high autonomy, low interaction). A den accommodates teams of people who produce group products (low autonomy, high interaction). A club atmosphere accommodates a workgroup where individual-intensive work and group

FIGURE 14.1
Two views of an office organized by function.

work occur at changing intervals; this group would rate highly on both autonomy and interaction, and their workspace would include private and common areas.

Windows help. Office occupants who have views of attractive scenery outdoors report less job stress than those who have views of parking lots and other buildings or no windows at all (McCoy, 2002). The impact of natural surroundings is considered later in this chapter.

Homes

The allocation of space to rooms in the private home has changed in the last decades of the 20th century compared with older structures. Homes that were built in the United States during the 17th century were designed around a large fireplace, which served as the central heating system and kitchen for cooking and baking. The fireplace was built so that its floor was flush to the floor of the room rather than raised; thus, it was relatively easy to walk in, hang cooking pots, and manage the fuel. The family spent the majority of its time around the fireplace, more so in the winter, and it is no surprise that the phrase *living room* came into the language. Bedrooms were relatively small and were often appointed with a wash-up stand.

The 19th century brought central heating, flush toilets, and electricity, the benefits of which were obvious. The distribution of heat allowed the dwellers to allocate more time to other rooms. In the late 20th century, the bedrooms became much larger than ever before. The children and adolescents have their personal worlds in their rooms, including their televisions, stereos, and computers. The master bedroom facilitates one's removal from the other household members, and many more functions of life take place in the bedroom that used to take place in the common room or rooms. Some modern homes are built with a study off the master bedroom, although many households prefer to utilize a room that is shaped like a separate bedroom for this purpose—or perhaps the den. The television room, or family room, has moved into the space that was known for several previous decades as the (formal) living room. The amount of space allocated to these two types of rooms is now frequently merged into a great room.

The point of the foregoing discourse is to emphasize that the house should be built, or rebuilt, according to the way people expect to use the rooms. Obviously, households will differ in their habits, and houses will vary accordingly. According to one report (Ahrentzen, Levine, & Michelson, 1989), women spend relatively more time in the locations that facilitate household chores, whereas the men spend more time in the spaces devoted to passive leisure.

The age of the occupants also matters. The most critical issues are captured by the statistics on accidental deaths in home environments. Falls are the second single most frequent cause of accident death after the automobile accident. Most victims of falling deaths are either younger than 5 years or older than 75 years (National Safety Council, 1989, 2002). The elderly often prefer fewer stairs. Ladders are another source of falling incidents, as mentioned in Chapter 4.

Poisoning and drug overdoses are the third leading cause of accidental death. Here most victims are between ages 25 and 44 years. On the other hand, the advisement to "keep this and all other medications out of reach of children" is a good one. Accidental misuse of medications, for instance picking up the wrong prescription bottle (when they all look alike) without reading (or fully processing) the label first, could be a source of mismedication. The odds of this type of casualty increase with age, as older people use more medications; complications among the elderly are probably more likely to be classified as death or illness due to their known preexisting medical condition, however, and not cataloged as accidental.

Drowning is the fourth single leading cause of death, and third in the home. The bathroom can be the most dangerous room in the house for slipping and falling and possibly drowning in the bathtub. Common countermeasures include slip-free floor, tub, and shower surfaces, adequate handrails, and possibly a phone located in the bathroom in case of emergency.

Fires and suffocation rank fifth and sixth. Again, most victims are younger than 5 years or older than 75 years (National Safety Council, 1989). Age makes them less likely to vacate a dangerous situation unassisted.

Kitchens

A great deal of work activity occurs in the kitchen, most of which involves food preparation, but not exclusively so. Food-preparation tasks are facilitated by a triangular arrangement among the stove, sink, and refrigerator. This work triangle facilitates the largest amount of work with the smallest amount of walking between points (Figure 14.2). The less attractive alternative would place the three main devices in a straight line.

There are two other critical elements of kitchen design, however: storage and countertops. On the other hand, the quantity of both will depend on the planned size of the dwelling and the costs associated with the size. On the other hand, added quantities of storage and countertops can be made possible through the efficient use of space.

Not all countertops are created equal. Some of the space is used for the placement of decorative items, storage of some other items, food preparation, and messes that are headed for the dishwasher as soon as someone unloads it. One common requirement is a surface on which one may place hot pans and pots. Many decorative countertops become discolored by utensils coming right off the stovetop or from the oven. With a bit of forethought, however, one can install heat-resistant tiles and thus turn the entire countertop area into a hotplate.

If the kitchen is large enough, some of the central floor area can be allocated to a work island, with a work surface on top and additional storage cabinets below. Homeowners who prefer glamour over function often select granite countertops for their islands. Granite looks good, but knives and other cutting tools scratch it easily. Other homeowners prefer the look and utility of butcher block, which is essentially a large cutting board. The large, heavy, and permanent slab of wood is friendly to knives and cleavers, easy to clean and maintain but probably requires some light sanding every 5 to 10 years.

The stove in Figure 14.3 represents a classic dilemma in control panels—where to put the controls for the stove. In this example, the user must reach across the pots and pans to use the controls, and a large pot might obscure the controls. To the users of this particular system, however, the disadvantages of the control placement are greatly outweighed by the smooth surface of the stovetop. Here the burners are located below the surface rather than above it, and the burns and spills are much easier to clean than would be the case with conventional stovetops where the wells around the burners frequently collect caked-on crud.

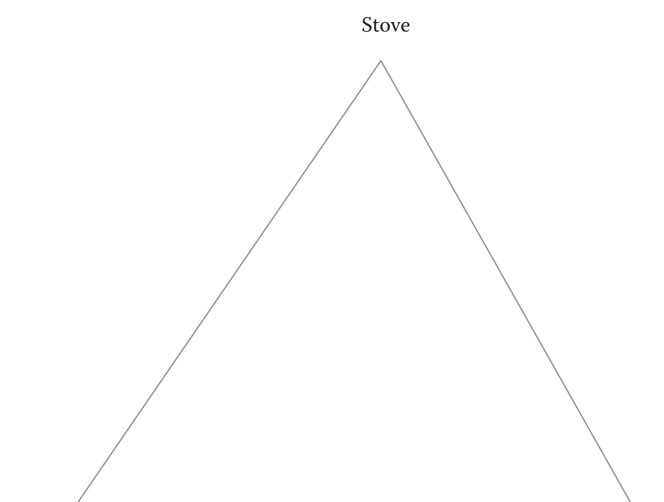

FIGURE 14.2
The kitchen work triangle.

FIGURE 14.3
Stove with flat even surface; controls located at the rear, behind the burners; countertops left and right are heat-resistant tiles. A portion of a butcher-block top on a work island is seen in the foreground.

Stairs

When this writer was in college, there was a set of outdoor stairs that ran from the street level up a hill to the main part of campus. The stairs were unusually shallow (riser height) and too long (tread depth), so that each step required more than a usual step, but was not wide enough to take two steps between stairs. Another student's voice was heard to say, "When the revolution comes, the guy who built these stairs will be shot!"

The point is that stairs should be built to accommodate normal patterns of walking. Figure 14.4 is a diagram of stairway features from the U.S. Department of Defense (2012) specifications. Nothing has ever prevented the staircase in a house from being designed

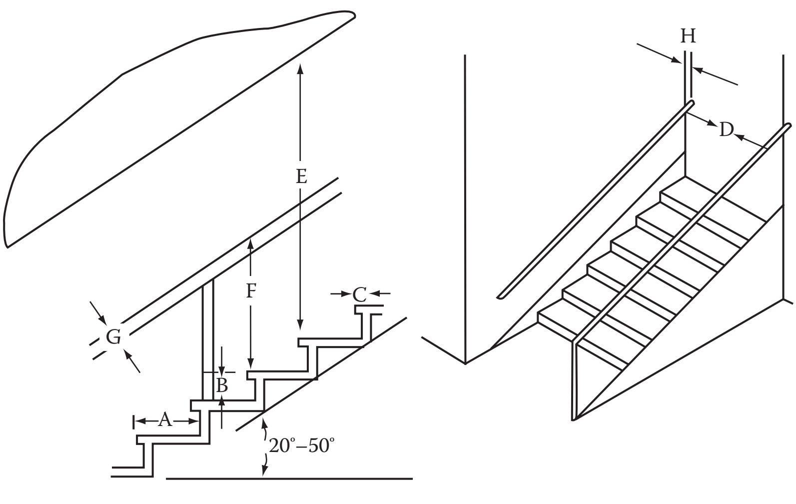

FIGURE 14.4
U.S. Department of Defense (2012) specifications for stairs: A, tread depth; B, riser height; C, depth of nosing; D, width between handrails; E, overhead clearance; F, height of the handrail from 3 leading edge of tread; G, handrail diameter; H, handrail clearance from wall.

differently; the spiral staircase is a case in point. Nonetheless, the DOD specifications are as follows: The tread depth should fall within 11.0 and 12.0 in (280–300 mm). The riser height should fall between 6.5 and 7.0 in (165–180 mm). The depth of the nosing should not be more than 1.0 inch (25 mm). The overhead clearance should be a minimum of 78 in (1980 mm).

The width between handrails should be 36 in (910 mm) for a one-way staircase and 51 in (1300 mm) for a two-way staircase. The recommended handrail diameter is 1.5 in (36 mm). The clearance between the rail and wall should be 3.0 in (75 mm).

Macroenvironments

Building and Facility Complexes

Many types of macroenvironment involve large facilities, which often, in turn, involve multiple buildings. Examples would include hospitals, manufacturing plants, universities, shopping malls, and airports. Obviously, there is a large variation in size and intended uses. As a general rule, the building complexes should be organized by function also. Linkages are also important; ideally, facilities users would have to travel only the shortest possible distance between two areas where tasks are performed. Instead of tracking linkages with helmets and eye movements, linkages could be determined by asking people to keep a log of which points they visited in which order during the course of a week. The needs of the employees and facilities managers might be different from those of customers and clients, however; thus, the design should balance the needs of the two groups of users.

Clark (1975) catalogued several common faults in facilities design. One group pertains to the interface between the inside and the outside. There should be sufficient space in and around loading docks to allow for the maneuvering of trucks and enough space in the platform area to accommodate the expected flows of materials going in and out. Rain canopies are recommended along with a buffer area between the inside, where climate is controlled for indoor comfort, and the outside, which is, well, outside.

The second group of design faults in facilities is related to walkways and production areas. Walkways should be wide enough to accommodate two-way traffic along with handcarts and other haulage devices that have to move throughout the space. Production areas often suffer from lack of sufficient floor space also. There might be enough room for the machines and people but not always enough room for materials and for storing work in progress. The results are clutter, tripping hazards, and other harmful encounters.

Although work supervisors spend a substantial portion of their time on the shop floor, they also need office space. The usual requirements are desks, relatively private and quiet areas to converse with employees or to use the telephone, and secure storage for records and reference materials. Unlike the standard office environment, the foreperson's office is often an isolated structure that is equipped with large windows for viewing the shop floor. Manufacturing facilities often find they have insufficient spaces of this type, especially since the technical organization of work has changed dramatically.

Facilities Management Systems

Large buildings often require electronic control systems. These facilities management systems typically contain five integrated systems: (a) heating, ventilation, and air conditioning;

(b) water utilization and flow; (c) fire and toxin safety; (d) building security and monitoring; and (e) energy management. Some of the controls and displays related to facilities management appeared in Chapters 4 and 11.

These systems rely on sensor technologies to measure events throughout the building complex, transform the sensations into electronic signals, and display the information in a suitable location along with appropriate controls. Part of the strategy in system design is to aggregate enough sensor information into a small number of centralized control stations so that a small number of vigilant humans can monitor and control large portions of the system. At the same time, the system should be designed with enough entry points so that emergency corrective actions can be taken near the location where the problem is occurring. It is not efficient to require maintenance personnel to walk a great distance to a centralized control station and back again to see if the control action had the desired effect. Alternatively, a route to a centralized control station might be impassable due to fire or other hazards in the area. In essence, the optimal system is hierarchically organized with different points of entry.

Figure 14.5 depicts a sensor display system that detects leaks in refrigeration gas. Leaks beyond a certain intensity are fatal to any human venturing into the area without a gas

FIGURE 14.5
Alarm unit that detects leaks in poisonous gas.

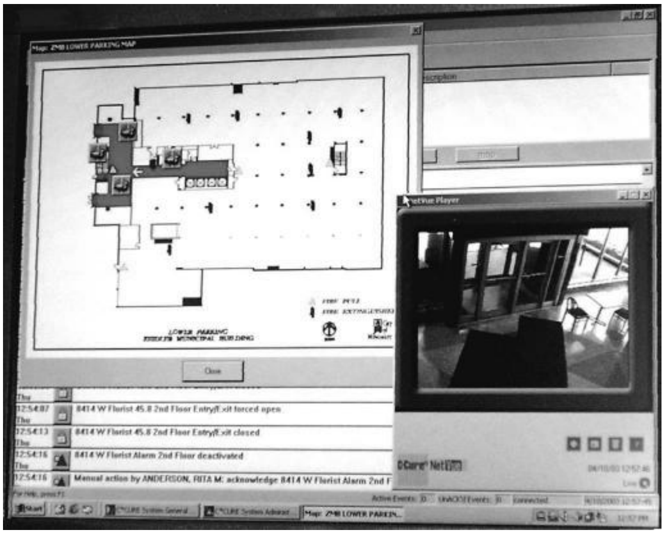

FIGURE 14.6
Integrated screen from a facilities management program showing open windows for building security camera and floor map.

mask. When the system detects a critical number of offending parts per million, a warning horn sounds repeatedly, and the red light at the top spins around and flashes. Many types of toxins can be detected and reported in the same manner.

Figure 14.6 shows a computer screen from an integrated security display. The menu that is more or less in the background lists each location in the system (a six-building complex in this case); sensors of different types are located in each of the areas. The operator may click on the screen to activate a pop-up map or perhaps see what a security camera in the area sees.

A sixth component of facilities management systems is communication throughout the building complex. To some extent, communication involves telephones and cell phones that are external to the remaining five areas of facility management systems. Computer technologies, however, allow for intranet communication among operators that are located at computer terminals. In any case, communication systems need to accommodate the volume and flow of communications among facilities management personnel and anticipate the need for communication during emergencies when one medium of communication may be impaired by the nature of the emergency.

System engineers often find a need to anticipate inappropriate actions of facilities users. Figure 14.7 depicts the atrium area of Milwaukee City Hall. Six out of seven balconies are shown. The building was erected in the 1890s. Shortly after the great stock market crash of 1929, however, people started using the balconies as diving boards. To prevent any further suicides, the city installed cage-like barriers all along each balcony. They were removed after 50 years, when aesthetics eventually won out over security concerns.

Defensible Space Theory

Some buildings, particularly large residential facilities, are more susceptible to criminal invasions than others. According to the theory of defensible space (Newman, 1972),

FIGURE 14.7
The atrium of Milwaukee City Hall showing six out of seven balconies.

facilities are less vulnerable if they are designed to allow residents to exert their territorial imperative. Figure 14.8, left, is a high-rise low-cost urban apartment building of early 1960s vintage. The entrance is secluded and not visible from the apartments. Several buildings of this type are lined up in rows to form a housing project. The passageways between the buildings are not easily monitored from the apartments or by police or private security patrols. It is also not possible to monitor what could be going on in the halls from the outside. Hall monitoring is typically not a concern for higher rent buildings, but higher

FIGURE 14.8
Two urban housing designs showing relatively less (left) and relatively greater (right) defensible space.

crime levels are anticipated here. In a variation of this design, hallways are arranged on the outer perimeter of the building where balconies are located in Figure 14.8 (left). They are protected on the outer edge by sturdy steel cages.

The building shown in Figure 14.8 (right) is an improved design that was introduced in the 1970s. The *X*-shaped layout allows visibility of the entrance and surrounding grounds from several apartments. The hallways around the central elevator are lit and visible from the outside. Buildings in the complex are staggered to enhance surveillance of one building from another and by the facility's security patrols.

Single-unit housing of the suburban variety is also susceptible to break-ins and theft. According to Newman (1972), houses would be less vulnerable if they were well maintained and appointed with symbols of territoriality such as low fences. The surrounding land should be well lit and with little dark area for perpetrators to hide in. Poorly maintained properties, according to the theory, would be more vulnerable because they would communicate that the owners were not exerting any control over the property.

MacDonald and Gifford (1989) tested the propositions of defensible space theory by preparing a set of photographs of homes that varied in defensible space characteristics. Photographs were shown to a sample of incarcerated home invaders who rated the homes for vulnerability as theft targets. The surveillance principle was supported; thieves avoided homes where three-fourths of the surroundings were visible from the street or adjacent homes by virtue of lighting and the absence of landscape barriers. Contrary to the theory, however, homes with poor upkeep were less likely to be targeted because they communicated that their owners did not have anything that they were concerned about protecting. The symbolic barriers were actually more inviting because they signified that there was something worth protecting inside and sometimes because the barriers themselves actually obscured visibility from the street.

Navigation through Facilities

Large and complex facilities often present a challenge to the users to find the place to which they are trying to go. Despite claims by architects and designers about the navigability of their creations, there is little behavioral research on record to support most claims (Carpman & Grant, 2002). We know from common experience, however, that people differ in their sense of direction; some people always get to where they are going, whereas others find themselves turning right when they should be turning left and vice versa. It is probable that any subjective ratings as to how well a person can navigate a facility will depend on this ability, whatever it is. People who tend to get lost often will have one opinion about the navigability of a building, whereas people with good directional sense would be less affected by unintuitive designs.

Tolman's (1932) principle of expectancy (Chapter 6) simultaneously gave rise to the concept of a *cognitive map*. The experiments with rats used radial mazes (Figure 14.9) designed so that the rat would be allowed to enter the maze from any point, but the (maximum) cheese would always be in the same target location. The rat would need to use different patterns of left and right turns to reach the target. When rats were successful in doing so, the conclusion was that they had attained a cognitive map of the maze. This conclusion contradicted the strict behavioral interpretation, which was that the cheese reinforced a pattern of left and right turns. This is not to say that a sequence of turns could not be learned, however.

Humans navigate through buildings and spaces using three modalities—route following, which is a sequence of turns, following a cognitive map of the building or space that

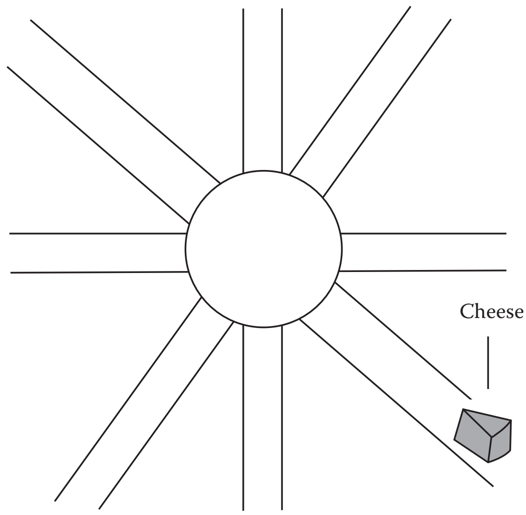

FIGURE 14.9
Radial maze.

is used as a reference, or movement relative to a landmark. A complex route within a building can disrupt the cognitive map (Carlson, Holsher, Shipley, & Dalton, 2010). The loss of orientation would happen with fewer or more turns depending on the completeness of the individuals' cognitive map and the complexity of the building's internal layout. Figure 14.10 depicts layouts of two hospitals showing primary, secondary, and tertiary routes. The layout on the left, which should *look* more compact, was the easier of the two for building users to navigate (Haq & Zimring, 2003).

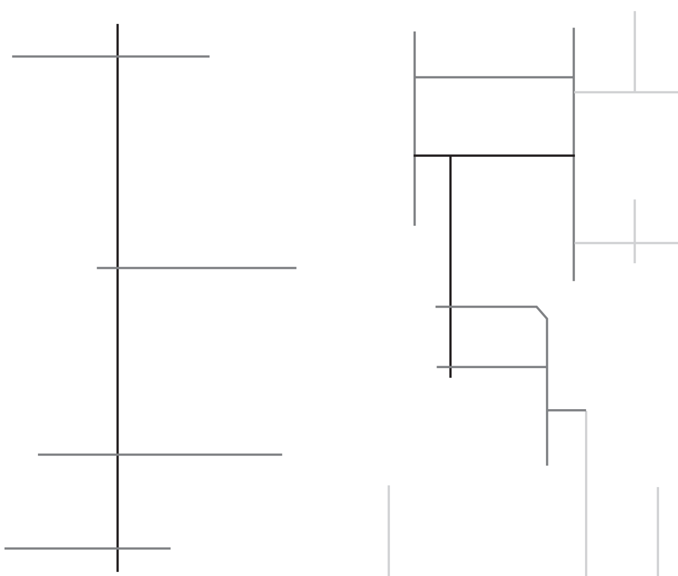

FIGURE 14.10
Two building layouts showing primary (black), secondary (thinner gray), and tertiary (thin light gray) corridors.

Building designs that conflict with users' expectations impede orientation. For instance, users expect the floor plans of multistory buildings to be the same on each floor and tend to become lost if not (Carlson et al., 2010). The Seattle Public Library is an extreme example; it has several floors that are designed as a book spiral. Although award-winning for its striking aesthetics, it still confuses long-term regular visitors.

Signage helps with navigation. The best design advice at present is to position direction information at critical choice points in the path of a user traversing the facility. Important choice points occur when hallways separate into two or more paths and where hallways intersect. At present there are no fixed rules of human factors to say when a list of locations with directional arrows is preferable to a kiosk that shows "you are here." It is important, however, that the signs be salient but not aesthetically annoying. The content of the signs should be complete enough to cover all the possible locations that users might like to visit; users might not be aware of all the possible places that they could visit. (How many shoe stores do they have here?) It also appears that people are less likely to become lost or misdirected when hallways intersect at right angles instead of narrower angles (Carpman & Grant, 2002). It is not currently known what makes an effective landmark for navigating a building, however (Carlson et al., 2010).

Special Populations

Residential facilities for the elderly with dementia present some special challenges. One's personal residence may look like all the other doors on the hallway with a number on it. One technique that seems to work is to place a small display outside the door to the personal room that contains a few of the resident's personal objects; those would be immediately recognized as signifying one's proper location in the facility (Day & Calkins, 2002).

Everyone needs some fresh air, but we do not want the residents getting lost or escaping either. One technique that appears to be successful is not to place exits at the end of the hallway where the residents can see the door from the opposite end. The vision of the door seems to present the wrong idea at the wrong time. Rather, a door to the outside should be located just to the side of the end of the hall; it will be where it is needed for emergency purposes, or loading and unloading, but it is less likely to trigger unscheduled exits (Day & Calkins, 2002).

Emergency Exits

OSHA has a set of standards for emergency exits for workplaces. Although they are written with employees in mind, they should fill the needs for public safety in facilities where the general public may be present.

The first set of regulations applies to exit routes. An exit route must be a permanent part of the facility construction and maintained during construction, alterations, and renovations. It should be made of fire-resistant materials. It should serve specific work locations only, so as not to create a bottleneck of traffic heading out one emergency door. Thus, multiple exits are probably necessary for larger facilities. Emergency doors should be unlocked from the inside so that employees will not have to find keys or look up special instructions to use the exit. The emergency doors must be hinged from the side, swing out, and support the expected occupant load. Once outside, the escaping personnel should not be confronted with dead ends; rather the walkways should lead to open areas to allow the employees to get a safe distance away from the building (U.S. Department of Labor, 2004).

Exit signs should say "EXIT" and be at least 6 in (15.2 cm) high with principle strokes on the lettering of at least 0.75 in (1.9 cm). Fire-retardant materials in the exit route should be good for 1 hr in buildings of three or fewer floors, and for 4 to 5 hr in larger buildings. Fire-retardant paints and chemicals must be refreshed periodically. Emergency alarms need to be in place and operating correctly. Employees must be informed and trained in emergency evacuation procedures (U.S. Department of Labor, 2004).

Sick Building Syndrome

The Centers for Disease Control and Prevention fielded numerous reports of sick building syndrome during the 1970s and 1980s. The buildings were typically newer office buildings with little outside air circulation. The workers in those buildings reported unusually high frequencies of symptoms such as skin irritations, watery eyes, runny noses, abnormal taste or smell sensations, tiredness and fatigue, or reduced mental concentration (Hedge, Erickson, & Rubin, 1992; Rothman & Weintraub, 1995). The first response was to take assessments of airborne chemical toxins, but the consistent finding was that chemical toxins were not present in levels above OSHA standards.

Workers reporting symptoms were more often female and incumbents of lower level jobs. Studies of their work content revealed that they were experiencing high amounts of job stress and lower job satisfaction, both of which were more extreme for workers who used computer terminals full time (Hedge et al., 1992). Hedge et al. (1992) did consider the possibility that benign levels of airborne particles and insufficient air exchange may be mitigating factors that brought about the particular symptoms. Job stress continues to be the primary explanation for the sick building phenomenon (McCoy, 2002; Sulsky & Smith, 2005). The large numbers of workers in a building, or a floor of a building, reporting symptoms reflects a form of contagion that is otherwise known as mass hysteria (Rothman & Weintraub, 1995).

The Great Outdoors

Aesthetics and Stress

There is some sense in the health psychology literature that pleasant outdoor environments reduce stress and enhance the feeling of well-being. Thus, offices are sometimes built around reflection ponds or in attractive semiwooded environments; most office dwellers do not rate one of those. Apartment dwellers that have a choice probably prefer to overlook a lake or a park instead of a brick wall or a neighbor's apartment window, all other aspects of the apartment being equal.

Herzog and Bosley (1992) investigated the features that make one outdoor environment more attractive than another. It appears the participants in their study responded to five contributing factors when identifying their favorites: tranquillity, mystery, coherence, spaciousness, and focus. *Tranquillity* was the sense of peace that one might get from being in the environment. *Mystery* was the impression that more could be seen if one walked into it further. *Coherence* was the sense of being able to structure the environment visually rather than perceive it as an endless jumble of rocks or trees. *Spaciousness* meant that there was plenty of room to wander around. *Focus* referred to the presence of a focal point that commanded one's attention, that was used as an experimental stimulus by Herzog and Bosley (1992). Figure 14.11 is an example of a focal point within a desert scene.

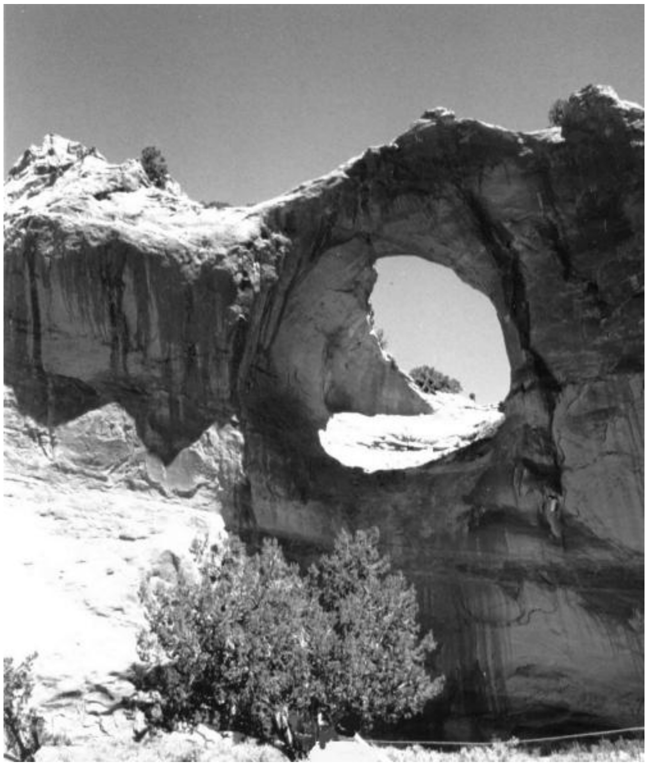

FIGURE 14.11
Window Rock in Arizona is an example of a focal point in a desert scene.

The analysis of ratings for 66 pictures showed that there were seven categories of image that were organized along four bipolar dimensions. They were: mountains versus fields and streams; desert versus large bodies of water; rushing water (e.g., waterfalls) versus gardens; and misty mountains versus anything that was not a misty mountain.

Navigation

Once again, people find their way by means of route knowledge, cognitive maps, and reference to landmarks. Lapeyre, Hourlier, Servantie, N'Kaoua, and Sauzéon (2011) assessed the extent to which synthetic vision systems improved wayfinding of pilots using synthetic vision systems. The nonpilot participants in the study operated a flight simulator and had to rely either on the standard display or the display that was graphically enhanced. The standard display superimposed direction metrics over a two-color background with blue for sky and brown for land. The synthetic vision display superimposed the same metrics over a more detailed picture of the terrain. Other experimental conditions varied the outdoor visibility conditions that accompanied the directional display.

The participants made a flight to eight waypoints, then after a rest period were asked to recall landmarks, routes, and directions using pictures taken from the simulated flight. In the scene recognition task, they had to pick 8 out of 13 pictures that represented correct approaches to the waypoints. In the scene recognition task, the 8 correct waypoints had to be put into the correct order. In the sketch map task, they were handed a sheet of graph paper with the departure point marked and were asked to draw the pattern of directional changes that took place in their trip with landmarks as a reference. To determine their wayfinding capability, they had to replicate the route on the simulator without the synthetic vision system in both high and low visibility conditions. The synthetic vision system made a positive impact on all three levels of spatial knowledge.

Other types of navigation tasks involve two or more people trying to coordinate their locations and directions. When people are approaching a target location from different directions, "turning left" has different meanings. One person working alone with a map might do better with north pointing up or the directional heading pointing up, depending on the task. When two people are working together, however, they should have their maps oriented in the same direction for best performance (van der Kleij & te Brake, 2010). Thus, in the case of multiuser electronic displays, the system should not permit different operators from using different orientations of the map.

Playing in Traffic

Some of the human factors issues related to automobile operation were covered in Chapter 2 concerning automation, Chapter 4 concerning visual displays, Chapter 7 concerning controls, Chapter 6 concerning cognitive workload, and Chapter 8 concerning anthropometry. This section is concerned with the issues connected to the roadway and the use thereof.

According to one undocumented story, there were four automobiles registered in the city of St. Louis in 1894. Two of them managed to crash at an intersection. The first fatal crash was documented in Britain in 1899 when a single vehicle lost a wheel and crashed. Witnesses reported that the vehicle was traveling unusually fast at the time of the accident (Evans, 2002). The counterpoint between drivers' and vehicles' faults took shape immediately. Inasmuch as the majority of road accidents today are associated with driver variables, driver variables are considered first here.

Exposure

The rates of motor vehicle fatalities are somewhat related to road miles of exposure, but there are other trends to indicate that other large effects are involved. It is not possible to obtain drivers' mileage of exposure from most countries for comparison; thus, the number of fatalities per 1,000 vehicles is a more manageable index. Moreover, data from single-vehicle crashes is preferred for many purposes of traffic accident analysis because they circumvent the problem of figuring out which driver was at fault.

Curiously, the nations with the lowest number of vehicles per 1,000 capital units have the highest fatality rates per 1,000 vehicles (e.g., Ethiopia, Malawi, Mozambique, Tanzania), whereas the most heavily motorized countries have the lowest rates (Australia, Germany, Japan, Sweden, United States; Evans, 2002). The nexus of sparse roads, poor-quality roads, and low utilization by motor vehicles is associated with poor regulation and poorly developed driving habits and risk perception. The situation is akin to a revenge effect: The absence of obvious dangers makes the situation all the more dangerous (Tenner, 1996).

Rates per 1,000 vehicles have declined in all nations since 1980, although the reduction rates vary substantially. The nations with the highest rates in 1980 showed the sharpest drop (Evans, 2002).

Driver's Age

It is a robust finding that drivers aged 18 to 25 years are twice as likely to suffer fatal crashes as drivers aged 40 to 50 years. The fatality rates for male drivers are three times that for females throughout the age span. Given that a one-car crash occurred and someone in the car died, however, the rate of female driver deaths is 50% greater than the rate for male drivers (Bedard, Guyatt, Stones, & Hirdes, 2002). In other words, women do not appear to survive the crash as well as men.

The popular opinion that elderly drivers are more dangerous than younger or middle-aged drivers may be ill founded for similar reasons. Evans (2002) reported a database in which the rate of single-vehicle crashes per million population were plotted against drivers' age; there was no increase in accident rate for drivers aged 70 to 80. Bedard et al. (2002) found, however, that given that a crash occurred with fatality resulting, drivers aged 80 years and older are five times more likely to die than younger drivers. In other words, the elderly drivers do not survive the experience well. Similar results were reported by the National Safety Council (2002) for fatalities that could have been either drivers or passengers; the highest morbidity rate was associated with people aged 75 years and older. Thus, the evidence favors the interpretation that the elderly are less likely to survive a crash than younger people, but not as likely to be responsible for one.

There is still some concern among some policymakers about elderly drivers in light of their increasing numbers. De Raedt and Ponjaert-Kristoffersen (2001) studied 84 licensed drivers aged 65 to 94 years who were referred by their insurance companies because of accident involvement. In-depth interviews with the participants indicated that 63% of them were at fault in at least one of their accidents. De Raedt and Ponjaert-Kristoffersen administered a set of neuropsychological tests, a driving simulation, and the Belgian standard driving test. They found that it was possible to predict accident involvement from those tests if each type of accident—for example, intersections while making left turns, intersections while making right turns, front-rear collisions, sideswipes, and parking—were considered separately. Separate tests predicted different types of accidents because of the specific visual, memory, and motor functions involved in each type of maneuver. The prediction rates were about 19% better than chance (50%). The accidents could thus be explained by declines in perceptual and motor skills. These prediction levels are probably inflated relative to the general population of drivers in that age bracket, however, because of the way the participants were referred to the study. Statistically, prediction levels are optimal when the chance rate is 50%. Real-world base rates are often very different from 50%, however, which would imply that De Raedt and Ponjaert-Kristoffersen's prediction accuracy could be seriously inflated relative to the general population of drivers.

Blood-Alcohol Concentration

Alcohol intoxication is thought to be responsible for approximately 55% of driving fatalities (National Safety Council, 1989). It is not certain, however, to what extent fault is assigned to a driver in the databases because that driver was intoxicated rather than because that driver actually did something to cause the accident.

In 2006, the legal limit for alcohol in the bloodstream while driving was 0.10% in 36 of the U.S. states. This level is the midpoint between 0.00% and 0.20%, which is the average level of intoxication for drivers involved in fatal crashes, if they were under the influence at all (Foss, Stewart, & Reinfurt, 2001). Small deficits in drivers' judgment actually set in at 0.07%.

Thus, 35 states lowered the legal limit to 0.08%. Federal law now required that U.S. states adopt a BAC limit of 0.08% or face cut-backs in federal highway funds (Duiattorneytab, 2013). Evans (2002) noted that when the U.S. states raised the legal age for the purchase and consumption of alcohol from 18 to 21 years, there was an 18% drop in traffic fatalities. Foss et al. (2001) reported that no effect could be detected in traffic fatality data after North Carolina lowered its legal limit to 0.08%, and the results from other states were too murky to draw a conclusion about the effectiveness of lowering the limit.

Seat Belts

Seat belts were required by law as standard equipment in automobiles sold in the United States beginning with the 1963 model year. The federal legislature was strongly influenced by the testimonies of John Paul Stapp and Edward Murphy Jr. Stapp was an Air Force test pilot for experimental aircraft. He claimed to have crashed so many times he would never have survived without his seat belt. Murphy was his chief engineer who is credited with the famous Murphy's law: If something can possibly go wrong, it will (Tenner, 1996).

State laws that required that seat belts actually be used did not appear until 1984 (Houston & Richardson, 2002). It is now estimated that the use of seat belts reduces a driver's rate of fatality by 42% to 54% (Bedard et al., 2002; Evans, 2002). By 1996, however, it was estimated that only 68% of drivers actually used their seat belts. Part of the low utilization was attributed to the states' rules for enforcement of their seat belt laws. Most states only permitted ticketing of drivers for failing to use seat belts if they were stopped for another traffic offense. The legislative trend, however, moved toward "primary enforcement," whereby police may stop and ticket a driver for seat belt failure alone. In the state of California, which switched to primary enforcement, the new law resulted in a 4.9% reduction in traffic injuries (Houston & Richardson). Michigan reported a 13% increase in seat belt use (Eby, Vivoda, & Fordyce, 2002), but no injuries were reported in that study.

Currently, 33 U. S. states have a primary enforcement law seat belts, 16 states use secondary enforcement, and New Hampshire has neither for drivers over age 18. States vary as to whether their requirements affect the front seat occupants only, all seats, or vary by the age of children present (Governors Highway Safety Association, 2013). Despite laws and occasional improvements in technologies, driver behavior often attenuates the impact of the innovation. This principle is known as *selective recruitment:* Careful drivers are likely to respond to a new law or safety idea, but risky drivers remain risky.

Speed Limits

According to Bedard et al. (2002), drivers crashing at speeds greater than 69 mph (111 kph) were 2.65 times more likely to die in the crash than drivers crashing at slower speeds (an average of 35 mph or 56 kph in that database). In two-car head-on crashes each vehicle only needs to be traveling at half those speeds to attain the same effect.

Speed limits for roadways are usually set in response to several variables: road capacity, traffic density, sharpness of curves, and population or pedestrian density. The latter includes the number of driveways or intersections opening onto a particular road. Prior to 1974, the speed limits for U.S. Interstate highways would range from 50 to 70 mph. The OPEC Oil Embargo in 1973 produced a petroleum supply shortage in the United States, which quickly resulted in automotive design changes and legislation. One important item was the drop in the national highway speed limit to 55 mph. This limit was more related to the fact that automobile engines at the time were more efficient at 55 mph than at higher speeds than it was to anything else. The public was more likely to hear about the projections of highway safety, however. In any case, the predictions were correct according to Evans (2002): The average travel speed dropped from 63.4 mph to 57.6 mph, with a 34% drop in highway fatalities.

The interstate speed limit in the United States was raised to 65 mph for rural areas in the late 1980s in response to the increased use of overland trucking, and the commercial value of increasing speed limits. The political pressure on petroleum supplies had dissipated during that epoch. Reports on the impact of the change in the national speed limit are not available, although the hypothesis would likely be that highway fatalities in rural areas would have increased. Currently 33 U. S. states have maximum speed limits of 70 mph or above on rural interstate highways, and another 13 have limits of 65. Interstate highways running through urban areas are often slower (National Motorists Association, 2013).

Excessive speed at lower levels, for instance in urban areas, can be problematic as well. According to one report (Thompson, Fraser, & Howarth, 1985), 36% of drivers exceed the speed limit in school zones when children are present. The presence of students in groups of 10 or more promotes a reduction in speed of only 1 mph. Drivers are thus not well prepared to respond to children darting into the street. Thompson et al. (1985) opined that, all other factors being equal, children are at greater risk of injury in a car–pedestrian accident because they are short and thus not as readily visible as adults: "A child involved in an accident is frequently assumed responsible for the accident's causation. Accident-involved drivers often say, and probably believe, that the sudden appearance of the child they collided with 'never gave them a chance' " (Thompson et al., p. 1473).

Risk Homeostasis Theory

Driving speed and the other antics of drivers are the end result of personal risk assessments, according to risk homeostasis theory (Wilde, 1988). Drivers evaluate the relative rewards and costs of the safe options and the risky options and choose a level of risk that is best suited to their proclivities. Getting to the office sooner might outweigh the odds of a speeding ticket. Municipalities vary as to whether the traffic police will stop a car at 5, 10, or 15 mph above the limit. Speeding in excess of 20 mph over the limit could result in more than detainment and fines; it could result in a mandatory court appearance, which is an inconvenience that is often far greater than arriving at most destinations sooner.

The odds of an accident are another deterrent, and drivers decide for themselves whether the odds of an accident outweigh the merits of arriving at a destination sooner or the sheer fun of opening up the engine. Evans (2002) noted that highly skilled drivers are not necessarily safer drivers. Drivers with higher skill levels could perceive their odds of a wreck being lower than average, and thus they would convince themselves to take higher risks than average. Similarly, straight, open roads may appear easier to drive than other roads, but that would only induce drivers to drive much faster than they ordinarily would (Tenner, 1996).

Shortly after the introduction of risk homeostasis theory, there were several studies that offered data that limited the validity of the basic proposition. The upshot was that individuals' risk assessments are not exact, but rather loose and not always well informed (Lonero et al., 1994). Salient positive incentives should contribute to safe roadways within the rationale of risk homeostasis theory. Risk homeostasis theory appears to explain, however, why some people are willing to work in hazardous occupations whereas others would not consider those jobs (Halgrave & Colvin, 1998).

Risk perception may be a complicating variable. For instance, many drivers imagine that speed and concentration are complimentary safety variables. They see speed as the more important risk factor when traveling at high speeds, but concentration as the more important risk factor at lower speeds. If the drivers were more objective, they would see both factors as contributing to their safety under all conditions (Jorgensen & Pedersen, 2002).

Recent efforts to automate the automobile and move more forms of control away from the driver to the car produced the idea of putting speed limit warnings in the car's operating system to alert drivers when they exceeded the speed limit. Berlin, Reagan, and Bliss (2012) outfitted government-owned cars with such a warning and tested the effect on experimental participants who drove them in a 30-mile radius in Michigan. There was some positive effect on reducing the time spent in excess of speed. Perhaps a more interesting result, however, was that drivers who regularly wore their seatbelts exceeded the speed limit by 6 to 10 mph more often those who did not wear their seatbelts.

The speeding alert system requires that the car be equipped with a GPS, an Internet connection, and a database of speed limits. Thus, the viability of rigging a car to give out warnings might be limited outside the driver's usual haunts.

Driver Distractions

Driving is a preattentive process, meaning that the activity does require attention to visual stimuli on a continuous basis, some well-honed emergency responses, and not a lot of memory or executive decision-making activity. In other words the task can be done with a great deal of excess channel capacity that could be devoted to talking, listening to the radio, and solving problems on the job. The list of alternative activities now includes operation of a car phone, cell phone, onboard computer, and geographic information system. One automobile manufacturer is advertising video–DVD systems that are meant to entertain the passengers, and the driver as well when the car is parked while waiting for something or someone. The video system at the driver's screen automatically turns off when the car is put in gear. Before the advent of the new gadgets, bored drivers had been observed spreading business papers across the dashboard and examining them while driving.

Driver distractions often have low-tech origins—use of conventional maps, talking to passengers, tuning radios, operating cassette players, eating, and drinking. Driver distractions were recorded in 3% of police reports of fatal accidents in England and Wales during the 1986–1995 period (Stevens & Minton, 2001) and in 11% of accidents in the United States with any level of police-reportable severity (Wang, Knipling, & Goodman, 1996). Both reports came from a period shortly before the widespread boom in the use of cell phones. By some estimates, driver distractions could be responsible for up to 30% of contemporary automobile crashes (Shelton, 2001). These crashes could be the result of drivers not attending to road stimuli in favor of something else. Alternatively, the driver may indeed be attending to the correct stimuli but not devoting enough channel capacity to driving (Kantowitz, 2001) and thus not taking appropriate action soon enough when needed.

The risks associated with some driving distractions are now available (Dingus et al., 2010), as reported in Chapter 6. It is not usually possible, however, to reconstruct what the driver was doing just prior to the crash, or to what extent drivers were doing the same things when they were not crashing. Legislatures have been responding by applying legal limits on driver distractions, for example, the driver must keep one hand on the steering wheel while using a phone. Manufacturers are introducing safety features such as voice-activated controls or a control system that prevents the driver from programming a geographic information system unless the vehicle is stopped.

By other standards of imagination, however, the drivers who might have had some personal privacy in their automobiles should be captive customers for advertising, interacting with products on their touchscreen panels (where the radio and ventilation system controls used to be). As of early 2012, the National Highway and Traffic Safety Association (NHTSA), which facilitates dialog between manufacturers and the U.S. Department of Transportation, had set a standard of not more than 30 characters of text should appear on a display screen for a nondriving task. Manufacturers were trying to push for more text. One scenario was to introduce advertising through the vehicles GPS system: While driving past a shopping mall, the driver can find out all the special deals going on and hopefully pull off the road to go buy something.

Another scenario comes from the innovations in the music industry. What used to be a free-service FM radio and a CD player (user-supplied content), is being replaced with digital FM and satellite radio (users pay for the service) and no CD player. The radio flashes the titles of the music across the screen, and the driver should be able to push a buy button to download the music into the system's hard drive. Getting the music out of the car has not really been explained yet, and it would appear that the driver's credit card would need to be on file somewhere for continual buying. In the broader range of new ideas, the central touchscreen display is a tablet that can be designed at different levels of complexity. The current functionalities seem to be limited to Internet-based services in combination with GPS, cell phones, text messaging, and e-mail. The expected market, which is somewhat captive and forced upon drivers who might ordinarily prefer high-end vehicles, has been estimated between $15 billion and $100 billion (Wiese & Lee, 2007).

While all the distractions are in various stages of development, so are means for counteracting the distractions by automating the vehicle (Merat & Lee, 2012). Possibilities include total automatic driving, or limited automation involving synthetic vision systems linked to a heads-up display on the windshield that will alert the driver to road hazards and options for avoiding them. Another group of ideas connected to semiautomatic driving involves alerts to drivers for following the car ahead of them too closely or for veering laterally out of the driving lane.

Whereas it might appear at first blush that automation could compensate for distraction and fatigue, drivers who engage the automation only become more bored and fatigued (Mkrtchyan et al., 2012); control over the level of automation does not lessen the additional fatigue or boredom, and could actually create an additional distraction as the driver's trust in automation becomes challenged (Neubauer, Matthews, Langheim, & Saxby, 2012). Others respond to automation by engaging in more nondriving tasks (Carsten, Lai, Barnard, Jamson, & Merat, 2012).

Driver Response Times

State driving manuals typically provide charts and graphs showing the stopping distance for cars at varying speeds. The response time that is required to gauge a situation and step on the brake pedal is usually around 0.75 s and figured into the projections of stopping distances. A relatively recent report (relative to driving manuals) indicates that the driver response time, as determined with a driving simulator, to hazards approaching from peripheral directions, ranges from 0.65 to 0.85 s (Lovsund, Hedin, & Tornros, 1991).

It is important to recognize that simulation drivers are primed to make optimal responses. Their alertness levels are high. The typical driver, on the other hand, is operating at suboptimal conditions with multiple distractions in place, and a probable experience of nothing bad ever happening while driving. If a driver is operating under such lax conditions and should suddenly encounter a surprise on the roadway, the response time could be considerably greater, according to Eubanks (1993). Nonmoving objects in the road (other than stopped vehicles with brake lights on) may take additional time to notice and process compared with other types of hazards. Pedestrians entering the street or car doors opening may be more challenging to detect for drivers entering the segment of road from around a blind corner rather than from the head-on direction. Situations with these perceptual challenges combined with suboptimal response readiness could pump the driver response time as high as 3 s.

Roadway Configurations

Figure 14.12 (from Evans, 2002) characterizes roadway fatalities for interstate, noninterstate, rural, and urban locations with different configurations. Interstate highways are less often fatal than other roadways, and rural environments are more often fatal than urban environments. Within the urban and rural environments, arterial, collector, and local routes have different orderings of fatality rates. Of additional interest, death rates at night are approximately triple those for daytime accidents in both urban and rural settings (National Safety Council, 1989).

The three-lane road design predisposed drivers to head-on collisions. The three lanes were north (or east), south (or west), and a third lane that was shared by the two directions for passing. The road markings would indicate when it was each lane's turn to use the lane for passing. The markings told drivers when their turn to use the lane would start but gave them no warning as to when their turn would end, hence the inevitable. The logic behind this road design is uncertain. Nevertheless, as the collision rates increased on these roads, often as a result of increasing traffic, they were slowly replaced by wider four-lane configurations. Three-lane roads still exist, although high-speed two-lane highways are probably more frequent. The latter are classic examples of straight, clear roads in the middle of nowhere where accident rates are comparatively high.

Despite the drama associated with head-on collisions, they represent only about 3% of all two-car collisions in the United States (National Safety Council, 1989). Front-to-rear collisions and angle collisions are each about eight times more frequent. Retting, Weinstein, Williams, and Preusser (2001) reported that 90% of urban arterial crashes in the Washington, DC area could be accounted for by seven different accident types: (a) making a left turn and

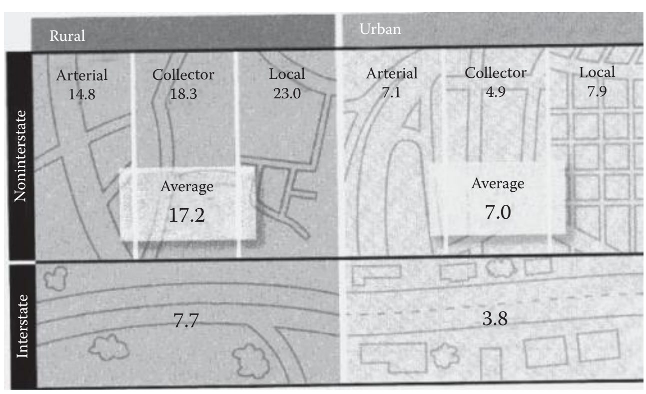

FIGURE 14.12
Roadway configurations and fatality rates per billion kilometers traveled. (From Evans, L., *American Scientist*, *90*, 248, 2002, illustration by Barbara Aulicino. Reprinted with permission. Copyright 2002 by Sigma Xi, the Scientific Research Society.)

colliding with an oncoming vehicle that has the superior right of way, (b) being rear-ended by another vehicle while waiting to make a left turn, (c) running red lights, (d) vehicle is stopped or just starting in a travel lane and is rear-ended by another vehicle, (e) changing lanes, (f) being run off the road, often into a parked vehicle, and (g) hitting a pedestrian (p. 725). Retting et al. offered several options for design correction including: the use of left-turn lanes, widening the roads, increasing signal duration, restriction of left turns by use of median barriers, and restriction of driveways onto the thoroughfare.

Lighting and Signals

Rural roads are often poorly lit, which may contribute to their crash rates, especially at night. When lighting is installed, accident rates on those roads do decrease, but a generalizable rate is not forthcoming because of all the other roadway factors that could be involved. Not making things any clearer is the finding that once the lighting has been installed, drivers tend to increase their driving speeds by 5% on straight roads and 1% on curved roads (Jorgensen & Pedersen, 2002).

The popular perception is that municipalities do not install traffic signals at intersections unless the crash or death rate is high enough. This perception is probably not entirely wrong, although congestion may be another contributing factor in the decision to install traffic light systems. On the other hand, increased use of an intersection, perhaps due to population mobility in a suburban area, increases the number of opportunities for accidents.

The unregulated intersection, particularly the four-way stop intersection, poses a challenge to drivers, especially in times of high traffic density. Most state driving manuals specify the driving rules in these situations, but only up to a point. Rule 1 in Wisconsin is that the car that stops at the intersection first proceeds first. Rule 2 is when in doubt about which vehicle arrived first or when two vehicles arrive simultaneously, the one on the right proceeds first. Right here we have the memory challenge: Which car arrived first? Given three or four cars, which car on the right of whom goes first? Then there is the popular variation where drivers proceeding in east–west directions alternate with those proceeding north–south. There is also the problem of making a left turn under some of these conditions because one car's turn by one rule may conflict with the right-of-way rules somewhere else in the driver's manual. In any case, the drivers must figure out which rule is in play, and when it is their turn. The cognitive processes are interesting, and suggest coordination dilemmas that could generalize to other types of situations (Guastello & Guastello, 1998), as discussed in Chapter 13.

Regulated intersections could possibly pose an opportunity for human factors issues to arise. Chen (1999) published a photograph that depicted a cluttered signal configuration in a city that was otherwise in the middle of a desert region of the southwestern United States. Newcomers could have substantial difficulty determining which signal is theirs.

How often do drivers observe the lights when there are no problems with signal clarity? Red-light running accounted for 22% of urban crashes in the early 1990s (Retting, Williams, Preusser, & Weinstein, 1995), and 3% of fatalities during the 1992–1996 period (Porter & Berry, 2001; Retting, Ulmer, & Williams, 1998). Porter and Berry conducted a telephone survey of a stratified sample of U.S. drivers, of whom 19.4% reported running a red light in the last 10 intersections they used. Drivers who failed to observe lights reported that they were in a hurry, had no passengers, and figured that the odds of being stopped by police were very low. The typical scenario was trying to squeeze through an intersection as the light was changing from yellow to red ("squeezing the tomato"). The survey results also indicated that drivers who run lights are likely to be younger (18–35 years), less likely to wear seat belts, and also predisposed to weaving through traffic, speeding, tailgating, and gesturing to other drivers. Contrary to their hypothesis, however, Porter and Berry found that running lights was not associated with reported frustration; weaving, tailgating, and gesturing were more closely correlated to reports of frustration.

Outer Space

The developments in manned spaceflight during the second half of the 20th century were outgrowths of engineering and design ideas, including basic terrestrial flight, from the first half of the century. Caprara (2000) and Dewaard and Dewaard (1984) are recommended illustrated histories. Kiselev, Medvedev, and Menshikov (2003) presented an illustrated compendium of technical and system design objectives. Out of necessity, this final section of the chapter sticks close to human factors issues as they have emerged.

Brief History

Space missions fall into three basic categories. Sortie missions are flights that go out to do a job and then return. Temporarily manned space stations are intended for habitation for 6 months or less. Permanently manned stations are intended for habitation for durations in excess of 6 months. Personnel in the latter group are expected to rotate periodically, but the operation of the station is meant to be continuous.

The first manned spaceflights (Project Mercury in the United States) were single-person vehicles that were launched into space at the tip of a three-stage rocket (Figure 14.13). The capsule landed in the ocean with a parachute for a braking device when it returned to Earth, and it had to be retrieved by an aircraft carrier. These capsules were not reusable. Reusable space vehicles that landed on solid ground were not available until the series of space shuttles in 1980. There were parallel missions conducted by the Soviet Union, which

FIGURE 14.13
Mercury capsule accommodating one astronaut.

is actually credited with the first manned spaceflight, the first space walk, and the record-setting spaceflight of 438 days.

The second series of spaceflights involved two astronauts (Project Gemini in the United States). Here we had the opportunity to see how well two people could get along when stuck in a tin can together (Figure 14.14). One technical goal of the Gemini series was for an astronaut to exit the space capsule and walk in space while tied to the craft with a cord. This goal was fundamental to eventual work in space and for making repairs.

The third series of spaceflights involved three astronauts (Project Apollo in the United States) and culminated with a landing on the moon in 1969 (Figure 14.15). The landing

FIGURE 14.14
Gemini capsule accommodating two astronauts.

FIGURE 14.15
Apollo capsule accommodating three astronauts.

was accomplished by launching a smaller two-person craft from the main capsule (Figure 14.16), which continued to orbit the moon under the auspices of the third astronaut. This landing unit was capable of lifting off from the moon and returning to the Apollo.

The launch of Skylab (Figure 14.17), a temporarily manned space station, similar Russian projects, and drone spaceflights to other planets dominated space adventuring during the

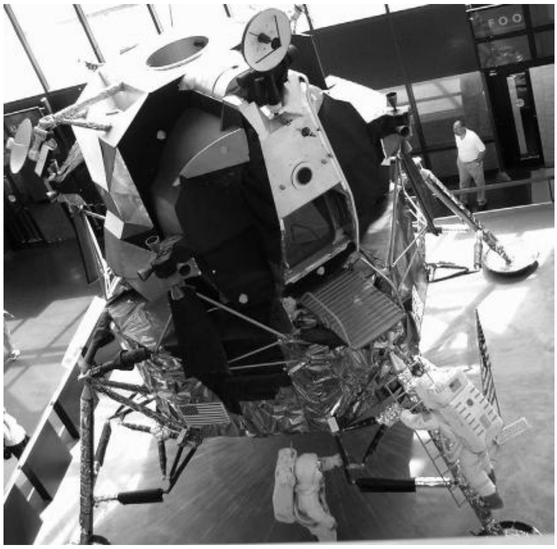

FIGURE 14.16
Eagle moon Lander.

FIGURE 14.17
Skylab, exterior view.

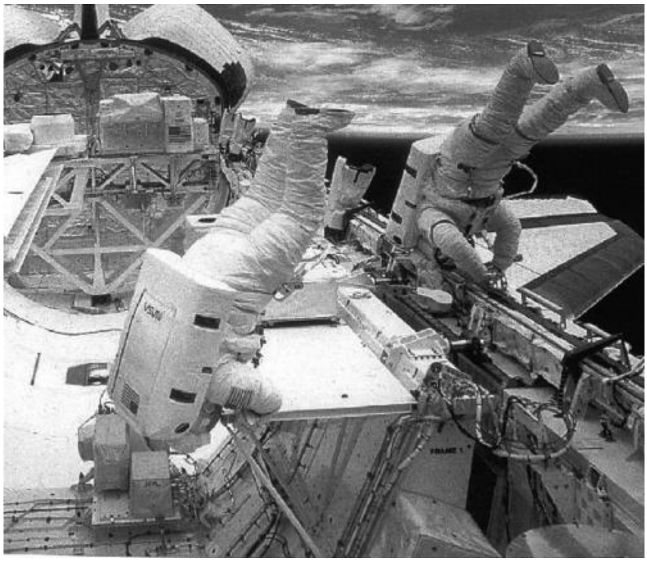

FIGURE 14.18
Astronauts working outside a space shuttle. (NASA photograph in the public domain.)

1970s. During this time there was undoubtedly much to be done to prepare the space shuttles for the next decade.

Space shuttles that were launched in the 1980s and 1990s typically involved seven-person crews. For the first time women and civilians were included in the personnel rosters. Both the U.S. and Russian space programs during this period included personnel from other interested spacefaring countries that did not have an operational space program of their own, short of launching some telecommunications satellites. Space shuttle flights have both scientific and commercial objectives, such as conducting biological and chemical experiments where microgravity is a variable and repairing or reorienting a telecommunications satellite. Figure 14.18 depicts astronauts working outside a space shuttle using jetpacks for mobility.

The 1990s and the 21st century saw the advent of permanently manned space stations, beginning with the Russian project Mir. An international space station is currently operational. Meanwhile, the United States has been sending robotic missions to Mars, and manned stations on the moon and Mars are being considered.

Overview of Human Concerns

The term *human factors* in this context means everything that is not hardware engineering and would include medical health and wellness, psychological or emotional health and work behavior, sociology of space communities, culture and living patterns, finances, earthbound politics, space law, specialized education and training, management, and communication systems development (Kring, 2001; Harris, 1989, 1992; Suedfeld, 2003). A comprehensive systems view of space travel is essential for any further development. Another outline of factors that contained items that are more germane to the core of human factors and ergonomics (Christensen & Talbot, 1986; Harris, 1992) includes limits of perceptual and motor performance, issues connected to emotional adjustment and social stress, spacecraft habitability, work–rest and sleep cycles, and hazards and emergency management.

Personnel Selection

When the space program first started in the United States in 1958, NASA considered the possibility that the ideal astronauts would be drawn from the small population of Air Force experimental aircraft pilots. Those individuals had already traveled beyond the speed of sound and into the stratosphere. As it turned out, the experimental aircraft pilots who were available at the time did not want the job (Dewaard & Dewaard, 1984). The first astronauts were selected from Air Force personnel who met the highest possible standards for physical and psychological capabilities as they were known at the time. The thinking was that if the top seven could not cope with a few days in space, no one could.

By the late 1970s enough information had accumulated from Antarctic expeditions to allow for some plausible generalizations to outer space personnel. Antarctic conditions are not the same as those found in outer space, but they are perhaps the most extreme terrestrial environments, featuring thin air, social isolation, and severe diurnal disruption. At the time of this writing, there are as many as 10,000 people living in Antarctica during its 6 months of summer, but only as few as 400 in its winter. The proportions alone suggest the level of environmental challenge before retelling the stories of the famous explorers of the early 20th century. The personality characteristics of the successful Antarctican include independence, task orientation, and achievement motivation. One group of successes is gung ho about the importance of their work, whereas another group manages to think of their work as just another 9-to-5 job (Natani, 1980).

Antarcticans who winter over are prone to *dysadaptation syndrome*. U.S. personnel who have fallen victim to this malady show all symptoms comparable to clinical depression on the Minnesota Multiphasic Personality Inventory, which is one of a few standard clinical assessment instruments. The Russian interpretation of their dysadaptation casualties was "cardiovascular insufficiency" (Natani, 1980). In recent times, some cultural differences in response to extreme environments have come to the foreground. To make a long story short, successful soldiers in severe combat, underwater divers, and bomb-disposal operators displayed lower levels of physiological activation when confronted with stressful situations than the less successful personnel. Thus, Halgrave and Colvin (1998) recommended the use of psychophysiological testing to screen personnel for work in extreme conditions. In the case of space personnel, it now appears that Russian personnel display better levels of adjustment than U.S. personnel (Boyd, 2001); differences in selection procedures may be explicatory, but follow-up studies are needed before drawing any conclusions.

There are issues beyond adjustment to the physical environment. A space crew is cooped up with each other in a small space for an extended period. Kanas et al. (2002) recommended that crews be selected as a group with attention to mutual compatibility, rather than individually. Psychological factors that can be determined during the selection process would include decision-making styles and preferences, work habits and leadership preferences, and communication patterns (Kring, 2001). Facility in a common language is also highly desirable.

Gravitational Forces

Although outer space travel is associated with microgravity, there are excessive gravitational forces on the operators during acceleration, especially at the levels required to break free of the Earth's orbit. Large gravitational forces are sometimes found in other unusual environments as well. Extreme forces can be harmful and painful; thus, anti-G suits were developed to give operators reasonable protection. Even with anti-G protection, human tolerance is only a matter of minutes and depends on the direction of the thrust. For 6 Gs pushing at an operator's chest, the average endurance time is 550 s; endurance drops to 250 s for a thrust of 10 Gs. For 6 Gs pushing at an operator's back, the average endurance is 350 s, and endurance drops to 100 s for a thrust of 10 Gs. For a downward thrust of 6 Gs, the average endurance is 150 s and drops to 0 s for a thrust of 9 Gs (Chambers, 1963).

The gravitational forces that impinge on an astronaut during takeoff have important effects that start under less extreme conditions than the exposure limits just described. As acceleration starts, blood begins to pool and drain from the upper head region to the lower body extremities. The result is a near blackout when acceleration reaches about 4 Gs. Operators do not lose consciousness during this time, but perception and motor coordination are impaired under these high-G conditions: "Linear and angular accelerations themselves produce vestibular, kinesthetic, tactual, and visual cues that, when compared against normal 1 G experiences, appear to be unusual, unique, and misleading" (Chambers, 1963, p. 218). The anti-G suit worn by the astronauts is a technical response to this phenomenon. The suits counteract the changes in blood flow by applying pressure to the lower extremities and preventing blood from flowing too far downward. Another technical response is to rely on automatic controllers during this critical phase of takeoff when the astronauts are experiencing psychomotor deficits from the G-forces.

Allocation of Function

The initial plan for the Mercury spacecraft allocated functions between mission control personnel and the equipment that they operated. The astronauts were expected to sit in their chairs and not do much of anything. The Mercury prototype did not even have a window. The astronaut team in training strongly objected, insisting that they wanted a set of controls in case something happened. They also figured if they were going to be shot into space, they wanted to look out the window, see where they were, and take a few pictures.

The spaceflight system quickly adopted a more complex view of the person–machine system whereby ground control would work its equipment and remotely control the spacecraft where possible, and the astronauts would control the spacecraft in all other circumstances. As it turned out, the astronauts had the right idea; ground control lost communication contact with John Glenn's Mercury capsule during the third spaceflight, and Glenn had to find his way back to Earth on his own.

The marked allocation of functions between the space crew and the ground control has its assets and limitations, especially in emergency situations. Apollo 13, which was meant to be a third trip to the moon, had to be turned around and returned to Earth because of life-threatening technological failures. Engineers in ground control were able to work out solutions to some of the problems and report back to the astronauts who were working on other aspects. Another astronaut worked in a spaceflight simulator in the ground station to figure out the successful control sequence that was needed to return the Apollo to Earth.

Space station personnel had some less fortunate experiences. The first Skylab mission met with some difficulty when ground control asked the astronauts to carry out some unplanned work activities. The astronauts experienced considerable difficulty convincing ground control of the adverse conditions under which they were working that made the requests impossible. The result was that Skylab refused to talk to ground control for an entire day (Bechtel, 2003).

Cosmonauts on Mir reported that their mission control often blamed them unfairly for things that went wrong, such as a fire and an atmosphere leakage. To compound the problem, Mir was operated by a private enterprise that provided bonuses for the cosmonauts for projects that went right, and pay reductions when something went wrong. Cosmonauts thus adopted the strategy of minimizing negative reports (Bechtel, 2003).

Anthropometry

Anthropometric issues showed up early. The first spacecraft were designed for a specially targeted group of seven astronaut trainees. Although a relatively personalized fit could be maintained for the operator's workstation and environment suit, designers did not realize until after the first spaceflight that the human body grows about 2 in. in zero gravity because the pressure on the spinal column is gone. Thus, design allowances were needed for subsequent flights (Louviere & Jackson, 1980).

Once in flight and allowed to move around the spacecraft, the floating human body takes on a semifetal position known as the *neutral body position*, as shown in Figure 14.19. This position results in a decline in the line of sight from –15° to –35°. Control panels need to be oriented accordingly (Louviere & Jackson, 1980).

The lack of gravitational forces in space results in loss of muscle strength. The impact is most clear after 4 weeks in space and continues through the first 3 months. Different parts of the body are affected to different extents, with strength losses of up to 20% in the legs and 10% in the biceps (Fujii & Patten, 1992). Astronauts thus spend part of their day on an exercise treadmill to build back their strength. To use the treadmill, the operator must be anchored with cords to prevent floating away (Joels, Kennedy, & Larkin, 1982). Cosmonauts reportedly once spent 2.5 hr per day exercising (Fujii & Patten). Comparable exercise rates from the United States are not available, except to say that some do engage

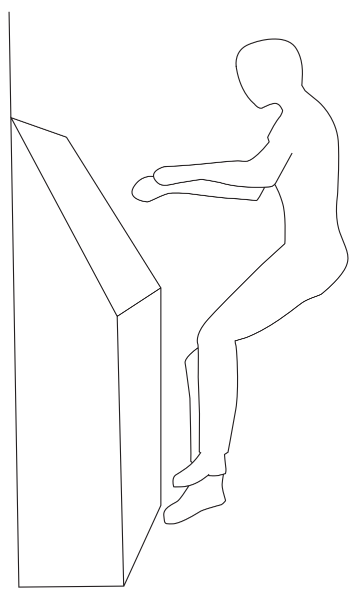

FIGURE 14.19
The neutral body position.

in exercise and some do not (Bacal, Billica, & Bishop, 2003); the duration of the mission is a factor that affects the astronauts' daily schedules.

Vision

Since the earliest spaceflights, astronauts have reported flashes of light that were not really there. The visual flashes are caused by radiation that depolarizes retinal neurons (Fujii & Patten, 1992).

The perception of *up* and *down* is significantly impaired during spaceflight. Visual cues that denote the relationship of objects to other objects are disrupted because the usual reference points, which are gravitationally centered, are not operating. Astronauts compensate for the disruption in visual information by relying on the long axis of their own bodies as reference points (Fujii & Patten, 1992).

Vestibular Sense and Motor Control

The disorganization of visual cues disrupts the coordination of the cerebellum and inner ear, which are responsible for body balance and motor coordination (*otolithic system*). The most common physical complaints reported by astronauts after they return to earth are clumsiness in movements, difficulty walking in a straight line, vertigo while walking or standing, nausea, and difficulty concentrating (Bacal et al., 2003; Hlavacka, Dzurková, & Polónyová, 2000). Some of these effects were ameliorated by in-flight exercise.

Disruption of the otolithic system is the primary explanation for *space adaptation syndrome*. Its symptoms include headache, nausea, vomiting, and sensitivity to head and body motion (Fujii & Patten, 1992, p. 1004). It is reported by 67% of astronauts shortly after take-off and tapers off after 4 days in space. Its symptoms often return after the return to Earth as the body must readjust itself once again. Space adaptation syndrome can be countered with medication that is taken minutes before blastoff. Seasoned space travelers are less likely to experience its effects, but they are not immune to them. The same malady is also known to occur as a result of using spaceflight or virtual reality simulators; in those contexts, it is known as *simulator sickness*.

Microgravity produces some sensorimotor impairments in pointing and tracking tasks. Pointing movements are slowed in tasks that involve a range of whole arm and finger movements (Bock et al., 2001). The primary culprit is thought to be the disruption of visual feedback, particularly when one movement ends and the next is meant to begin.

Tracking tasks performed with a computer screen and joystick show increased errors during spaceflight. The impairment shows up around the fourth day and extends to the 16th day then appears to subside. Impairments return after the cosmonaut has returned to Earth for about another 10 days (Manzey, Lorenz, Heuer, & Sangals, 2000). The tracking effect is explained in part by disrupted circadian rhythms, disrupted visual cues, and microgravity itself, resulting in a miscalculation of one's muscular forces while working the controls (Heuer, Manzey, Lorenz, & Sangals, 2003).

Sleep

The loss of the 24-hr day–night cues on earth promotes disruption of sleep onset and duration. Sleep reduces to 6 hr per night during the first 30 days of spaceflight, then extends to 7 hr per night afterward. Deep sleep shortens proportionately. Six-hour sleeps do not promote performance problems, but performance problems tend to occur after one or two nights of 5 hr sleep or less (Gundel, Drescher, & Polyakov, 2001).

Space Habitats

The first capsules were barely large enough to hold a human, and were not what anyone would call a *habitat*. The two-person Gemini capsules were similarly proportioned, but physical crowdedness now became social crowdedness as well. Analysis of recordings of Gemini astronauts' conversations indicated that they oscillated between times of talking socially and quietude, apparently for privacy (Sieber, 1980). It is interesting that these two needs had to be met through strictly psychological means; it was not possible to say "good night" and go to one's own room.

The living environment in space offers only a limited amount of space per person and barely enough room to accomplish all the daily personal tasks during extended missions. It is a closed system rather than an open system. Not only is the environment computer controlled, the computers monitor the humans unusually closely. In the earlier chapter on workspace design, the concern was to have enough space to carry out all the intended work movements. Social space was not an issue, as it was assumed that the worker could go home to whomever and whatever. The early space travelers trained together and otherwise knew each other fairly well before their mutual space missions. The space personnel of the future are likely to be more diverse with respect to national origin and cultural habits. Additional space must be allocated to accommodate their differences in food preparation, social habits, recreation, religious observances, and habitat aesthetics (Kring, 2001).

The living quarters for the three astronauts of Skylab are shown in Figure 14.20. The space was equivalent to a 90-m^2 (about 825-ft^2) apartment (Caprara, 2000). The four areas are for eating and sleeping, toilet facility, food preparation, and physical activity and medical needs. It is possible to walk through Skylab at the National Aeronautics and Space Museum in Washington, DC. These minimal facilities accommodated the astronauts for a maximum of 84 days. Space stations such as the International Space Station require larger layouts for larger crews and longer term living.

Space toilets operate on a vacuum system. Each droplet and particle must be properly accounted for and not allowed to float around. The equipment compels space travelers to relearn body functions that they thought they had mastered at age 2 years.

Showers must be taken in closed bags. At the time of Skylab, the process of setting up the shower and dissipating the used water required about 2.5 hr (Caprara, 2000). Residents of Mir resorted to using large antiseptic wipes instead of soap and water for stretches of up to 5 months (Bechtel, 2003).

Space stations can receive supplies of food and other necessities from shuttles, but longer term intergalactic flights present a problem with food (assuming one knows where one is going). Some thought has been given to developing a closed ecosystem wherein the astronauts raise food crops, eat, and recycle waste. Neither the biochemistry nor the agricultural requirements have been fully worked out, and there is a concern that traces of toxins will build up to critical levels after too many cycles have taken place (Bechtel, 2003; Modell & Spurlock, 1980; Rambaut, 1980).

Between the two extremes, we have the colonization of Mars. Because of the varying distance between the two planets as they orbit the sun, it is only possible to send flights from Earth to Mars every 26 months. Each 26-month period is somewhat different in the

FIGURE 14.20
Living areas in Skylab.

distance calculations, which translate into timing calculations for different parts of the mission. Missions are expected to run upward of 900 days, which would be approximately twice the length of the longest mission in space to date. Time and distance create a challenge for the transportation of building supplies, food, and water, and for communication. It is likely that space stations will play important roles in the development and maintenance of a Martian colony. Figure 14.21 shows two views of Martian landscapes taken from drone spacecraft.

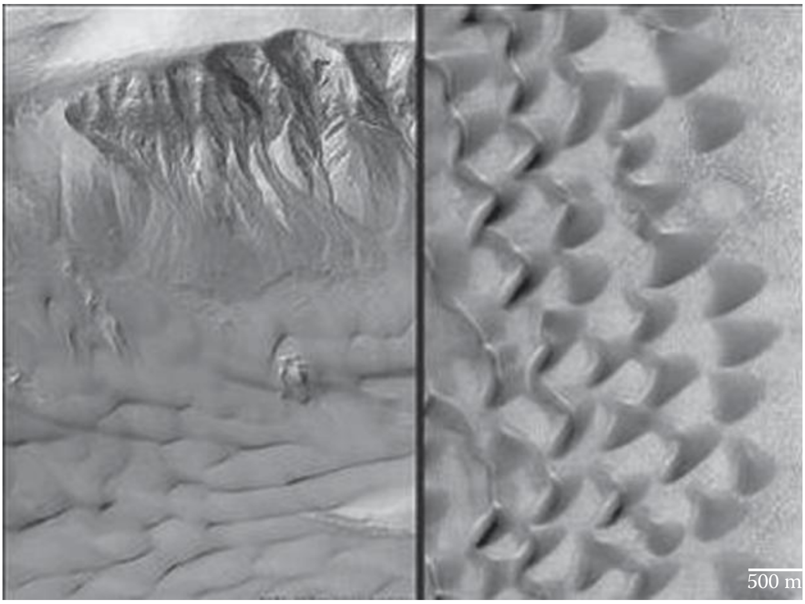

FIGURE 14.21
Martian landscapes. (NASA photograph in the public domain.)

DISCUSSION QUESTIONS

1. Where else on the stove could you locate the controls for burners? Describe the assets and limitations of the possible configurations.
2. How can the design of a facility such as a factory be a source of stress in the workplace? How about a high-rise apartment building?
3. Compare and contrast the problem of navigation through a large building complex and navigation through a web site.
4. Suggest a role for insurance companies to enhance roadway safety. Hint: If risk homeostasis theory is not convincing, consider what works in the prevention of occupational accidents.
5. What are some of the current traffic laws that pertain to the use of onboard computers, telephones, or geographic information systems?
6. Motor vehicles have been equipped with a middle rear brake light that is usually mounted in the rear windshield since 1985. Why was this light thought to be necessary? How effective is this device?
7. Consider this piece of accident litigation: A car is proceeding down a city street in the left lane of two lanes. The posted speed limit is 35 mph. A 10-year-old boy enters the street from the sidewalk on the right in an attempt to cross the street and is struck by the car. The car drags the boy 41 ft before stopping. The driver, who is in an emotional state at this time, puts the car in reverse and runs over him again. The boy has substantial medical expenses associated with this accident, and is suing the driver's insurance company. The insurance company's position is that they are not obligated to pay because the accident was the boy's fault. Analyze the human factors involved. What pieces of technical information are required to determine whether substantial fault resides with the driver?

References

Abbott, A., Button, C., Pepping, G.-J., & Collins, D. (2005). Unnatural selection: Talent identification and development in sport. *Nonlinear Dynamics, Psychology, and Life Sciences, 9*, 61–88.

Abedin, S., Lewis, M., Brooks, N., Owens, S., Scerri, P., & Sycara, K. (2011). SUAVE: Integrating UAV using a 3D model. *Proceedings of the Human Factors Society, 55*, 91–95.

Ackerman, P. (1987). Individual difference in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3–27.

Ackerman, P. L. (Ed.). (2011). *Cognitive fatigue*. Washington, DC: American Psychological Association.

Adams, A. E., Rogers, W. A., & Fisk, A. D. (2012). Choosing the right task analysis tool. *Ergonomics in Design, 20*(1), 4–10.

Adams, J. A. (1989). *Human factors engineering.* New York: Macmillan.

Adams, J. S., & Trucks, L. (1976). A procedure for evaluating auditory warning signals. *Proceedings of the Human Factors Society, 20*, 166–172.

Adams, M. A., & Dolan, P. (2005). Spine biomechanics. *Journal of Biomechanics, 38*, 1972–1983.

Agervold, M. (1976). Shiftwork: A critical review. *Scandinavian Journal of Psychology, 17,* 181–189.

Aggarwal, C. C., & Yu, P. S. (Eds.). (2008). *Privacy-preserving data mining: Models and algorithms*. New York: Springer.

Ahrentzen, S., Levine, D. W., & Michelson, W. (1989). Space, time, and activity in the home: A gender analysis. *Journal of Environmental Psychology, 9,* 89–101.

Aiello, J. R., & Kolb, K. J. (1995). Electronic performance monitoring: A risk factor for workplace stress. In S. L. Sauter & L. R. Murphy (Eds.), *Organizational risk factors for stress* (pp. 163–180). Washington, DC: American Psychological Association.

Akerstedt, T. (1977). Inversion of sleep–wakefulness pattern: Effects on circadian variations in psychophysical activation. *Ergonomics, 10,* 459–474.

Allen, P. A. (2009). Complexity, evolution, and organizational behavior. In S. J. Guastello, M. Koopmans, & D. Pincus. (Eds.), *Chaos and complexity in psychology: Theory of nonlinear dynamical systems* (pp. 452–474). New York: Cambridge University Press.

Alves, E. E., & Kelsey, C. M. (2010). Combating vigilance decrement in a single-operator radar platform. *Ergonomics in Design, 18*(2), 6–9.

Alzheimer's Association. (2010). Alzheimer's disease facts and figures. Retrieved from http://www.alz.org/alzheimers_disease_facts_and_figures.asp

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human–computer speech interaction. *International Journal of Man–Machine Studies, 38*, 547–566.

American Civil Liberties Union. (1999). *Drug testing: A bad investment.* New York: Author.

Amirabdollahian, F., Loureiro, R., & Harwin, W. (2002). Minimum jerk trajectory control for rehabilitation and haptic applications. In *Proceedings of the 2002 IEEE-RAS Conference, Washington, DC* (pp. 11–15). Los Alimitos, CA: IEEE.

Anderson, P., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice, 10,* 240–247.

Andreadis, N., & Quinlan, P. T. (2010). Task switching under predictable and unpredictable circumstances. *Attention, Perception, and Psychophysics, 72*, 1776–1790.

Andreassi, J. L. (1980). *Psychophysiology.* New York: Oxford Press.

Angell, L. S., & Lee, J. D. (2011). Science and technology of driver distraction: Introduction to the special issue. *Ergonomics in Design, 19*(4), 3.

Anthony, K. H., & Watkins, N. J. (2002). Exploring pathology: Relationships between clinical and environmental psychology. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 129–146). New York: Wiley.

Ao, S.-I. (2008). *Data mining and applications in genomics*. Dordrecht, the Netherlands: Springer.

Arecchi, F. T. (2011). Phenomenology of consciousness: From apprehension to judgment. *Nonlinear Dynamics, Psychology, and Life Sciences, 15*, 359–376.

Arthur, W., Jr., Barrett, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance, 4,* 89–105.

Arthur, W., Jr., Barrett, G. V., & Doverspike, D. (1990). Validation of an information-processing-based test battery for the prediction of handling accidents among petroleum-product transport drivers. *Journal of Applied Psychology, 75,* 621–628.

Arthur, W., Jr., & Doverspike, D. (1992). Locus of control and auditory selective attention as predictors of driving accident involvement: A comparative longitudinal investigation. *Journal of Safety Research, 23,* 73–80.

Ash, J. E. (1914). Fatigue and its effects upon control. *Archives of Psychology, 31*, 1–61.

Ashby, W. R. (1956). *Introduction to cybernetics.* New York: Wiley.

Atkins, W. B., Wood, R. E., & Rutgers, P. J. (2002). The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes, 88*, 587–604.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Baber, C. (2008). Evaluating mobile human–computer interaction. In J. Lumsden (Ed.), *Handbook of research on user interface design and evaluation for mobile technology* (pp. 731–743). Hershey, PA: Information Science Reference.

Baber, C., & Noyes, J. M. (2001). Speech control. In K. Baumann & B. Thomas (Eds.), *User intercase design for electronic appliances* (pp. 190–208). New York: Taylor & Francis.

Bacal, K., Billica, R., & Bishop, S. (2003). Neurovestibular symptoms following space flight. *Journal of Vestibular Research, 13,* 93–102.

Bahr, G. S., & Ford, R. A. (2011). How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior, 27*, 776–763.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors, 48*, 693–709.

Baillargeon, J., & Davis, C. (1984). Barnum meets the computer: A critical test. *Journal of Personality Assessment, 48,* 418–419.

Bak, P. (1996). *How nature works.* New York: Springer.

Balagué, N., Hristovski, R., Agagonés, D., & Tenenbaum, G. (2012). Nonlinear model of attention focus during accumulated effort. *Psychology of Sport and Exercise, 13*, 591–597.

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Perception & Psychophysics, 73*, 640–657.

Baldwin, C. L., & Struckman-Johnson, D. (2002). Impact of speech presentation level on cognitive task performance: Implications for auditory display design. *Ergonomics, 45,* 41–74.

Bales, R. F., & Cohen, S. P. (1979). *SYMLOG: A system for the multiple level observation of groups*. New York: Free Press.

Balkin, T. J., Horrey, W. J., Graeber, R. C., Czeisler, C. A., & Dinges, D. F. (2011). The challenges and opportunities of technological approaches to fatigue management. *Accident Analysis and Prevention, 43*, 565–572.

Bankes, S., & Lempert, R. (2004). Robust reasoning with agent-based modeling. *Nonlinear Dynamics, Psychology, and Life Sciences, 8*, 231–258.

Banks, A. P., & Millward, L. J. (2000). Running shared mental models as a distributed cognitive process. *British Journal of Psychology, 91,* 513–531.

Barabási, A.-L. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus.

Baram, M. (2001). Legal borders necessary for a safe genetic technology. In J. Aithoff (Ed.), *Safety of modern technical systems* (pp. 645–652). Saarbrücken, Germany: TÜV-Saarland Foundation.

Barnes, C. M., Hollenbeck, J. R., Wagner, D. T., DeRue, D. S., Nahgang, J. D., & Schwind, K. M. (2008). Harmful help: The costs of backing-up behavior in teams. *Journal of Applied Psychology, 90*, 529–539.

Barofsky, G. L., & Smith, M. (1993). Reductions in turnover, accidents, and absenteeism: The contribution of a pre-employment screening inventory. *Journal of Clinical Psychology, 49,* 109–116.

Bar-Yam, Y. (1997). *Dynamics of complex systems.* Reading, MA: Addison-Wesley.

Baumann, M. R., Sniezek, J. A., & Buerkle, C. A. (2001). Self-evaluation, stress, and performance: A model of decision making under acute stress. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 139–158). Mahwah, NJ: Lawrence Erlbaum Associates.

Bechtel, R. B. (2003). On to Mars! In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 676–684). New York: Wiley.

Bedard, M., Guyatt, G. H., Stones, M. J., & Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident Analysis and Prevention, 34,* 717–728.

Benner, L., Jr. (1975). Accident investigation: Multinational events sequencing methods. *Journal of Safety Research, 7,* 67–73.

Bennett, K. B., & Flach, J. M. (2011). *Display and interface design: Subtle science, exact art*. Boca Raton, FL: CRC Press/Taylor & Francis.

Bennett, K. B., & Flach, J. M. (2013). Configural and pictorial displays. In J. D. Lee & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 517–533). New York: Oxford University Press.

Bergamasco, M., Bardy, B., & Gopher, D. (2013). *Skill training in multimodal virtual environments*. Boca Raton, FL: CRC Press/Taylor & Francis.

Berlin, S., Reagan, I. J., & Bliss, J. P. (2012). Participant characteristics and speeding behavior during an advisory warning and cash incentive intervention. *Proceedings of the Human Factors and Ergonomics Society, 56*, 1044–1048.

Bernstein, N. (1967). *The coordination and regulation of movements.* Oxford, UK: Pergamon.

Berringer, D. B. (1978). Collision avoidance response stereotypes in pilots and non-pilots. *Human Factors, 20,* 529–536.

Bertin, J. (1981). *Graphics and graphic information processing.* Berlin, Germany: de Gruyter.

Betrancourt, M., & Tversky, B. (2000). Effect of computer animation on users' performance: A review. *Travail Humain, 63,* 311–329.

Beveles, A. (1948). A mathematical model for group structures. *Human Organization, 7*, 16–30.

Billinghurst, S. S., Morgan, C., Rorie, R. C., Kiken, A., Bacon, L. P., Vu, K.-P. L., Strubel, T. Z., & Battiste, V. (2011). Should students learn general sir traffic control management skills before NextGen tools. *Proceedings of the Human Factors and Ergonomics Society, 56*, 16–20.

Bills, A. G. (1931). Blocking: A new principle in mental fatigue. *American Journal of Psychology, 67*, 230–245.

Binik, Y. M., Servan-Schreiber, D., Freiwald, S., & Hall, K. S. (1988). Intelligent computer-based assessment and psychotherapy: An expert system for sexual dysfunction. *Journal of Nervous and Mental Disease, 176,* 387–400.

Blankenburger, S., & Hahn, K. (1991). Effects of icon design on human–computer interaction. *International Journal of Man–Machine Studies, 35,* 363–377.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human–Computer Interaction, 4*, 11–44.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*, 175–178.

Bock, O., Fowler, B., & Comfort, D. (2001). Human sensorimotor coordination during spaceflight: An analysis of pointing and tracking responses during the "Neurolab" space shuttle mission. *Aviation, Space, and Environmental Medicine, 72,* 877–883.

Bond, C. F. Jr., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 265–292.

Bond, S., & Cooper, S. (2006). Modeling emergency decisions: Recognition-primed decision making. The literature in relation to an ophthalmic critical incident. *Journal of Clinical Nursing, 15*, 1023–1032.

Borges, L. A. J., & Guastello, S. J. (1998). Chaos theory applied to TQM: A survey. In W. L. Baker (Ed.), *Proceedings of the 52nd Annual Quality Congress* (pp. 578–585). Philadelphia, PA: American Society for Quality Control.

Botella, J., Barriopedro, M. L., & Suero, M. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1452–1467.

Bousfield, W. A. (1953). The recurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229–240.

Boyd, K. (2001). Psychological, emotional studies of Mir Space Station missions show Russians fared better than Americans. *Journal of Human Performance in Extreme Environments, 5,* 96–97.

Braarud, P. O. (2001). Subjective task complexity and subjective workload: Criterion validity for complex team tasks. *International Journal of Cognitive Ergonomics, 5,* 261–273.

Brachman, R. (1985). "I lied about the trees," or, defaults and definitions in knowledge representation. *A. I. Magazine, 6,* 80–93.

Branche, C. M., Sniezek, J. E., Sattin, R. W., & Mirkin, I. R. (1991). Water recreation-related spinal injuries: Risk factors in natural bodies of water. *Accident Analysis and Prevention, 23,* 13–18.

Brannick, M. T., Prince, A., Prince, P., & Salas, E. (1995). The measurement of team process. *Human Factors, 37,* 641–651.

Brannick, M. T., Roach, R. M., & Salas, E. (1993). Understanding team performance: A multimethod study. *Human Performance, 6,* 287–308.

Braun, T. W. (2008). Prevention through design (PtD) from the insurance perspective. *Journal of Safety Research, 39,* 137–139.

Brave, S. B. (2003). Agents that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Dissertation Abstracts International, 64*(5-B), 2437 (UMI No. 2003-95022-206).

Breeden, J. L., Dinkelacker, F., & Hubler, A. (1990). Noise in the modeling and control of dynamical systems. *Physics Review A, 42,* 5827–5836.

Brehmer, B. (1987). Development of mental models for decision in technological systems. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 111–120). New York: Wiley.

Brehmer, B. (2005). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomic Science, 6,* 73–94.

Brewer, C. A. (2006). Basic mapping principles for visualizing cancer data using geographic information systems (GIS). *American Journal of Preventive Medicine, 30,* 525–536.

Brewer, C. A., Hatchard, G. W., & Harrower, M. A. (2003). ColorBrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Systems, 30,* 5–32.

Bronkhorst, A. W., Veltman, J. A., & van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors, 38,* 23–33.

Bruce, C. J., & Atkins, F. J. (1993). Efficiency effects of premium-setting regimes under Workers' Compensation: Canada and the United States. *Journal of Labor Economics, 77*(Suppl.), S38–S69.

Brunnstein, K. (2001). About actual threats from the Internet: About viruses, worms, Trojan horses and other computer beasts. In J. Aithoff (Ed.), *Safety of modern technical systems* (pp. 585–586). Saarbrücken, Germany: TÜV-Saarland Foundation.

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of telegraphic language. *Psychological Review, 4,* 27–53.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345–375.

Buchholz, R. A. (1977). The belief structure of managers relative to work concepts measured by a factor analytic model. *Personnel Psychology, 30,* 567–587.

Buck, J. R., & Kantowitz, B. H. (1983). Decision making with applications to inspection and maintainability. In B. H. Kantowitz & R. L. Sorkin (Eds.), *Human factors: Understanding people-system relationships* (pp. 428–453). New York: Wiley.

Bunn III, W. B., Pikelny, D. B., Slavin, T. J., & Paralkar, S. (2001). Health, safety, and productivity in a manufacturing environment. *Journal of Occupational and Environmental Medicine, 43,* 47–55.

Buntine, W. (1987). Induction of Horn clauses: Methods and the plausible generalization algorithm. *International Journal of Man–Machine Studies, 26,* 499–519.

Buntine, W., & Stirling, D. (1988). *Induction as interaction* (Tech. Rep.). NSW, Australia: University of New South Wales.

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology* (2nd ed.). New York: Wiley Interscience.

Burt, R. (2004). Structural holes and good ideas. *American Journal of Sociology, 110,* 349–399.

Bush, T. R., & Hubbard, R. P. (2008). A comparison of four office chairs using biomechanical measures. *Human Factors, 50*, 629–642.
Bushby, S. T. (2011, November). Information model standard for integrating facilities with smart grid. *BACnet Today & the Smart Grid, ASHRAE Journal, nv*, B18–B22.
Butler, R. J. (1994). *Safety incentives in Workers' Compensation: A review of the evidence and some new findings* (Tech. Rep.). Minneapolis, MN: City University of Minnesota, Industrial Relations Center.
Campion, M. A. (1983). Personnel selection for physically and demanding jobs: Review and recommendations. *Personnel Psychology, 36*, 527–550.
Campion, M. A., Papper, E. M., & Medsker, G. J. (1996). Relations between work team characteristics and effectiveness: A replication and extension. *Personnel Psychology, 49*, 429–452.
Cannon-Bowers, J., Salas, E., & Converse, S. (1990). Cognitive psychology and team training: Training shared mental models of complex systems. *Human Factors Bulletin, 33*, 1–4.
Cannon-Bowers, J. Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Hillsdale, NJ: Lawrence Erlbaum Associates.
Cao, L. (Ed.). (2009). *Data mining and multi-agent integration*. New York: Springer.
Cao, L., Ni, J., Wang, J., & Zhang, C. (2004, September). Agent services-driven plug and play in the F-TRADE. In *17th Australian Joint Conference on Artificial Intelligence* (pp. 917–922). New York: Springer.
Caplin, S. (2001). *Icon design*. New York: Watson-Guptill.
Caprara, G. (2000). *Living in space: From science fiction to the International Space Station.* Buffalo, NY: Firefly Books.
Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601–613.
Card, S., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Carlson, L. A., Holscher, C., Shipley, T. F., & Dalton, R. C. (2010). Getting lost in buildings. *Current Directions in Psychological Science, 19*, 284–289.
Carlsson, H., & van Damme, E. (1993). Equilibrium selection in stag hunt games. In K. Binmore, A. Kirman, & P. Tani (Eds.), *Frontiers of game theory* (pp. 237–253). Cambridge, MA: MIT Press.
Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors, 54*, 747–761.
Carter, N., & Menckel, E. (1985). Near-accident reporting: A review of Swedish research. *Journal of Occupational Accidents, 7*, 61–64.
Carter, R. C., Jr. (1978). Knobology underwater. *Human Factors, 20*, 641–647.
Carpman, J. R., & Grant, M. A. (2002). Wayfinding: A broad view. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 427–442). New York: Wiley.
Carroll, B. (1987). Expert systems for clinical diagnosis: Are they worth the effort? *Behavioral Science, 32*, 274–292.
Carroll, J. M., & Rosson, M. B. (1987). Paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction* (pp. 80–111). Cambridge, MA: MIT Press.
Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semi-automated driving: Does it matter what aspects are automated? *Human Factors, 54*, 747–761.
Casner, S. M. (1994). Understanding the determinants of problem-solving behavior in a complex environment. *Human Factors, 36*, 580–596.
Catrambone, R., & Seay, A. F. (2002). Using animation to help students learn computer algorithms. *Human Factors, 44*, 495–511.
Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*, 63–65.
Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the sixteen personality factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.

Centers for Disease Control and Prevention. (2007). Number of deaths for leading causes of death. Retrieved from http://www.cdc.gov/nchs/fastats/lcod.html

Chaffin, D. B., Herrin, G. D., Keyserling, W. M., & Foulke, J. A. (1977). *Pre-employment strength testing in selecting workers for materials handling jobs* (Report No. CDC 99-74-62). Cincinnati, OH: National Institute for Occupational Safety and Health, Physiology and Ergonomics Branch.

Chambers, R. M. (1963). Operator performance in acceleration environments. In N. M. Burns, R. M. Chambers, & E. Hendler (Eds.), *Unusual environments and human behavior* (pp. 193–320). New York: Free Press.

Chandrasekaran, B., & Goel, A. (1988). From numbers to symbols to knowledge structures: Artificial intelligence perspectives on the classification task. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*, 415–424.

Chapanis, A. (1975). Engineering psychology. In M. Dunnette (Ed.), *Handbook of industrial–organizational psychology* (pp. 697–744). Chicago: Rand McNally.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72*, 271–280.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271–280.

Cho, A., & Histon, J. (2012). Factors affecting the learning of a new air traffic control sector for experienced air traffic controllers. *Proceedings of the Human Factors and Ergonomics Society, 56*, 90–94.

Chen, B. (1999). *Needle in a haystack.* Retrieved from www.baddesigns.com/manylts.html

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Chien, J. M., & Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science, 21*, 78–84.

Christensen, J. M., & Talbot, J. M. (1986). Psychological aspects of space flight. *Aviation, Space, and Environmental Medicine, 57*, 203–212.

Chueng, J. (1983). Interdependence and coordination in organizations. A rule–system analysis. *Academy of Management Journal, 26*, 156–162.

Cigler, B. A. (2007). The "big questions" of Katrina and the 2005 great flood of New Orleans. *Public Administration Review, 67*, 64–76.

Clark, T. B. (1975, May). Common shortcomings in facilities design. *Industrial Management, X*, 1–4.

Clarke, S. (2006). The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Occupational Health Psychology, 11*, 315–327.

Clayton, K., & Frey, B. B. (1997). Studies of mental "noise." *Nonlinear Dynamics, Psychology, and Life Sciences, 1*, 173–180.

Cloninger, C. (2002). *Fresh styles for web designers: Eye candy from the underground*. Indianapolis, IN: New Riders.

Cohen, P. R., & Howe, A. E. (1989). Toward AI research methodology: Three case studies in evaluation. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 634–646.

Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin, 88*, 82–108.

Collins, J. (2008). *Nonlinear dynamics and enhanced sensorimotor function*. Retrieved from http://www.labome.org/grant/r01/hd/nonlinear/dynamics/nonlinear-dynamics-and-enhanced-sensorimotor-function

Colquhoun, W. P. (1971). *Biological rhythms and human performance.* London: Academic Press.

Colquhoun, W. P. (1975). Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Human Factors, 17*, 425–437.

Comer, D. R. (1995). A model of social loafing in real work groups. *Human Relations, 48*, 647–661.

Comfort, L. (1996). Self-organization in disaster response: Global strategies to support local action. In G. Koehler (Ed.), *What disaster response management can learn from chaos theory* (pp. 94–112). Sacramento, CA: California Research Bureau, California State Library.

Comfort, L. (1999a). Nonlinear dynamics in disaster response: The Northridge, California earthquake, January 17, 1994. In L. D. Kiel & E. Elliott (Eds.), *Nonlinear dynamics, complexity, and public policy* (pp. 139–152). Commack, NY: Nova Science.

Comfort, L. (1999b). *Shared risk: Complex systems in seismic response.* Kindlington, UK: Pergamon/Elsevier.

Conrad, R. (1951). Speed and load stress in a sensorimotor skill. *British Journal of Industrial Medicine, 8,* 1–7.

Constant, D., Sproull, L., & Kiesler, S. (1997). The kindness of strangers: On the usefulness of electronic weak ties for technical advice. In S. Kiesler (Ed.), *Culture of the Internet* (pp. 303–322). Mahwah, NJ: Lawrence Erlbaum Associates.

Contractor, N. S., & Siebold, D. R. (1993). Theoretical framework for the study of structuring processes in group decision support systems. *Human Communication Research, 23,* 528–563.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Cooke, N. J., Amazeen, P. G., Gorman, J. C., Guastello, S. J., Likens, A., & Stevens, R. (2012). Modeling the complex dynamics of teamwork from team cognition to neurophysiology. *Proceedings of the Human Factors and Ergonomics Society, 56*, 183–187.

Cooke, N. J., & Shope, S. M. (2004). Designing a synthetic task environment. In S. G. Schiflett, L. R. Elliott, E. Salas, & M. D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 263–296). Burlington, VT: Ashgate.

Corker, K. M., Cramer, M. L., Henry, E. H., & Smoot, D. E. (1992). *Methodology for evaluation of automation impacts on tactical command and control (C-2) systems: Final report* (U.S. AFHRL Tech. Rep. No. 90-91). US: AL/HRPP.

Cornum, R., Matthews, M. D., & Seligman, M. E. P. (2011). Comprehensive soldier fitness: Building resilience in a challenging institutional context. *American Psychologist, 66,* 4–9.

Costa, G. (2003). Factors influencing health of workers and tolerance to shift work. *Theoretical Issues in Ergonomic Science, 4,* 263–288.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effect of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94,* 428–434.

Crawley, S. L. (1926). An experimental investigation of recovery from work. *Archives of Psychology, 85.*

Culbert, C., Riley, G., & Savely, R. T. (1989). An expert system development methodology that supports verification and validation. *ISA Transactions, 28,* 15–18.

Culley, K. E., & Madhavan, P. (2012). Affect and time pressure in a simulated luggage screening paradigm: A fuzzy signal detection theory analysis. *Proceedings of the Human Factors and Ergonomics Society, 56*, 338–342.

Czermak, P. (2001). Safety measures in laboratories and production plants using genetic technology. In J. Althoff (Ed.) *Safety of modern technical systems* (pp. 617–618). Saarbrücken, Germany: TÜV-Saarland Foundation.

Czerwinski, M. P., & Larson, K. (2003). Cognition and the web: Moving from theory to web design. In J. Ratner (Ed.), *Human factors and web development* (2nd ed., pp. 147–165). Mahwah, NJ: Lawrence Erlbaum Associates.

Daily, R. C. (1980). A path analysis of R & D team coordination and performance. *Decision Sciences, 11,* 357–369.

d'Amato, M. R. (1973). *Experimental psychology: Methodology, psychophysics and learning.* New York: McGraw-Hill.

Dang, N. T., Tavanti, M., Rankin, I., & Cooper, M. (2009). A comparison of different input devices for a 3D environment. *International Journal of Industrial Ergonomics, 39*, 554–563.

Darnell, M. J. (1997). *Yellow street lights.* Retrieved from www.baddesigns.com/strlt.htm

Darnell, M. J. (1999). *Don't go to the right?* Retrieved from www.baddesigns.com/dontgo.html

Dawson, D. A. (1994). Heavy drinking and the risk of occupational injury. *Accident Analysis and Prevention, 26,* 655–665.

Dawson, D., Noy, Y. I., Härmä, M., Akerstedt, T., & Belensky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis and Prevention, 43*, 549–564.

Day, K., & Calkins, M. P. (2002). Design and dementia. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 374–393). New York: Wiley.

Deatherage, B. (1972). Auditory and other sensory forms of information presentation. In H. Van Cott & R. Kinkade (Eds.), *Human engineering guide to equipment design* (pp. 123–160). Washington, DC: U.S. Government Printing Office.

DeGreene, K. B. (1991). Emergent complexity and person–machine systems. *International Journal of Man–Machine Studies, 35,* 219–234.

Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human–Computer Studies, 52,* 1–22.

Deininger, R. L. (1960). Human factors engineering studies of the design and use of pushbutton telephone sets. *Bell System Technical Journal, 39,* 986–1012.

Denning, P. J. (1986). The science of computing: Expert systems. *American Scientist, 74,* 18–20.

Dennis, A. R., & Valacich, J. S. (1993). Computer brainstorms: More heads are better than one. *Journal of Applied Psychology, 78,* 531–537.

Denny, V. E., Gilbert, K. J., Erdmann, R. C., & Rumble, E. T. (1978). Risk assessment methodologies: An application to underground mine systems. *Journal of Safety Research, 10,* 24–34.

de Pontbriand, R. J., Allender, L. E., & Doyle, F. J. III. (2008). Temporal regulation and temporal cognition: Biological and psychological aspects of time stress. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (pp. 143–162). Abingdon, Oxon, UK: Ashgate.

Depue, R. A., & Monroe, S. M. (1986). Conceptualization and measurement of human disorder in life stress research: The problem of chronic disturbance. *Psychological Bulletin, 99,* 36–51.

De Raedt, R., & Ponjaert-Kristoffersen, I. (2001). Predicting at-fault car accidents of older drivers. *Accident Analysis and Prevention, 33,* 809–820.

Derr, J., Forst, L., Chen, H. Y., & Conroy, L. (2001). Fatal falls in the U.S. construction industry, 1990–1999. *Journal of Occupational and Environmental Medicine, 43*, 921–931.

de Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & De Carolis, B. (2003). From Greta's mind to her face: Modeling the dynamics of affective states in a conversational embodied agent. *International Journal of Human–Computer Studies, 59,* 81–118.

Derthick, M. (2007). Where federalism didn't fail. *Public Administration Review, 67*, 36–47.

Desari, D., Crowe, C., Ling, C., Zhu, M., & Ding, L. (2010). EEG pattern analysis for physiological indicators of mental fatigue in simulated air traffic control tasks. *Proceedings of the Human Factors and Ergonomics Society, 54*, 205–209.

Deutsch, C. P. (1964). Auditory discrimination and learning: Social factors. *Merrill–Palmer Quarterly of Behavior and Development, 10,* 277–296.

Deutsch, D. (1992). Hie tritone paradox: Implications for the representation and communication of pit structures. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 115–138). Washington, DC: American Psychological Association.

Dewaard, J. E., & Dewaard, N. (1984). *History of NASA: America's voyage to the stars.* New York: Bison/Exter/Simon & Schuster.

Dey, A., & Mann, D. D. (2010). Sensitivity and diagnosticity of NASA-TLX and simplified SWAT to assess the mental workload associated with an agricultural sprayer. *Ergonomics, 53*, 848–857.

Dillon, R. L., & Tinsley, C. H. (2007). How near-misses influence decision making under risk: A missed opportunity for learning. *Management Science, 54*, 1425–1440.

Dingus, T. A., Hanowski, R. J., & Klauer, S. G. (2011). Estimating crash risk. *Ergonomics in Design, 19*(4), 8–12.

DiNocera, F., Ferlazzo, F., & Gentilomo, A. (1996). [Inter-individual stability in the P300 effect of event-related brain potentials under double-task conditions.] *Psychologie Francaise, 41*, 365–374.

DiPietro, A. (1976). *An analysis of the OSHA inspection program in manufacturing 1972–73.* Washington, DC: U.S. Department of Labor.

Doll, T. J., Folds, D. J., & Leiker, B. S. (1984). *Auditory information systems in military aircraft: Current configurations versus the state of the art* (Rep. No. USAFSAM-TR-84-15). Brooks Air Force Base, TX: U.S. Air Force School of Aerospace Medicine.

Doll, T. J., Hanna, T. E., & Russotti, J. S. (1992). Masking in three-dimensional auditory displays. *Human Factors, 34,* 255–265.

Dooley, K. J. (1997). A complex adaptive systems model of organization change. *Nonlinear Dynamics, Psychology, and Life Sciences, 1*, 69–97.

Dooley, K. J. (2004). Complexity science models of organizational change and innovation. In M. S. Poole & A. H. Van de Ven (Eds.), *Handbook of organizational change and innovation* (pp. 354–373). New York: Oxford University Press.

Dooley, K. J. (2009). The butterfly effect of the "butterfly effect." *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*, 279–288.

Doos, M., Backstrom, T., & Samuelson, S. (1994). Evaluation of strategy: Preventing accidents with automated machinery through targeted and comprehensive investigations conducted by safety engineers. *Safety Science, 17*, 187–206.

Dore, M. H. I. (2009). The impact of Edward Lorenz: An introductory overview. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13,* 241–242.

Dore, M. H. I., & Singh, R. G. (2009). Turning points in nonlinear business cycle theories, financial crisis, and the global 2007–2009 downturn. *Nonlinear Dynamics, Psychology, and Life Sciences, 13*, 423–444.

Douglas, M., & Wildavsky, A. (1982). *Risk and culture: An essay on the selection of environmental dangers.* Berkeley, CA: University of California Press.

Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: Cognitive aging. *Journal of Geriatric Psychiatry, 23*, 75–93.

Druskat, V. U., & Pescosolido, A. T. (2002). The content of effective teamwork mental models in self-managing teams: Ownership, learning and heedful interrelating. *Human Relations, 55,* 283–314.

Duan, Q. (2013). Service-oriented network virtualization for convergence of networking and cloud computing in next-generation networks. In A.-S. K. Pathen, M. M. Monowar, & Z. M. Fadlullah, (Eds.), *Building next-generation converged networks: Theory and practice* (pp. 393–416). Boca Raton, FL: CRC/Taylor & Francis.

Dubitzky, W. (Ed.). (2008). *Data mining techniques in grid computing environments*. Chichester, UK: Wiley-Blackwell.

Duffy, F. (1997). *The new office.* London: Conran Octopus.

Duiattorneytab. (2013). Blood alcohol levels by state. Retrieved from http://www.duiattorneytab.com/blood-alcohol-limits-by-state.html

Dunbar, E. 1993. The role of psychological stress and prior experience in the use of personal protective equipment. *Journal of Safety Research*, *24*, 181–187.

Dureman, E. I., & Boden, C. (1972). Fatigue in simulated car-driving. *Ergonomics, 15,* 299–305.

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors*, *50,* 442–448.

Dux, P. E., & Marios, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*, 1683–1700.

Eastman, M. C., & Kamon, E. (1976). Posture and subjective evaluation at flat and slanted desks. *Human Factors*, *18*, 15–26.

Eberts, R., & Salvendy, G. (1986). The contributions of cognitive engineering to the safe design and operation of CAM and robotics. *Journal of Occupational Accidents*, *8*, 49–67.

Eby, D. W., Vivoda, J. M., & Fordyce, T. A. (2002). The effects of standard enforcement on Michigan safety belt use. *Accident Analysis and Prevention*, *34,* 815–824.

Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics*, *25*, 202–210.

Ein-Dor, T., Mikulincer, M., Doron, G., & Shaver, P. R. (2010). The attachment paradox: How can so many of us (the insecure ones) have no adaptive advantages? *Perspectives on Psychological Science*, *5*, 123–141.

Elg, F. (2005). Leveraging intelligence for high performance in complex dynamic systems requires balanced goals. *Theoretical Issues in Ergonomic Science, 6,* 63–72.
Elliott, E., & Kiel, L. D. (Eds.). (2004). Agent-based modeling [Special issue]. *Nonlinear Dynamics, Psychology, and Life Sciences, 8,* 121–302.
Ellis, K. K. E., Kramer, L. J., Shelton, K. J., Arthur, J. J. III, & Prinzel, L. J. III. (2011). Transition of attention in terminal area NextGen operations using synthetic vision systems. *Proceedings of the Human Factors and Ergonomics Society, 55*, 46–60.
Enander, A. (1989). Effects of thermal stress on human performance. *Scandinavian Journal of Work and Environmental Health, 15*(1), 27–33.
Endres, L. S., Guastello, S. J., & Rieke, M. L. (1992). Meta-interpretive reliability of computer-based test interpretations: The Karson Clinical Report. *Journal of Personality Assessment, 59,* 448–467.
Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.
Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design*. Boca Raton, FL: CRC Press.
Epstein, I. (2010). *Clinical data-mining: Integrating practice and research*. New York: Oxford University Press.
Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies.* Cambridge, MA: MIT Press.
Epstein, Y., & Karlin, R. A. (1975). Effects of acute experimental crowding. *Journal of Applied Social Psychology, 5,* 34–53.
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.
Eubanks, J. (1993). *Pedestrian accident prevention.* Tucson, AZ: Lawyers and Judges.
Eubanks, J. L., & Killeen, P. R. (1983). An application of signal detection theory to air combat training. *Human Factors, 25,* 449–456.
Evans, L. (2002). Traffic crashes. *American Scientist, 90*, 244–253.
Evans, M., Hastings, N., & Peacock, B. (1993). *Statistical distributions* (2nd ed.). New York: Wiley.
Evans, R. E., & Kortum, P. (2010). The impact of voice characteristics on user response in an interactive voice response system. *Interacting with Computers, 22*, 606–614.
Eysenck, M. W. (1997). *Anxiety and cognition: A unified theory*. East Essex, UK: Psychology Press/Taylor & Francis.
Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: Are they viable? *Human Factors, 43,* 173–193.
Fans, S. M. (1997). *Evolution of 3-D technology.* Retrieved from http://vrex.com
Farazmand, A. (2007). Learning from the Katrina crisis: A global and international perspective with implications for future crisis management. *Public Administration Review, 67*, 149–159.
Fechner, G. T. (1964). *Elemente der psychophysik* [Elements of psychophysics]. Amsterdam, the Netherlands: E. J. Bonset.
Feinauer, D. M., & Havlovic, S. J. (1993). Drug testing as a strategy to reduce occupational accidents: A longitudinal analysis. *Journal of Safety Research, 24,* 1–7.
Fiedler, F. E., Bell, C. P. L., Chemers, M. M., & Patrick, D. (1984). Increasing mine productivity and safety through management training and organizational development: A comparative study. *Basic and Applied Social Psychology, 5,* 1–18.
Fine, P. E. M. (1979). John Brownlee and the measurement of infectiousness: An historical study in epidemic theory. *Journal of the Royal Statistical Society A, 142*, 347–362.
Fine, S. A., & Getkate, M. (1995). *Benchmark tasks for job analysis: A guide for functional job analysis (FJA) scales.* Mahwah, NJ: Lawrence Erlbaum Associates.
Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors, 52*, 203–224.
Fischer, U., McDonnell, L. K., Davison, J., Haars, K., Villeda, E., & Van Aken, C. (2001). *Detecting errors inflight.* Retrieved from http://olias.nasa.gov/reports/detecting.html
Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47,* 381–391.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46,* 199–210.

Flach, J. M., Guisinger, M. A., & Robinson, A. G. (1996). Fitts' law: Nonlinear dynamics and positive entropy. *Ecological Psychology, 8,* 281–325.

Flache, A., & Macy, M. W. (1997). The weakness of strong ties: Collective action failure in a highly cohesive group. In P. Doreian, & F. B. Stokman (Eds.), *Evolution of social networks* (pp. 19–44). Amsterdam, the Netherlands: Gordon & Breach.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327–358.

Flanders, V. (2013). Web pages that suck. Retrieved from www.webpagesthatsuck.com

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience, 5,* 1688–1703.

Fleishman, E. A. (1954). Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology, 48,* 437–454.

Fleishman, E. A. (1964). *The structure and measurement of physical fitness.* Englewood Cliffs, NJ: Prentice Hall.

Flin, R., Slaven, G., & Stewart, K. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors, 38,* 262–277.

Flowers, J. H., Buhman, D. C., & Turnage, K. D. (1997). Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors, 39,* 341–351.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge, MA: MIT Press.

Foss, R. D., Stewart, J. R., & Reinfurt, D. W. (2001). Evaluation of the effects of North Carolina's 0.08% BAC law. *Accident Analysis and Prevention, 33,* 507–518.

Fowler, B. (1994). P300 as a measure of workload during a simulated aircraft landing task. *Human Factors, 36*, 670–683.

Fox, J. (1984). Formal and knowledge-based methods in decision technology. *Acta Psychologica, 56,* 303–331.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239.

Freeny, M. (2007). Whatever happened to clinical privacy? *Annals of the American Psycotherapy Association, 10*(2), 13–17.

Frensch, P. A., & Runger, D. (2003). Implicit learning. *Current Directions in Psychological Science, 12,* 13–18.

Frey, P. W., & Sears, R. J. (1978). Model of conditioning incorporating the Rescorla–Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review, 85,* 321–340.

Frick, K., Jensen, P., Quinlan, M., & Wilthagen, T. (Eds.). (2000). *Systematic occupational health and safety management: Perspectives on an international development.* Oxford, UK: Pergamon.

Friedman, J. W., & Samuelson, L. (1994). The "folk theorem" for repeated games and continuous decision rules. In J. W. Friedman (Ed.), *Problems of coordination in economic activity* (pp. 103–128). Boston: Kluwer.

Fröhlich, B., Plate, J., Wind, J., Wesche, G., & Gobel, M. (2000). Cubic mouse-based interaction in virtual environments. *IEEE Computer Graphics and Applications, X,* 12–15.

Fu, W.-T., & Pirolli, P. (2013). Establishing the micro-to-macro link in cognitive engineering: Multilevel models of socio-computer interaction. In J. D. Lee & A. Kirklik (Eds.), *Oxford handbook of cognitive engineering* (pp. 501–513). New York: Oxford University Press.

Fuchs, C. (2012). *Implications of deep packet inspection (DPI) internet surveillance for society*. Technical Report No. 1. Uppsala, Sweden: Department of Informatics and Media, Uppsala University.

Fujii, M. D., & Patten, B. M. (1992). Neurology of microgravity and space travel. *Neurologic Clinics, 10,* 999–1013.

Fullarton, C., & Stokes, M. (2007). The utility of a workplace injury instrument in prediction of workplace injury. *Accident Analysis and Prevention, 39,* 28–37.

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors, 54*, 36–51.

Furnham, A., & Schofield, S. (1987). Accepting personality test feedback: A review of the Barnum effect. *Current Psychological Research and Reviews, 6,* 62–178.

Galam, S. (1996). When humans interact like atoms. In E. White, & J. H. Davis (Eds.), *Understanding group behaviour* (Vol. 1, pp. 293–312). Mahwah, NJ: Lawrence Erlbaum Associates.

Galitz, W. O. (1993). *User–interface screen design.* New York: Wiley.

Gallegati, M., Greenwald, B., Richiadi, M. G., & Stiglitz, J. E. (2008). The asymmetric effect of diffusion processes: Risk sharing and contagion. *Global Economy Journal, 8*(3). doi:10.2202/1524-5861.1365.

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis and Prevention, 43*, 573–590.

Gandit, M., Kouabeanan, D. R., & Caroly, S. 2009. Road-tunnel fires: Risk perception and management strategies among users. *Safety Science, 47*, 105–114.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human–Computer Interaction, 2,* 167–277.

Gaver, W. W. (1994). Using and creating auditory icons. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 417–446). Reading, MA: Addison-Wesley.

Gell-Mann, M. (1994). *The quark and the jaguar*. New York: W. H. Freeman.

Gethmann, C. F. (2001). Ethical aspects of technical safety. In J. Althoff (Ed.), *Safety of modern technical systems* (pp. 45–46). Saarbrücken, Germany: TÜV-Saarland Foundation.

Getty, D. J. (1983). Three dimensional displays: Perceptual research and applications. *Psychological Documents, 13,* 178 (Manuscript No. 2570).

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gilmore, R. (1981). *Catastrophe theory for scientists and engineers.* New York: Wiley.

Gilster, P. (1993). *The Internet navigator.* New York: Wiley.

Gittell, J. H. (2000). Organizing work to support relationship coordination. *International Journal of Human Resource Management, 11*, 517–539.

Glass, D. C., & Singer, J. E. (1972). *Urban stress: Experiments on noise and social stressors.* New York: Academic Press.

Glover, H. (1982). Four syndromes of post-traumatic stress disorder: Stressors and conflicts of the traumatized with special focus on the Vietnam Combat Veteran. *Journal of Traumatic Stress, 1,* 57–78.

Gnoni, M. G., Andriulo, S., Maggio, G., & Nardone, P. (2013). "Lean occupational" safety: An application for a near-miss management system design. *Safety Science, 53*, 96–104.

Gödel, K. (1962). *On formally undecidable propositions.* New York: Basic Books.

Goldberg, L. R. (1968). Simple models or simple process? Some research on clinical judgments. *American Psychologist, 23,* 483–496.

Goldberg, L. R. (1969). Man versus model of man: A rationale, plus some evidence, for a method of improving clinical inferences. *Psychological Bulletin, 6,* 422–432.

Goldstein, J. (2011). Emergence in complex systems. In P. Allen, S. Maguire, & B. McKelvey (Eds.), *The Sage handbook of complexity and management* (pp. 79–92). Thousand Oaks, CA: Sage.

Gontier, E., LeDantec, C., Leieu, A., Paul, I., Charvin, H., Bernard, C., Lalonde, R., Rebai, M. (2007). Frontal and parietal ERPs associated with duration discrimination with or without task interference. *Brain Research, 1170*, 79–89.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamics decision making. *Computers in Human Behavior, 21*, 273–286.

Goodrich, M. A. (2013). Multitasking and multi-robot management. In J. D. Lee & A. Kirklik (Eds.), *Oxford handbook of cognitive engineering* (pp. 379–394). New York: Oxford University Press.

Gordon, J. E. (1978). *Structures, or why things don't fall down.* New York: Penguin Books.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences, 14,* 265–290.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G., & Fouse, S. (2012). Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors, 54*, 503–517.

Governors High Safety Association. (2013). Seat belt laws. Retrieved from http://www.ghsa.org/html/stateinfo/laws/seatbelt_laws.html

Grant, A. M., & Parker, S. K. (2009). Redesigning work design theories: The rise of relational and proactive perspectives. *Academy of Management Annals, 3*, 317–375.

Greenleaf, W. (1996). VR solutions for rehabilitation, ergonomics, and disability. *Virtual Reality*, 3(1), 13–17.

Greenwood, D. (1991). An overview of neural networks. *Behavioral Science, 36*, 1–33.

Gregerson, N. P., Brehmer, B., & Moren, B. (1996). Road safety improvement in large companies: An experimental comparison of different measures. *Accident Analysis and Prevention, 28*, 297–306.

Gregg, S., Martin, L., Homola, J., Lee, P. U., Mercer, J., Brasil, C., Cabrall, C., & Lee, H. (2012). Shifts in air traffic controllers' situation awareness during high-altitude mixed equippage operations. *Proceedings of the Human Factors and Ergonomics Society, 56*, 95–99.

Gregson, R. A. M. (1988). *Nonlinear psychophysical dynamics.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gregson, R. A. M. (1992). *n-Dimensional nonlinear psychophysics.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gregson, R. A. M. (1998). Herbart and "Psychological Research on the Strength of a Given Impression." *Nonlinear Dynamics, Psychology, and Life Sciences, 2*, 157–167.

Gregson, R. A. M. (2009). Psychophysics. In S. J. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and complexity in psychology: The theory of nonlinear dynamical systems* (pp. 108–131). New York: Cambridge University Press.

Gregson, R. A. M. (2013). Symmetry-breaking, grouped images and multistability with transient unconsciousness. *Nonlinear Dynamics, Psychology, and Life Sciences, 17*, 325–344.

Grieg, G. L. (1990). On the shape of energy-detection ROC curves. *Perception and Psychophysics, 48*, 77–81.

Gruber, T. R., & Cohen, P. R. (1987). Design for acquisition: Principles of knowledge-system design to facilitate knowledge acquisition. *International Journal of Man–Machine Studies, 26*, 143–150.

Guastello, D. D., & Guastello, S. J. (1987a). A climate for safety in hazardous environments: A psychosocial approach (Report No. 2389). *Social and Behavioral Sciences Documents, 17*, 67.

Guastello, D. D., & Guastello, S. J. (1987b). The relationship between work group size and occupational accidents. *Journal of Occupational Accidents, 9*, 1–9.

Guastello, S. J. (1981). Catastrophe modeling of equity in organizations. *Behavioral Science, 26*, 63–74.

Guastello, S. J. (1982). Color matching and shift work: An industrial application of the cusp-difference equation. *Behavioral Science, 27*, 131–139.

Guastello, S. J. (1985). Euler buckling in a wheelbarrow obstacle course: A catastrophe with complex lag. *Behavioral Science, 30*, 204–212.

Guastello, S. J. (1987). A butterfly catastrophe model of motivation in organizations: Academic performance. *Journal of Applied Psychology, 72*, 165–182.

Guastello, S. J. (1988). Catastrophe modeling of the accident process: Organizational subunit size. *Psychological Bulletin, 103*, 246–255.

Guastello, S. J. (1989). Catastrophe modeling of the accident process: Evaluation of an accident reduction program using the Occupational Hazards Survey. *Accident Analysis and Prevention, 21*, 61–77.

Guastello, S. J. (1991). Psychosocial variables related to transit accidents: A catastrophe model. *Work and Stress, 5*, 17–28.

Guastello, S. J. (1992). Accidents and stress-related health disorders: Forecasting with catastrophe theory. In J. C. Quick, J. J. Hurrell, & L. M. Murphy (Eds.), *Work and well-being: Assessments and interventions for occupational mental health* (pp. 262–269). Washington, DC: American Psychological Association.

Guastello, S. J. (1993). Do we really know how well our occupational accident prevention programs work? *Safety Science, 16*, 445–463.

Guastello, S. J. (1995). *Chaos, catastrophe, and human affairs: Applications of nonlinear dynamics to work, organizations, and social evolution.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Guastello, S. J. (1998). Creative problem solving groups at the edge of chaos. *Journal of Creative Behavior, 32,* 38–57.

Guastello, S. J. (2000). Symbolic dynamic patterns of written exchange: Hierarchical structures in an electronic problem solving group. *Nonlinear Dynamics, Psychology, and Life Sciences, 4,* 169–188.

Guastello, S. J. (2001). Nonlinear dynamics and occupational accidents. In J. Althoff (Ed.), *Safety of modem technical systems* (pp. 315–324). Saarbriicken, Germany: TÜV-Saarland Foundation.

Guastello, S. J. (2002). *Managing emergent phenomena: Nonlinear dynamics in work organizations.* Mahwah, NJ: Lawrence Erlbaum Associates.

Guastello, S. J. (2003). Nonlinear dynamics, complex systems, and occupational accidents. *Human Factors and Ergonomics in Manufacturing, 13,* 293–304.

Guastello, S. J. (2005). Statistical distributions in self-organizing phenomena: What conclusions should be drawn? *Nonlinear Dynamics, Psychology, and Life Sciences, 9,* 463–478.

Guastello, S. J. (2006). Comment: Motor control research requires nonlinear dynamics. *American Psychologist, 61,* 77–78.

Guastello, S. J. (2007). Nonlinear dynamics and leadership emergence. *Leadership Quarterly, 18,* 357–369.

Guastello, S. J. (2009). Chaos as a psychological construct: Historical roots, principal findings, and current growth directions. *Nonlinear Dynamics, Psychology, and Life Sciences, 13,* 289–310.

Guastello, S. J. (2010). Nonlinear dynamics of team performance and adaptability in emergency response. *Human Factors, 52,* 162–172.

Guastello, S. J. (2011). Discontinuities: SETAR and catastrophe models with polynomial regression. In S. J. Guastello, & R. A. M. Gregson (Eds.). *Nonlinear dynamical systems analysis for the behavioral sciences using real data* (pp. 251–280). Boca Raton, FL: CRC Press/Taylor & Francis.

Guastello, S. J. (2013). Modeling illness and recovery with nonlinear dynamics. In J. P. Sturmberg & Carmel M. Martin (Eds.), *Handbook on complexity in health* (pp. 147–156). New York: Springer.

Guastello, S. J., Bock, B. R., Caldwell, P., & Bond, R. W., Jr. (2005). Origins of group coordination: Nonlinear dynamics and the role of verbalization. *Nonlinear Dynamics, Psychology, and Life Sciences, 9,* 297–333.

Guastello, S. J., Boeh, H., Gorin, H., Huschen, S., Peters, N. E., Fabisch, M., & Poston, K. (2013). Cusp catastrophe models for cognitive workload and fatigue: A comparison of seven task types. *Nonlinear Dynamics, Psychology, and Life Science, 17,* 23–47.

Guastello, S. J., Boeh, H., Schimmels, M., Gorin, H., Huschen, S., Davis, E., Peters, N. E., Fabisch, M., & Poston, K. (2012). Cusp catastrophe models for cognitive workload and fatigue in a verbally-cued pictorial memory task. *Human Factors, 54,* 811–825.

Guastello, S. J., Boeh, H., Shumaker, C., & Shimmels, M. (2012). Catastrophe models for cognitive workload and fatigue. *Theoretical Issues in Ergonomics Science, 13,* 586–602.

Guastello, S. J., & Bond, R. W., Jr. (2004). Coordination learning in Stag Hunt games with application to emergency management. *Nonlinear Dynamics, Psychology, and Life Sciences, 8,* 345–374.

Guastello, S. J., Gershon, R. M., & Murphy, L. R. (1999). Catastrophe model for the exposure to blood-borne pathogens and other accidents in health care settings. *Accident Analysis and Prevention, 31,* 739–750.

Guastello, S. J., Gorin, H., Huschen, S., Peters, N. E., Fabisch, M., & Poston, K. (2012). New paradigm for task switching strategies while performing multiple tasks: Entropy and symbolic dynamics analysis of voluntary patterns. *Nonlinear Dynamics, Psychology, and Life Sciences, 16,* 471–497.

Guastello, S. J., Gorin, H., Huschen, S., Peters, N. E., Fabisch, M., & Poston, K., & Weinberger, K. (2013). The minimum entropy principle and task performance. *Nonlinear Dynamics, Psychology, and Life Sciences, 17,* 405–423.

Guastello, S. J., & Guastello, D. D. (1986). The relation between the locus of control construct and involvement in traffic accidents. *Journal of Psychology: Interdisciplinary and Applied, 120,* 293–298.

Guastello, S. J., & Guastello, D. D. (1988). *The occupational hazards survey: Manual and case report.* Milwaukee, WI: Authors.

Guastello, S. J., & Guastello, D. D. (1998). Origins of coordination and team effectiveness: A perspective from game theory and nonlinear dynamics. *Journal of Applied Psychology, 83,* 423–437.

Guastello, S. J., & Gregson, R. A. M. (Eds.). (2011). *Nonlinear dynamical systems analysis for the behavioral sciences using real data*. Boca Raton, FL: CRC Press/Taylor & Francis.

Guastello, S. J., Koehler, G., Koch, B., Koyen, J., Lilly, A., Stake, C., & Wozniczka, J. (2008). Risk perception when the tsunami arrived. *Theoretical Issues in Ergonomic Science*, *9*, 95–114.

Guastello, S. J., & Johnson, E. A. (1999). The effect of downsizing on hierarchical work flow dynamics in organizations. *Nonlinear Dynamics, Psychology, and Life Sciences, 3,* 347–378.

Guastello, S. J., Korienek, G., & Traut, M. (1989). Verbal versus pictorial representation of objects in a human–computer interface. *International Journal of Man–Machine Studies, 31,* 99–120.

Guastello, S. J., & Liebovitch, L. S. (2009). Introduction to nonlinear dynamics and complexity. In S. J. Guastello, M. Koopmans, & D. Pincus. (Eds.), *Chaos and complexity in psychology: Theory of nonlinear dynamical systems* (pp. 1–40). New York: Cambridge University Press.

Guastello, S. J., & Lynn, M. (in press). Catastrophe model of the accident process, safety climate, and anxiety. *Nonlinear Dynamics, Psychology, and Life Sciences*.

Guastello, S. J., Malon, M., Timm, P., Weinberger, K., Gorin, H., Fabisch, M., & Poston, K. (2012, July). *Cusp catastrophe models for cognitive workload and fatigue in a vigilance task.* Paper presented to the 22nd Annual International Conference of the Society for Chaos Theory in Psychology & Life Sciences, Baltimore.

Guastello, S. J., & McGee, D. W. (1987). Mathematical modeling of fatigue in physically demanding jobs. *Journal of Mathematical Psychology, 31,* 248–269.

Guastello, S. J., Nathan, D. E., & Johnson, M. J. (2009). Attractor and Lyapunov models for reach and grasp movements with application to robot-assisted therapy. *Nonlinear Dynamics, Psychology, and Life Sciences, 13,* 99–121.

Guastello, S. J., Nielson, K. A., & Ross, T. J. (2002). Temporal dynamics of brain activity in human memory processes. *Nonlinear Dynamics, Psychology, and Life Sciences, 6,* 323–334.

Guastello, S. J., & Philippe, P. (1997). Dynamics in the development of large information exchange groups and virtual communities. *Nonlinear Dynamics, Psychology, and Life Sciences, 1,* 123–149.

Guastello, S. J., Pincus, D., & Gunderson, P. R. (2006). Electrodermal arousal between participants in a conversation: Nonlinear dynamics for linkage effects. *Nonlinear Dynamics, Psychology, and Life Sciences, 10,* 365–399.

Guastello, S. J., & Rieke, M. L. (1990). The Barnum effect and validation for computer-based test interpretations: The Human Resource Development Report. *Psychological Assessment: Journal of Consulting and Clinical Psychology, 2,* 186–190.

Guastello, S. J., & Rieke, M. L. (1994). Computer-based test interpretations as expert systems: Validity and viewpoints from artificial intelligence theory. *Computers in Human Behavior, 4,* 435–495.

Guastello, S. J., Shircel, A., Poggi, D., Malon, M., Timm, P., Weinberger, K., Reiter, K., & Fabisch, M. (2013, July). *Cusp catastrophe models for cognitive workload and fatigue: A new paradigm for studying biases in financial decision making.* Paper presented to the 23rd Annual International Conference of the Society for Chaos Theory in Psychology & Life Sciences, Portland.

Guilford, J. P. (1967). *The structure of intellect.* New York: McGraw-Hill.

Guion, R. M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gun, R. T. (1993). The role of regulations in the prevention of occupational injury. *Safety Science, 16,* 47–66.

Gundel, A., Drescher, J., & Polyakov, V. V. (2001). Quantity and quality of sleep during the record manned space flight of 438 days. *Human Factors and Aerospace Safety, 1,* 87–98.

Gunnars Ergo, H. D. (2012). Gunnar digital performance eyewear, optics by Carl Zeiss Vision. Retrieved from http://czvgunnars.com/?a=-1

Gureckis, T. M., & Goldstone, R. L. (2006). Thinking in groups. *Pragmatics & Cognition, 14*, 293–311.

Hackman, J. R. (Ed.). (1990). *Groups that work (and those that don't).* San Francisco: Jossey-Bass.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 15*, 250–279.

Haigh, R. D. (1977). Mechanical safety and electrical control gear. In W. Handley (Ed.), *Industrial safety handbook* (2nd ed., pp. 251–259). New York: McGraw-Hill.

Haken, H. (1984). *The science of structure: Synergetics*. New York: Van Nostrand Reinhold.

Haken, H. (1988). *Information and self-organization: A macroscopic approach to complex systems.* New York: Springer.

Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology, 70*, 657–669.

Halgrave, R. J., & Colvin, C. (1998). Selection of personnel for stressful occupations: The potential utility of psychophysiological measures as selection tools. *Human Performance in Extreme Environments, 9*, 121–139.

Halpern, C. A., & Dawson, K. D. 1997. Design and implementation of a participatory ergonomics program for machine sewing tasks. *International Journal of Industrial Ergonomics, 20*, 429–440.

Hammer, J. M. (1999). Human factors of functionality and intelligent avionics. In D. Garland, J. Wise, & V. D. Hopkins (Eds.), *Handbook of human factors in aviation* (pp. 549–565). Mahwah, NJ: Lawrence Erlbaum Associates.

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution.* New York: HarperCollins.

Hammer, M., & Champy, J. (1996). Reengineering the corporation: The enabling role of information technology. In J. M. Shafritz & J. S. Ott (Eds.), *Classics of organizational theory* (pp. 561–577). Belmont, CA: Wadsworth.

Hancock, M. F. Jr. (2012). *Practical data mining.* Boca Raton, FL: CRC Press.

Hancock, P. A. (2007). On the process of automation transition in multitask human–machine systems. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 37*, 586–598.

Hancock, P. A., & Desmond, P. A. (Eds.). (2001). *Stress, workload, and fatigue*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human Factors, 49*, 851–877.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*, 519–537.

Haq, S., & Zimring, C. (2003). Just down the road a piece: The development of topological knowledge of building layouts. *Environment and Behavior, 35*, 132–160.

Harary, F., Norman, R. Z., & Cartright, D. (1965). *Structural models: An introduction to the theory of directed graphs*. New York: Wiley.

Hardy, L. (1996a). Testing the predictions of the cusp catastrophe model of anxiety and performance. *Sport Psychologist, 10,* 140–156.

Hardy, L. (1996b). A test of catastrophe models of anxiety and sports performance against multidimensional anxiety theory models using the method of dynamic differences. *Anxiety, Stress, and Coping, 9,* 69–86.

Hardy, L., & Parfitt, C. G. (1991). A catastrophe model of anxiety and performance. *British Journal of Psychology, 82,* 163–172.

Harms-Ringdahl, L. (1987). Safety analysis in design: Evaluation of a case study. *Accident Analysis and Prevention, 19,* 305–317.

Haroush, K., Deouell, L.Y., & Hochstein, S. (2011). Hearing while blinking: Multisensory attentional blink revisited. *The Journal of Neuroscience, 31*, 922–927.

Harris, P. R. (1989). Behavioral science space contributions. *Behavioral Science, 34,* 207–227.

Harris, P. R. (1992). *Living and working in space: Human behavior, culture, and organization.* New York: Ellis Horwood/Simon & Schuster.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA task load index (TLX): Results of experimental and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human workload* (pp. 138–183). Amsterdam, the Netherlands: North-Holland.

Haubner, P. (1992). Design of dialogues for human–computer interaction. In M. Galer, S. Harker, & J. Ziegler (Eds.), *Methods and tools in user-centered design for information technology* (pp. 203–236). Amsterdam, the Netherlands: North-Holland.

Hausdorff, J. M., Ashkenazy, Y., Peng, C.-K., Ivanov, P. C., Stanley, H. E., & Goldberger, A. I. (2001). When human walking becomes random walking: Fractal analysis and modeling of gait rhythm fluctuations. *Physica A, 302*, 138–147.

Havlovic, S. J., & Feinauer, D. (1990). Worker participation groups and safety committees: Is there an interaction effect on OSHA recordable accidents? *Proceedings of the Industrial Relations Research Association, 42,* 514.

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 802–811.

Hazy, J. K. (2008). Toward a theory of leadership in complex systems: Computational model explorations. *Nonlinear Dynamics, Psychology, and Life Sciences, 12*, 281–310.

Hazy, J. K., Goldstein, J. A., & Lichtenstein, B. (Eds.). 2007. *Complex systems leadership theory*. Litchfield Park, AZ: ISCE.

Head, J., Helton, W., Russell, P., & Neumann, E. (2012). Text-speak processing impairs tactile location. *Acta Pscyhologica, 141*, 48–53.

Heath, R. A. (2002). Can people predict chaotic sequences? *Nonlinear Dynamics, Psychology, and Life Sciences, 6,* 37–54.

Hedge, A., Erickson, W. A., & Rubin, G. (1992). The effects of personal and occupational factors on sick building syndrome reports in air-conditioned offices. In J. M. Quick & J. J. Hurrell, Jr. (Eds.), *Stress and well-being at work* (pp. 286–298). Washington, DC: American Psychological Association.

Hedge, A., & James, T. (2012). Gender effects on musculoskeletal symptoms among physician computer users in outpatient diagnostic clinics. *Proceedings of the Human Factors and Ergonomics Society, 56*, 887–891.

Heinrich, H. W. (1931). *Industrial accident prevention.* New York: McGraw-Hill.

Helander, M. G. (2007). Using design equations to identify sources of complexity in human–machine interaction. *Theoretical Issues in Ergonomics Science, 8*, 123–146.

Helander, M. G., & Burri, G. J. (1995). Cost effectiveness of ergonomics and quality improvements in electronics manufacturing. *International Journal of Industrial Ergonomics, 15,* 137–151.

Hellesoy, O., Gronhaug, K., & Kvitastein, O. 1998. Profiling the high hazards perceivers: An exploratory study. *Risk Analysis, 18*, 253–259.

Helton, W. S., Weil, L., Middlemiss, A., & Sawers, A. (2010). Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Consciousness and Cognition, 19*, 77–85.

Henao, O. (2002). The ability of competent and poor readers to remember information from hypermedia and printed texts [Capacidad de lectores competentes 7 lectores poco habiles para recordard informacion de un texto hipermedial e impresso]. *Infacia y Aprendizaje, 25,* 315–328.

Heppenheimer, T. A. (1990). How Von Neumann showed the way. *American Heritage of Innovation and Technology,* 6(2), 8–16.

Hertel, G., Kerr, N. L., & Messé, L. A. (2000). Motivation gains in performance groups: Paradigmatic and theoretical developments on the K'hler effect. *Journal of Personality and Social Psychology, 79*, 580–601.

Herzog, T. R., & Bosley, P. J. (1992). Tranquillity and preference as affective qualities of national environments. *Journal of Environmental Psychology, 12,* 115–128.

Hess, R. F. (1990). The Eldridge–Green lecture: Vision at low light levels: Role of spatial, temporal and contrast filters. *Ophthalmic and Physiological Optics, 10,* 351–359.

Heuer, H., Manzey, D., Lorenz, B., & Sangals, J. (2003). Impairments of manual tracking performance during spaceflight are associated with specific effects of microgravity on visuomotor transformations. *Ergonomics, 46,* 920–934.

Hlavacka, F., Dzurkova, O., & Pólonyová, A. (2000). CNS adaptation to microgravity and its influence on human balance after spaceflight. *Homeostasis, 40,* 132–133.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive–energetical framework. *Biological Psychology, 45*, 73–93.

Holcom, M. L., Lehman, W. E. K., & Simpson, D. D. (1993). Employee accidents: Influences of personal characteristics, job characteristics, and substance use in jobs different in accident potential. *Journal of Safety Research, 24,* 205–221.

Holding, D. H. (1983). Fatigue. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 145–167). New York: Wiley.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Cambridge, MA: Perseus.

Hollis, G., Kloos, H., & Van Orden, G. C. (2009). Origins of order in cognitive activity. In S. J. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and complexity in psychology: The theory of nonlinear dynamical systems* (pp. 206–242). New York: Cambridge University Press.

Hollnagel, E. (2011). The score of resilience engineering. In E. Hollnagel, J. Periès, D. D. Woods, & J. Wreathall (Eds.), *Resilience engineering in practice: A guidebook* (pp. xxix–xxxix). Burlington, VT: Ashgate.

Hollnagel, E. (2012). *FRAM: The functional resonance analysis method*. Burlington, VT: Ashgate.

Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2006). *Resilience engineering*. Burlington, VT: Ashgate.

Holmberg, D. (2011). Demand response and standards: New role for buildings in the smart grid. *BACnet Today & the Smart Grid, ASHRAE Journal, November, nv*, B23–B28.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research, 11*, 213–218.

Hong, S. L. (2010). The entropy conservation principle: Applications in ergonomics and human factors. *Nonlinear Dynamics, Psychology, and Life Sciences, 14,* 291–315.

Hong, W., Thong, J. M., & Tarn, K. Y. (2004). Does animation attract online users' attention? The effects of flash on information search performance and perceptions. *Information Systems Research, 15,* 60–86.

Hormann, H. (2001). Safety at sea. In J. Althoff (Ed.), *Safety of modern technical systems* (pp. 361–369). Saarbrücken, Germany: TÜV-Saarland Foundation.

Horrey, W. J. (2011). Assessing the effects of in-vehicle tasks on driving performance. *Ergonomics in Design, 19*(4), 4–7.

Houston, D. J., & Richardson, L. E., Jr. (2002). Traffic safety and the switch to a primary seat belt law: The California experience. *Accident Analysis and Prevention, 34,* 743–807.

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Boston: PWS-Kent.

Hoyert, M. S. (1992). Order and chaos in fixed-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 57*, 339–363.

Hudlick, E., & McNeese, M. D. (2002). Assessment of user affective and belief states for interface adaptation: Application to an Air Force pilot task. *User Modeling and User-Adapted Interaction, 12,* 1–47.

Hudson, P. T. W. (2001). Automation: A problematic solution. In J. Alfthoff (Ed.), *Safety of modern technical systems* (pp. 393–402). Saarbrücken, Germany: TÜV-Saarland Foundation.

Huffington Post (2012, October 8). Texas school district reportedly threatening students who refuse tracking ID, can't vote for homecoming. Retrieved from http://www.huffingtonpost.com/2012/10/08/texas-school-district-rep_n_1949415.html

Hurrell, J. J., Jr., & Murphy, L. R. (1996). Occupational stress intervention. *American Journal of Industrial Medicine, 29,* 338–341.

Husemann, B., Von Mach, C. Y., Borsotto, D., Zepf, K. I., & Scharnbacher, J. (2009). Comparison of musculoskeletal complaints and data entry between a sitting and sit-stand workstation paradigm. *Human Factors, 51*, 310–320.

Ijspeert, A., Nakanishi, J., & Schaal, S. (2003). *Learning attractor landscapes for motor primitives.* Cambridge, MA: MIT Press.

Illuminating Engineering Society. (1981). *IES lighting handbook.* New York: Author.

International Electrotechnical Commission. (2005). IEC *60601-1-8: Medical electrical equipment—General requirements, tests, and guidance for alarm systems in medical electrical equipment and medical electrical systems*. Geneva, Switzerland: Author.

International Standards Organization. (1986). *Mechanical vibration: Guidelines for the measurement and assessment of human exposure to hand-transmitted vibration* (Standard No. S3.34). Geneva, Switzerland: Author.

Ioteyko, J. (1920). *La fatigue* [Fatigue] (2nd ed.). Paris: Flammarion.

Ivanoff, R. V., Yakimov, P. I., & Djamiykov, T. S. (2011). Implementation of smart metering as an essential part of advanced metering infrastructure by end-customers. *Annual Journal of Electronics, nv,* 17–20.

Jacko, J. A., & Sears, A. (2003). *The human–computer interaction handbook*. Mahwah, NJ: Lawrence Erlbaum Associates.

Jackson, E. A. (1991a). Control of dynamic flows and attractors. *Physical Review A, 44,* 4839–4853.

Jackson, E. A. (1991b). On the control of complex dynamic systems. *Physica D, 50,* 341–366.

Jacobs, R. R., Conte, J. M., Day, D. V., Silva, J. M., & Harris, R. (1996). Selecting bus drivers: Multiple predictors, multiple perspectives on validity, and multiple estimates of utility. *Human Performance, 9, 199*–217.

Jacobsen, J. J., & Guastello, S. J. (2011). Diffusion models for innovation: S-curves, networks, power laws, catastrophes, and entropy. *Nonlinear Dynamics, Psychology, and Life Sciences, 15,* 307–333.

Jagacinski, R. J., & Flach, J. M. (2003). *Control theory for humans.* Mahwah, NJ: Lawrence Erlbaum Associates.

Janata, P., & Edwards, W. H. (2013). A novel sonification strategy for auditory display of heart rate and oxygen saturation changes in clinical settings. *Human Factors, 55*, 356–372.

Jarrett, R. F. (1948). Percent increases in output of selected personnel as an index of test efficiency. *Journal of Applied Psychology, 32,* 135–145.

Jensen, P. L. (1997). Can participatory ergonomics become 'the way we do things in this firm'?: The Scandinavian approach to participatory ergonomics. *Ergonomics, 40*, 1078–1087.

Ji, Y. G. (Ed.). (2013). *Advances in affective and pleasurable design*. Boca Raton, FL: CRC Press/Taylor & Francis.

Jirsa, V. K., & Kelso, J. A. S. (Eds.). (2004a). *Coordination dynamics: Issues and trends*. New York: Springer.

Jirsa, V. K., & Kelso, J. A. S. (2004b). Integration and segregation of perceptual and motor behavior. In V. K., Jirsa & J. A. S. Kelso (Eds.), *Coordination dynamics: Issues and trends* (pp. 243–260). New York: Springer.

Jobidon, M.-E., Rousseau, R., & Breton, R. (2005). The effect of variability in temporal information on the control of a dynamic task. *Theoretical Issues in Ergonomic Science, 6,* 49–62.

Joels, K. M., Kennedy, G. P., & Larkin, D. (1982). *The space shuttle operator's manual.* New York: Ballantine.

Johnson, C. (2002). *Novel visualization for the presentation of accident reports.* Retrieved from http://www.dcs.gla.uk/~johnson/papers

Johnson, S. E. (2007). The predictive validity of safety climate. *Journal of Safety Research, 38*, 511–521.

Jones, L. (2001). *Human factors and haptic interfaces.* Retrieved from www.ims.umn.edu/talks/workshops/6-14-15.2001/jones/jones.pdf

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors*, *50*, 90–111.

Jorgensen, F., & Pedersen, P. A. (2002). Drivers' response to the installation of road lighting: An economic interpretation. *Accident Analysis and Prevention, 34,* 601–608.

Juarrero, A. (1999). *Dynamics in action.* Cambridge, MA: MIT Press.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 253–291.

Kanas, N., Salnitskiy, V., Grand, E. M., Gushin, V., Weiss, D. S., Kozerenko, O. et al. (2002). Lessons learned from shuttle/Mir: Psychosocial countermeasures. *Aviation, Space, and Environmental Medicine*, *73,* 607–611.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 617–671.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66–71.

Kanfer, R. (2011). Determinant and consequences of subjective cognitive fatigue. In P. Ackerman (Ed.), *Cognitive fatigue* (pp. 189–208). Washington, DC: American Psychological Association.

Kantowitz, B. (2001). *Reducing driver distractions in the age of overload.* Retrieved from http://www.umich.edu/~newsinfo/Releases/2001/Mar01/r030501c.html

Kantowitz, B., & Sorkin, D. (1983). *Human factors: Understanding people–system relationships.* New York: Wiley.

Kantowitz, B. H. (1985). Channels and stages in human information processing: A limited analysis of theory and methodology. *Journal of Mathematical Psychology, 29,* 135–174.

Kaplan, D., & Glass, L. (1995). *Understanding nonlinear dynamics.* New York: Springer.

Karayanni, D. A., & Baltas, G. A. (2003). Web site characteristics and business performance: Some evidence from international business-to-business organizations. *Marketing Intelligence and Planning, 21,* 105–114.

Karwowski, W. (2000). Purely editorial. *Theoretical Issues in Ergonomic Science, 1,* 1–2.

Karwowski, W. (Ed.). (2006). *Handbook of standards and guidelines in ergonomics and human factors.* Mahwah, NJ: Lawrence Erlbaum Associates.

Karwowski, W. (2012). A review of human factors challenges of complex adaptive systems: Discovering and understanding chaos in human performance. *Human Factors, 54*, 983–995.

Karwowski, W., Rahmi, M., & Mihalyi, T. (1988). Effects of computerized automation and robotic safety performance of a manufacturing plant. *Journal of Occupational Accidents, 10*, 217–233.

Katsavelis, D., Muhkerjee, M., Decker, L., & Stergiou, N. (2010a). Variability of lower extremity joint kinematics during backward walking in a virtual environment. *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 165–178.

Katsavelis, D., Muhkerjee, M., Decker, L., & Stergiou, N. (2010b). The effect of virtual reality on gait variability. *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 239–256.

Kauffman, S. A. (1993). *The origins of orders: Self-organization and selection in evolution.* New York: Oxford University Press.

Kaufman, L. (1974). *Sight and mind: An introduction to visual perception.* New York: Oxford University Press.

Kelly, A. J., & Dux, P. E. (2011). Different attentional blink tasks reflect distinct information processing limitations: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1867–1873.

Kelly, K. (1994). *Out of control: The new biology of machines, social systems and the economic world.* Reading, MA: Addison-Wesley.

Kelly, R. V., Jr. (1995). Abundant worlds and virtual personas. *Virtual Reality,* 2(5), 35–43.

Kemeny, J. (1979). *The need for change: The legacy of TMI. Report of the Presidential Commission on Three Mile Island.* New York: Pergamon Press.

Kerr, R., McHugh, M., & McCrory, M. 2009. HSE management standards and stress-related outcomes. *Occupational Medicine, 59*, 584–479.

Khalid, H. M. (Ed.). (2004). Conceptualizing effective human factors design [Special issue]. *Theoretical Issues in Ergonomics Science, 5*(1).

Kieras, D. (1988). Toward a practical GOMS model methodology for user interface design. In M. Helander (Ed.), *Handbook of human–computer interaction* (pp. 135–157). Amsterdam, the Netherlands: Elsevier/North-Holland.

Kieras, D. (2004). GOMS models for task analysis. In D. Diaper & N. A. Stanton (Eds.), *The handbook of task analysis for human–computer interaction* (pp. 83–116). Mahwah, NJ: Lawrence Erlbaum Associates.

Kirlik, A. (2006). *Adaptive perspectives on human–technology interaction: Methods and models for cognitive engineering and human–computer interaction.* New York: Oxford University Press.

Kirwan, B. (2001). Transportation safety: A changing landscape. In J. Althoff (Ed.), *Safety of modern technical systems* (pp. 325–334). Saarbrücken, Germany: TÜV-Saarland Foundation.

Kiselev, A., Medvedev, A. A., & Menshikov, V. A. (2003). *Astronautics: Summary and prospects.* Vienna, Austria: Springer.

Kjellen, U. (1984a). The deviation concept in occupational accident control—1: Definition and classification. *Accident Analysis and Prevention, 16,* 289–306.

Kjellen, U. (1984b). The deviation concept in occupational accident control—2: Data collection and assessment of significance. *Accident Analysis and Prevention, 16,* 307–323.

Kjellen, U. (1990). Safety control in design: Experiences of an offshore project *Journal of Occupational Accidents, 12*, 49–61.

Kjellén, U., & Hovden, J. (1993). Reducing risks by deviation control—A retrospection into a research strategy. *Safety Science, 16*, 417–438.

Kjellen, U., & Larsson, T. J. (1981). Investigating accidents and reducing risks: A dynamic approach. *Journal of Occupational Accidents, 3*, 129–140.

Klahr, P., & Waterman, D. A. (1986). *Artificial intelligence: A Rand perspective* (Tech. Rep. No. P-7172). Santa Monica, CA: Rand.

Klein, G. (1989). Recognition-primed decisions. *Advances in Man–Machine Systems Research, 5,* 47–92.

Klein, J., Moon, Y., & Picard, R. W. (2002). The computer responds to user frustration: Theory, design, and results. *Interacting with Computers, 14,* 119–140.

Klen, T., & Vayrynen, S. T. (1983). The protective influence of logger's safety equipment. *Scandinavian Journal of Work and Environmental Health, 9,* 201–203.

Koehler, G. (1995). Fractals and path-dependent processes: A theoretical approach for characterizing emergency medical responses to major disasters. In R. Robertson & A. Combs (Eds.), *Chaos theory in psychology and the life sciences* (pp. 199–216). Hillsdale, NJ: Lawrence Erlbaum Associates.

Koehler, G. (1996). What disaster response management can learn from chaos theory. In G. Koehler (Ed.), *What disaster response management can learn from chaos theory* (pp. 2–41). Sacramento, CA: California Research Bureau, California State Library.

Koehler, G. (1999). The time compacted globe and the high-tech primitive at the millennium. In E. Elliott & L. D. Kiel (Eds.), *Nonlinear dynamics, complexity, and public policy* (pp. 153–174). Commack, NY: Nova Science.

Koehler, G. (2011). Q-consciousness: Where is the flow? *Nonlinear Dynamics, Psychology, and Life Sciences, 15*, 335–358.

Koelega, H. S., Brinkman, J.-A., Hendriks, L., & Marinus, N. (1989). Processing demands, effort, and individual differences in four different vigilance tasks. *Human Factors, 31*, 45–62.

Koesling, H., Zoellner, M., Sichelschmidt, L., & Ritter, H. (2009). With a flick of the eye: Assessing gaze-controlled human–computer interaction. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Human centered robot systems: Cognition, interaction, technology* (pp. 111–120). New York: Springer.

Kohonen, T. (1989). *Self-organization and associative memory.* New York: Springer.

Koller, F. (1992). Multimedia interfaces. In M. Galer, S. Harker, & J. Ziegler (Eds.), *Methods and tools in user-centered design for information technology* (pp. 299–315). Amsterdam, the Netherlands: North-Holland.

Koppen, M., & Doignan, J.-P. (1990). How to build a knowledge space by querying an expert? *Journal of Mathematical Psychology, 34*, 311–331.

Kovalerchuk, B., & Vityaev, E. (2000). *Data mining in finance: Advances in relational and hybrid methods*. Boston: Kluwer.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 1–78). New York: Oxford University Press.

Kremen, M. A. (1966). The effectiveness of tracking behavior as a function of the mode of the input signal. *Voprosy Psychologii, 5*, 83–92.

Kring, J. P. (2001). Multicultural factors for international spaceflight. *Journal of Human Performance in Extreme Environments, 5,* 11–32.

Kroll, W. (1981). The C. H. McCoy research lecture: Analysis of local muscular fatigue patterns. *Research Quarterly for Exercise and Sport, 52*, 523–539.

Kumar, S. (2002). Trunk rotation: Ergonomic and evolutionary perspectives. *Theoretical Issues in Ergonomic Science, 3,* 235–256.

Kundi, M. (2003). Ergonomic criteria for the evaluation of work schedules. *Theoretical Issues in Ergonomic Science, 4,* 302–318.

Kurz, M. J., Markopoulou, K., & Stergiou, N. (2010). Attractor divergence as a metric for assessing walking balance. *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 151–164.

Lacherez, P., Seah, E. L., & Sanderson, P. (2007). Overlapping melodic alarms are almost indiscriminable. *Human Factors, 49*, 637–645.

Lagi, M., Bar-Yam, Y., Bertrand, K. Z., & Bar-Yam, Y. (2011). The food crises: A quantitative model of food prices including speculators and ethanol conversion. Retrieved from http://arxiv.org/abs/1109.4859

Landauer, T. K. (1987). Relations between cognitive psychology and computer system design. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction* (pp. 1–25). Cambridge, MA: MIT Press.

Landauer, T. K., Galotti, K. M., & Hartwell, S. (1983). Natural command names and initial learning: A study of text-editing terms. *Communications of the CAM, 26,* 495–503.

Landsbergis, P. A., Schnall, P. L., Schwartz, J. E., Warren, K., & Pickering, T. G. (1995). Job strain, hypertension, and cardiovascular disease: Empirical evidence, methodological issues, and recommendations for further research. In S. K. Sauter & L. R. Murphy (Eds.), *Organizational risk factors for job stress* (pp. 97–112). Washington, DC: American Psychological Association.

Laner, S., & Sell, R. G. (1960). An experiment on the effect of specially designed safety posters. *Occupational Psychology, 34,* 153–169.

Lange, R., Oliva, T., & McDade, S. (2000). An algorithm for estimating multivariate catastrophe theory models: GEMCAT II. *Studies in Nonlinear Dynamics in Economics, 4,* 138–168.

Lanyon, R. I. (1984). Personality assessment. *Annual Review of Psychology, 35,* 667–701.

Lanyon, R. I. (1987). The validity of computer-based personality assessment products: Recommendations for the future. *Computers in Human Behavior, 3*, 225–238.

Lapeyre, B., Hourlier, S., Servantie, X., N'Kaoua, B., & Sauzéon, H. (2011). Using the landmark-route-survey framework to evaluate spatial knowledge obtained from synthetic vision systems. *Human Factors, 53*, 647–661.

Lardent, C. L., Jr. (1991). Pilots who crash: Personality constructs underlying accident prone behavior of fighter pilots. *Multivariate Experimental Clinical Research, 10,* 1–25.

Larsson, T. J. (1993). Investigating accidents and reducing risks: A dynamic approach (Kjellen & Larsson, 1981); its relevance for injury prevention. *Safety Science, 16,* 439–444.

Laughlin, P. R. (1996). Group decision making and collective induction. In E. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (pp. 61–80). Mahwah, NJ: Lawrence Erlbaum Associates.

Leary, M. R. (1990). Anxiety, cognition, and behavior: In search of a broader perspective. *Journal of Social Behavior and Personality, 5*, 39–44.

Leary, M. R., & Kowalski, R. M. (1995). *Social anxiety*. New York: Guilford Press.

LeBlanc, L. A., & Rucks, C. T. (1996). A multiple discriminant analysis of vessel accidents. *Accident Analysis and Prevention, 28*, 501–510.

Lee, P. U., & Prevor, T. (2012). Prediction of traffic complexity and controller workload in mixed-equippage NextGen environments. *Proceedings of the Human Factors and Ergonomics Society, 56*, 100–104.

Lehner, P. E. (1989). Toward an empirical approach to evaluating the knowledge base of an expert system. *IEEE Transactions on Systems, Man, and Cybernetics, 19,* 658–662.

Lehtola, M. M., van der Molen, H. F., Lappalainen, J., Hoonakker, P. L. T., Hsiao, H., Haslam, R. A., Hale, A. R., & Verbeek, J. H. (2008). The effectiveness of interventions for preventing injuries in the construction industry: A systematic review. *American Journal of Preventive Medicine, 35*, 77–85.

Leonhardt, J., Macchi, L., Hollnagel, E., & Kirwan, B. (2009). A white paper on resilience engineering for ATM. Eurocontrol. Retrieved from http://www.eurocontrol.int/esp/gallery/content/public/library.

Lerch, F. J., & Harter, D. E. (2001). Cognitive support for real-time dynamic decision making. *Information Systems Research, 12*, 63–82.

Lesgold, A. (2000). On the future of cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 451–465). Mahwah, NJ: Lawrence Erlbaum Associates.

Levi, K. (1989). Expert systems should be more accurate than human experts: Evaluation procedures from human judgment and decision making. *IEEE Transactions on Systems, Man, and Cybernetics, 19,* 658–662.

Levine, J. H., & Spedaro, J. (1988). Occupational mobility: A structural model. In B. Wellman, & S. D. Berkowitz (Eds.), *Social structures: A network approach* (pp. 452–475). New York: Cambridge University Press.

Levy, S. (1992). *Artificial life: A report from the frontier where computers meet biology.* New York: Random House.

Li, T., Ogihara, M., & Tzanetakis, G. (Eds.). (2012). *Music data mining*. Boca Raton, FL: CRC Press.

Liao, M.-J., & Johnson, W. W. (2004). Characterizing the effects of droplines on target acquisition performance on a 3-D perspective display. *Human Factors, 46*, 476–496.

Liberty Mutual Insurance. (2005). *Liberty Mutual materials handling tables*. Boston: Author. Retrieved from http://libertymmhtables.libertymutual.com/CM_LMTablesWeb

Lidwell, W., Holden, K., & Butler, J. (2011). *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Beverly, MA: Rockport.

Liebovitch, L. S. (2004, July). *Arrival time of e-mail and viruses provide information about the structure and dynamics of the Internet.* Paper presented at the 14th annual international conference of the Society for Chaos Theory in Psychology and Life Sciences, Milwaukee, WI.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Likert, R. (1961). *New patterns of management*. New York: McGraw-Hill.

Lin, Y., & Cai, H. (2009). A method for building a real-time cluster-based continuous mental workload scale. *Theoretical Issues in Ergonomic Science, 10*, 531–544.

Linde, L. K. (1986). What is domain skill? Frames in information seeking. *Behavioral Science, 31,* 89–102.

Lipscomb, H. J., Li, I., & Dement, J. (2003). Work-related falls among union carpenters in Washington State before and after the Vertical Fall Arrest Standard. *American Journal of Industrial Medicine, 44*, 157–161.

Lipscomb, J. S. (1996). *Stereoscopic and head-mounted display on IBM RISC System/6000 workstations, and on IBM PCs.* Retrieved from www.research.ibm.com/people/lipscomb/www-stereo-print-export.html

Little, K. B., & Schneidman, E. S. (1959). Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs,* 75 (6, Whole No. 476).

Liu, M. (2004). Examining the performance and attitudes of sixth graders during their use of a problem-based hypermedia learning environment. *Computers in Human Behavior, 20,* 357–379.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors, 49*, 376–399.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science, 20*, 240–245.

Lonero, L. P., Clinton, K., Wilde, G. J. S., Roach, K., McKnight, A. J., MacLean, H. et al. (1994). *The roles of legislation, education and reinforcement in changing road user behavior.* Toronto, Canada: Ministry of Transportation, Safety Research Office, Safety Policy Branch.

Lorenz, E. N. (1963). Deterministic periodic flow. *Journal of the Atmospheric Sciences, 20*, 130–141.

Lorist, M. M., & Faber, L. G. (2011). Consideration of the influence of mental fatigue on controlled and automatic cognitive processes. In P. Ackerman (Ed.), *Cognitive fatigue* (pp. 105–126). Washington, DC: American Psychological Association.

Lorist, M. M., Klein, M., Nieuwenhaus, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology, 37*, 614–625.

Louviere, A. J., & Jackson, J. T. (1980). Man–machine design for spaceflight. T. S. Cheston & D. L. Winter (Eds.), *Human factors of outer space production* (pp. 97–112). Boulder, CO: Westview Press and American Association for the Advancement of Science.

Love, L., & Johnson, C. (2002). *Using diagrams to support the analysis of system "failure" and operator "error."* Retrieved from http://www.dcs.gla.uk/~johnson/papers/aft.html

Lovsund, P., Hedin, A., & Tornros, J. (1991). Effects on driving performance of visual field effects: A driving simulator study. *Accident Analysis and Prevention, 23,* 231–342.

MacDonald, J. E., & Gifford, R. (1989). Territorial cues and defensible space theory: The burglar's point of view. *Journal of Environmental Psychology, 9,* 193–205.

Macek, A. J. (1987). Successful implementation of an injury prevention *program. Proceedings of the Human Factors Society, 31,* 1138–1140.

Madhaven, P., & Wiegman, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8,* 277–302.

Madhyastha, T. M., & Reed, D. A. (1994). A framework for sonification design. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 267–290). New York: Oxford University Press.

Maikala, R. V., Cirello, V. M., Dempsey, P. G., & O'Brien, N. V. (2009). Psychophysiologcial reponse in women during cart pushing on different frictional walkways. *Human Factors, 51*, 681–693.

Malaviya, P., & Ganesh, K. (1976). Shift work and individual differences in the productivity of weavers in an Indian textile mill. *Journal of Applied Psychology, 61,* 714–176.

Malaviya, P., & Ganesh, K. (1977). Individual differences in productivity across type of work shift. *Journal of Applied Psychology, 62*, 527–528.

Mandelbrot, B. B. (1983). *The fractal geometry of nature*. New York: Freeman.

Manzey, D., Lorenz, B., Heuer, H., & Sangals, J. (2000). Impairments of manual tracking performance during spaceflight: More converging evidence from a 20-day space mission. *Ergonomics, 43,* 589–609.

March, J. G. (2002). Forward. In A. Lomi & E. R. Larsen (Eds.), *Dynamics of organizations: Computational modeling and organization theories* (pp. ix–xvii). Cambridge, MA: MIT Press.

Marken, R. S. (1991). Degrees of freedom in behavior. *Psychological Science, 2*, 86–91.

Marklin, R. W., Saginus, K. A., Seeley, P., & Freier, S. H. (2010). Comparison of anthropometry of U.S. electric utility field-workers with North American general populations. *Human Factors, 52*, 463–662.

Marks, E. S. (1935). Individual differences in work curves. *Archives of Psychology, 186.*

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology, 87*, 3–13.

Marks, M. L., Mirvis, P. H., Hackett, E. J., & Grady, J. F., Jr. (1986). Employee participation in a quality circle program: Impact on quality of work life, productivity, and absenteeism. *Journal of Applied Psychology, 71,* 61–69.

Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying attentional blink and psychological refractory period. *NeuroImage, 59*, 2883–2898.

Martinez, R. M., & Brito, S. V. (1994). *Looking for participation: Cuban experiences applying the feedback model* (Tech. Rep.). Havana, Cuba: Faculty of Industrial Engineering, ISPJAE.

Masalonis, A. J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46*, 1045–1074.

Masud, M., Khan, L., & Thuraisingham, B. (2012). *Data mining tools for malware detection*. Boca Raton, FL: CRC Press/Taylor & Francis.

Matsumoto, A., & Szidarovszky, F. (2012, March). *Discrete-time delay dynamics of boundedly rational monopoly*. Paper presented to the 5th International Nonlinear Science Conference, Barcelona, Spain.

Matthews, G., & Campbell, S. E. (2009). Sustained performance under overload: Personality and individual differences in stress and coping. *Theoretical Issues in Ergonomics Science, 10*, 417–443.

Matthews, G., & Campbell, S. E. (2009). Sustained performance under overload: Personality and individual differences in stress and coping. *Theoretical Issues in Ergonomics Science, 10*, 417–443.

Matthieu, J. E., Heffher, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273–283.

Mavrommati, I. (2001). Creativity techniques. In K. Baumann & B. Thomas (Eds.), *User interface design for electronic appliances* (pp. 51–76). New York: Taylor & Francis.

Mayer, J. D. (2001). A field guide to emotional intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (Eds.), *Emotional intelligence in everyday life: A scientific inquiry* (pp. 3–24). Philadelphia, PA: Taylor & Francis.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). New York: Basic Books.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95,* 806–812.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187–198.

Mayer-Kress, G. (1994, November). *Global brains as paradigm for a complex adaptive world.* Paper presented at the conference on Evolving Complexity: Challenges to Society, Economy and the Individual, University of Texas at Dallas, Richardson.

Mayer-Kress, G., Newell, K. M., & Liu, Y.-T. (2009). Nonlinear dynamics of motor learning. *Nonlinear Dynamics, Psychology, and Life Sciences, 13*, 3–26.

Mayhew, D. J. (2003). Introduction. In J. Ratner (Ed.), *Human factors and web development* (2nd ed., pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge, UK: Cambridge University Press.

McAfee, R. B., & Winn, A. R. (1989). The incentives/feedback to enhance work place safety: A critique of the literature. *Journal of Safety Research, 20,* 7–19.

McAnally, K. I., & Martin, R. L. (2007). Spatial audio displays improve the detection of target messages in a continuous monitoring task. *Human Factors, 49*, 688–695.

McCallum, M., & Richard, C. (2008). Human factors analysis of pipeline monitoring and control operations: Final technical report. Retrieved from http://primis.phmsa.dot.gov/crm/FinalTechnicalReportNov2008.pdf

McComb, S., Kennedy, D., Perryman, R., & Warner, N. (2010). Temporal patterns of mental model convergence: Implications for distributed teams interacting in electronic collaboration spaces. *Human Factors, 52*, 264–281.

McCormick, E. J. (1976). *Human factors engineering and design* (4th ed.). New York: McGraw-Hill.

McCoy, J. M. (2002). Work environments. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 443–360). New York: Wiley.

McDaniel, R. R., Jr., & Driebe, D. J. (Eds.). (2005). *Uncertainty and surprise in complex systems*. New York: Springer-Verlag.

McEvoy, L. K., Smith, M. E., & Gervins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cerebral Cortex, 8*, 563–574.

McGrath, J. E. (1976). Stress and behavior in organizations. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1031–1062). Chicago: Rand McNally.

McGregor, D. M. (1960). *The human side of enterprise*. New York: McGraw-Hill.

McKelvey, B., & Lichetenstein, B. B. (2007). Leadership in the four stages of emergence. In J. K. Hazy, B. B. Lichtenstein, & J. Goldstein (Eds.), *Complex systems leadership theory* (pp. 93–107). Litchfield Park, AZ: ISCE.

McKendree, J., & Anderson, J. R. (1987). Effect of practice on knowledge and use of basic Lisp. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction* (pp. 236–259). Cambridge, MA: MIT Press.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*, 196–204.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1956). Wanted: A good cookbook. *American Psychologist, 11,* 262–272.

Meister, D. (1977). Implications of the system concept for human factors research methodology. *Proceedings of the Human Factors Society, 21,* 453–456.

Menckel, E., & Carter, N. (1985). Development and evaluation of accident prevention routines: A case study. *Journal of Safety Research, 20,* 73–82.

Menckel, E., Carter, N., & Hellbom, M. (1993). The long-term effects of two group routines on accident prevention activities and accident statistics. *International Journal of Industrial Ergonomics, 12,* 301–309.

Mendeloff, J. (1979). *Regulating safety: An economic and political analysis of occupational safety and health policy.* Cambridge, MA: MIT Press.

Merat, N., & Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. *Human Factors, 54*, 681–686.

Meshkati, N. (1989). An etiological investigation of micro- and macro-economic factors in the Bhopal disaster: Lessons for both industrialized and developing countries. *International Journal of Industrial Ergonomics, 4*, 161–175.

Meyer, D. E., Smith, J. E., Kornblum, S., Abrams, R. A., & Wright, C. E. (1990). Speed–accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and performance* (Vol. 13, pp. 173–226). Hillsdale, NJ: Lawrence Erlbaum Associates.

Michalewicz, Z. (1993). A hierarchy of evolution programs: An experimental study. *Evolutionary Computation, 1,* 51–76.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60–67.

Militello, L. G., & Klein, G. (2013). Decision-centered design. In J. D. Lee & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 261–271). New York: Oxford University Press.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Bulletin, 63,* 81–97.

Minelli T. A. (2009). Neurodynamics and Electrocortical Activity. In S. J. Guastello, M. Koopmans, and D. Pincus (Eds.), *Chaos and complexity in psychology: Theory of nonlinear dynamical systems* (pp. 73–104). New York: Cambridge University Press.

Mintz, A., & Blum, M. L. (1949). A re-examination of the accident proneness concept. *Journal of Applied Psychology, 33,* 195–221.

Mkrtchyan, A. A., Macbeth, J. C., Solovey, E. T., Ryan, J. C., & Cummings, M. L. (2012). Using variable-rate alerting to counter boredom in human supervisory control. *Proceedings of the Human Factors and Ergonomics Society, 56*, 1441–1445.

Modell, M., & Spurlock, J. M. (1980). Rationale for evaluating a closed-loop food chain for space habitats. In T. S. Cheston & D. L. Winter (Eds.), *Human factors of outer space production* (pp. 133–146). Boulder, CO: Westview Press and American Association for the Advancement of Science.

Moemeng, C., Gorosetsky, V., Zuo, Z., Yang, Y., & Zhang, C. (2009). Agent-based distributed data mining: A survey. In L. Cao (Ed.), *Data mining and multi-agent integration* (pp. 49–60). New York: Springer.

Moemeng, P., Cao, L., & Zhang, C. (2008). F-TRADE 3.0: An agent-based integrated framework for data mining experiments. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 612–615). Los Alamitos, CA: IEEE Computer Society.

Mohr, A., & Clemmer, D. L. (1989). Evaluation of an occupational injury intervention in the petroleum drilling industry. *Accident Analysis and Prevention, 21,* 263–271.

Moore, J. S., & Garg, A. 1998. The effectiveness of participatory ergonomics in the red meat packing industry: Evaluation of a corporation. *International Journal of Industrial Ergonomics, 21*, 47–58.

Morgan, H. M., Klein, C., Boehm, S. G., Shapiro, K. L., & Linden, D. E. J. (2008). Working memory load for faces modulates P300 N170 N250r. *Journal of Cognitive Neuroscience, 20*, 989–1002.

Morineau, T., Frénod, E., Blanche, C., & Tobin, L. (2009). Turing machine as an ecological model for task analysis. *Theoretical Issues in Ergonomics Science, 10*, 511–530.

Morris, C. (1906). *The San Francisco calamity, earthquake, and fire.* City Unknown: W. E. Scull.

Morris, J. C., Morris, E. D., & Jones, D. M. (2007). Reaching for the philosopher's stone: Contingent coordination and the military's response to hurricane Katrina. *Public Administration Review, 67*, 94–106.

Morrow, P. M., & Fiore, S. M. (2013). Team cognition: Coordination across individuals and machines. In J. D. Lee & A. Kirklik (Eds.), *Oxford handbook of cognitive engineering* (pp. 200–215). New York: Oxford University Press.

Mosso, A. (1894). *La fatigue intellectuelle et physique* [Mental and physical fatigue]. Paris, France: Alcon.

Mosso, A. (1915). *Fatigue.* New York: Putnam.

Murphy, L. R. (1984). Occupational stress management: A review and appraisal. *Journal of Occupational Psychology, 57*, 1–16.

Murphy, L. R., DuBois, D., & Hurrell, J. J. (1986). Accident reduction through stress management. *Journal of Business and Psychology, 1*, 5–18.

Murray, D. J. (1990). Fechner's later psychophysics. *Canadian Psychology, 31,* 54–60.

Mykytyn, P. P., Jr. (1989). Group embedded figures test (GEFT): Individual differences, performance, and learning effects. *Educational and Psychological Measurement, 49*, 951–959.

Nadolski, J. N., & Sandonato, C. E. (1987). Evaluation of an employee assistance program. *Journal of Occupational Medicine, 29,* 32–37.

Nagashima, S., Suwazono, Y., Okubo, Y., Mirei, U., Kobayaski, E., Kido, T., & Nogawa, K. (2007). Working hours and mental and physical fatigue in Japanese workers. *Occupational Medicine, 57*, 449–452.

Naitoh, P., Kelly, T. L., Hunt, P., & Goforth, H. W. (1994). *Sleep during SEAL delivery vehicle (SDV)/dry dock shelter exercises analyzed by a graphic approach* (Tech. Rep. No. 94–30). San Diego, CA: U.S. Naval Health Research Center.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics, 54,* 286–295.

Natani, K. (1980). Future directions for selecting personnel. T. S. Cheston & D. L. Winter (Eds.), *Human factors of outer space production* (pp. 25–63). Boulder, CO: Westview Press and American Association for the Advancement of Science.

Nathan, D. E., Guastello, S. J., Prost, R. W., & Jeutter, D. C. (2012). Understanding neuromotor strategy during functional upper extremity tasks using symbolic dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences, 16,* 37–59.

Nathan, D. E., Johnson, M. J., & McGuire, J. R. (2009). Design and validation of low-cost assistive glove for hand assessment and therapy during activity of daily living-focused robotic stroke therapy. *Journal of Rehabilitation Research and Development, 46*, 587–602.

Nathan, D. E., Prost, R. W., Guastello, S. J., & Jeutter, D. C. (in press). Understanding the importance of natural neuromotor strategy in upper extremity neuroprosthetic control. *International Journal of Bioinformatics Research and Applications*.

National Aeronautics and Space Administration. (2004). *Lesson 2: Technology interface design.* Retrieved from http://humanfactors.arc.nasa.gov/web/hfl01/technology.html

National Institute of Occupational Safety and Health. (1994). *Applications manual for the revised NIOSH lifting equation*. Washington, DC: Author. Retrieved from www.cdc/gov/niosh/docs/94–110/

National Institute for Occupational Safety and Health. (1997). *Elements of ergonomics programs: A primer based on workplace evaluations of musculoskeletal disorders* (DHHS NIOSH Publication No. 97–117). Cincinnati, OH: Author.

National Institute of Occupational Safety and Health. (2007). *Assistant grain elevator supervisor dies after being engulfed in shelled corn—North Carolina*. Washington, DC: Author. Retrieved from http://www.cdc.gov/niosh/face/In-house/full9416.html

National Institute of Occupational Safety and Health. (2011). *Confined spaces*. Washington, DC: Author. Retrieved from http://www.cdc.gov/niosh/topics/confinedspace/

National Motorist Association. (2013). State speed limit chart. Retrieved from http://www.motorists.org/speed-limits/state-chart

National Safety Council. (1989). *Accident facts, 1989 edition.* Chicago: Author.

National Safety Council. (2002). *International accident facts* (3rd ed.). Lyle, IL: Author.

Nayar, S. N., & Anand, V. N. (2007). 3D display using passive optical scatterers. *IEEE Computer Magazine*, *40*(7), 54–63.

Nayar, S. N., & Anand, V. N. (2013). Volumetric displays: Passive optical scatters. Retrieved from http://www.cs.columbia.edu/CAVE/projects/3d_display

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology, 91*, 946–953.

Neerincx, M. A., & Griffioen, E. (1996). Cognitive task analysis harmonizing tasks to human capabilities. *Ergonomics, 39,* 543–561.

Neitzel, R. L., Seixas, N. S., Harris, M. J., & Camp, J. (2008). Exposure to fall hazards and safety climate in the aircraft maintenance industry. *Journal of Safety Research, 39*, 391–402.

NEOS Technologies. (2004). *3-D volumetric display.* Retrieved from http://neostech.com/neos/3d.htm

Neubauer, C., Matthews, G., Langham, L., & Saxby, D. (2012). Fatigue and voluntary utilization of automation in simulated driving. *Human Factors, 54*, 734–746.

Neuhoff, J. G., Wayand, J., & Kramer, G. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied*, *8,* 1725.

Newell, A., & Card, S. K. (1985). The prospects for psychological science in human–computer interaction. *Human–Computer Interaction, 1,* 209–242.

Newhouse, R., Ruelle, D., & Takens, F. (1978). Occurrence of strange attractors: An axiom near quasiperiodic flows on Tm, m≥3. *Communications in Mathematical Physics*, *64*, 35–41.

Newman, M., Barabási, A.-L., & Watts, D. J. (Eds.). (2006). *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.

Newman, O. (1972). *Defensible space.* New York: Macmillan.

Ngo, M. K., Vu, K.-P. L., Thorpe, E., & Strybel, T. Z. (2012). Intuitiveness of symbol features for air traffic management. *Proceedings of the Human Factors and Ergonomics Society, 56*, 1804–1808.

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity*: New York: Freeman.

Nielsen, K. T. (2000). Organization theories implicit in various approaches to OHS management. In K. Frick, P. Jensen, M. Quinlan, & T. Wilthagen (Eds.), *Systematic occupational health and safety management: Prospectives on an international development* (pp. 99–123). Oxford, UK: Pergamon Press.

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics.* Cambridge, MA: MIT Press.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Norman, K. L. (2008). Better design of menu selection systems through cognitive psychology and human factors. *Human Factors, 50*, 556–559.

Noyes, J. M., & Bruneau, D. P. J. (2007). A self-analysis of the NASA-TLX workload measure. *Ergonomics, 50*, 514–519.

Noyes, J. M., & Frankish, C. R. (1992). Speech recognition technology for individuals with disabilities. *Augmentative and Alternative Communication, 8,* 297–303.

Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, *39*, 36–45.

Occupational Safety and Health Administration. (1996). *Occupational noise exposure compliance standard 1910.95.* Retrieved from www.osha.gov/SLTC/noisehearingconservation/compliance.html

Occupational Safety and Health Administration. (1996). *Specifications for accident prevention signs and tags 1910-. 145.* Retrieved from www.osha.gov/pls/oshaweb/uwadisp.show_document?p_table = standards&p_id-9794

Occupational Safety and Health Administration. (2002). *Effective ergonomics: A four-pronged comprehensive approach. U.S. Department of Labor.* Retrieved from http://osha.gov/ergonomics/ergofact02.html

O'Connor, O'Dea, & Melton, J. (2007). A methodology for identifying human error in U.S. navy diving accidents. *Human Factors, 49*, 214–226.

Ohsawa, Y., & Yada, K. (2009). *Data mining and marketing*. Boca Raton, FL: CRC Press.

Oldham, G. R., & Fried, Y. (1987). Employee reactions to workspace characteristics. *Journal of Applied Psychology, 72,* 75–80.

Oliver, M. N., & Berke, E. (2010, January). *The raw and the cooked: Smoothing spatial data*. Paper presented to the 22nd annual Primary Care Research Methods & Statistics Conference, San Antonio, TX.

Olson, J. R. (1987). Cognitive analysis of people's use of software. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction* (pp. 260–293). Cambridge, MA: MIT Press.

Omodei, M. M., McLennan, J., & Wearing, A. J. (2005). How expertise is applied in real-world decision environments: Head-mounted video and cued recall as a methodology for studying routines of decision making. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 271–288). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neill, B. (1984). Structures for nonhierarchical organizations. *Behavioral Science, 29*, 61–77.

Open Ergonomics. (2000). *PeopleSize 2000: The shape of customers to come.* Retrieved from www.openerg.com/psz/onestop.htm

Ophir-Arbelle, R., Oron-Gilad, T., Borowsky, A., & Parmet, Y. (2013). Is more information better? How dismounted soldiers use video feed from unmanned vehicles: Attention allocation and information extraction considerations. *Journal of Cognitive Engineering and Decision Making, 7,* 26–48.

Ott, E., Grebogi, C., & Yorke, J. A. (1990). Controlling chaos. *Physical Review Letters, 64,* 1196–1199.

Ott, E., Sauer, T., & Yorke, J. A. (Eds.). (1994). *Coping with chaos.* New York: Wiley.

Palinkas, L. A. (1987). *Antarctica as a model for the human exploration of Mars* (Tech. Rep. No. 87-16). San Diego, CA: Naval Health Research Center.

Papantonopoulos, S., & Galvendy, G. (2008). Analytic cognitive task allocation: A decision model for cognitive task allocation. *Theoretical Issues in Ergonomics Science, 9,* 155–185.

Parasuraman, R., Masalonis, A. J., & Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636–659.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, A: Systems and Humans, 50*, 468–474.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 50*, 511–520.

Park, S. H., & Woldstad, J. C. (2000). Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks. *Human Factors, 42,* 592–601.

Parkes, K. R. (1998). Psychosocial aspects of stress, health and safety on North Sea installations. *Scandinavian Journal of Work, Environment & Health, 24*, 321–333.

Pascuale-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

Pathak, S., Pokharel, M. P., & Mahadevan, S. (2013). Hyper-competition, collusion, free-riding or coopetition: Basins of attraction when firms simultaneously complete and cooperate. *Nonlinear Dynamics, Psychology, and Life Sciences, 17,* 133–157.

Patrick, J., & Morgan, P. L. (2010). Approaches to understanding, analyzing and developing situation awareness. *Theoretical Issues in Ergonomics Science, 11*, 41–57.

Pauchant, T. C., & Mitroff, I. I. (1992). *Transforming the crisis-prone organization*. San Francisco: Jossey-Bass.

Pazzani, M. (1987a). Explanation-based learning for knowledge-based systems. *International Journal of Man–Machine Studies, 26,* 413–433.

Pazzani, M. (1987b). *Inducing causal and social theories: A prerequisite for explanation-based learning* (Tech. Rep.). Los Angeles: University of California Artificial Intelligence Laboratory.

Peterson, J. L. (1981). *Petri net theory and the modeling of systems.* Englewood Cliffs, NJ: Prentice Hall.

Peterson, L. E., Pepper, L. J., Hamm, P. B., & Gilbert, S. L. (1993). Longitudinal study of astronaut health: Mortality in the years 1959–91. *Radiation Research, 133,* 257–264.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *IRE Professional Group on Information Theory PGIT-4, X*, 171–212.

Pheasant, S. (1996). *Bodyspace: Anthropometry, ergonomics, and the design of work.* Bristol, PA: Taylor & Francis.

Pheasant, S., & Haslegrave, C. (2005). *Bodyspace: Anthropometry, ergonomics and the design of work* (3rd ed.). London: Taylor & Francis.

Pidgeon, N. F. (1991). Safety culture and risk management in organizations. *Journal of Cross-Cultural Psychology, 22,* 129–140.

Pimental, K., & Texiera, K. (1995). Virtual reality: Through the new looking glass (2nd ed.). New York: McGraw-Hill.

Pinch, T., & Trocco, F. (2002). *Analog days: The invention and impact of the Moog synthesizer*. Cambridge, MA: Harvard University Press.

Pincus, D., & Metten, A. (2010). Nonlinear dynamics in biopsychosocial resilience. *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 353–380.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Procedings of the National Academic of Sciences USA, 88*, 2297–2301.

Pincus, S. M., & Goldberger, A. L. (1994). Physiological time-series analysis: What does regularity quantify? *American Journal of Physiology, 266*(Pt. 2), H1643–1656.

Pirolli, P. (2007). *Information foraging theory*. New York: Oxford University Press.

Poffenberger, A. T. (1928). The effects of continuous work upon output and feelings. *Journal of Applied Psychology, 12*, 259–467.

Porter, B. E., & Berry, T. D. (2001). A nationwide survey of self-reported red light running: Measuring prevalence, predictors, and perceived consequences. *Accident Analysis and Prevention, 33,* 735–742.

Post, D. L., Geiselman, E. E., & Goodyear, C. D. (1999). Benefits of color coding weapons symbology for an airborne helmet-mounted display. *Human Factors, 41*, 515–523.

Postman, K., Hermens, H. J., De Vries, J., Koopman, H. F. J. M., & Eisma, W. H. (1997). Energy storage and release of prosthetic feet, part 1: Biomechanical analysis related to user benefits. *Prosthetics and Orthotics International, 21*, 17–27.

Potvin, J. R. (2012). Predicting maximum acceptable efforts for repetitive tasks: An equation based on duty cycle. *Human Factors, 54*, 175–188.

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of working memory on visual selective attention: Behavioral and electrophysiological evidence. *Frontiers in Human Neuroscience, 5*, Article D57.

Prince, R. J., & Guastello, S. J. (1990). The Barnum effect in a computerized Rorschach interpretation system. *Journal of Psychology: Interdisciplinary and Applied, 124,* 217–222.

Priplata, A. A. (2006). Noise-enhanced balance control in patients with diabetes and patients with stroke. *Annals of Neurology, 59*, 4–12.

Quick, J. C., & Quick, J. D. (1984). *Organizational stress and preventive management*. New York: McGraw-Hill.

Quick, J. C., Murphy, L. R., & Hurrell, J. J., Jr. (Eds.). (1992). *Stress and well-being at work*. Washington, DC: American Psychological Association.

Quinlan, M. (1988). Psychological and sociological approaches to the study of occupational illness: A critical review. *Australia and New Zealand Journal of Sociology, 24,* 189–207.

Ralph, J., Gray, W. D., Schoelles, M. J. (2010). Squeezing the balloon: Analyzing the unpredictable effects of cognitive workload. *Proceedings of the Human Factors and Ergonomics Society, 54*, 299–303.

Rambaut, P. C. (1980). Nutritional criteria for closed–loop space food systems. In T. S. Cheston & D. L. Winter (Eds.), *Human factors of outer space production* (pp. 113–132). Boulder, CO: Westview Press and American Association for the Advancement of Science.

Rankin, F. W., Van Huyck, J. B., & Battalio, R. C. (2000). Strategic similarity and emergent conventions: Evidence from similar stag hunt games. *Games and Economic Behavior, 32*, 315–337.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: Wiley.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional link? *Journal of Experimental Psychology: Human Perception, and Performance, 18*, 849–860.

Reagan, R. T., Mostellar, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology, 74,* 433–442.

Reason, J. (1997). *Managing the risks of organizational accidents.* Brookfield, VT: Ashgate.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 308–324.

Rees, J. V. (1988). *Reforming the workplace: A study of self-regulation in occupational safety.* Philadelphia, PA: University of Pennsylvania Press.

Rege, A., & Agogino, A. M. (1988). Topological framework for representing and solving probabilistic inference problems in expert systems. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 402–414.

Reinerman-Jones, L., Taylor, G., Sprouse, K., Barber, D., & Hudson, I. (2011). Adaptive automation as a task switching and task congruence challenge. *Proceedings of the Human Factors and Ergonomics Society*, *55*, 197–201.

Reising, J. M., & Curry, D. G. (1987). A comparison of voice and multifunctional controls: Logic design is the key. *Ergonomics, 30,* 1063–1077.

Reiss, E. (2012). *Usable usability: Simple steps for making stuff better*. Indianapolis, IN: Wiley.

Reivich, K. J., Seligman, M. E. P., & McBride, S. (2011). Master resilience training in the U.S. army. *American Psychologist, 66,* 25–34.

Rempel, D. (2008). The split keyboard: An ergonomics success story. *Human Factors, 50*, 385–392.

Rempel, D., Willms, K., Anshel, J., Jaschinski, W., & Sheedy, J. (2012). The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Human Factors, 49*, 830–838.

Renshaw, P. F., & Wiggins, M. W. (2007). A self-report critical incident assessment tool for army night vision goggle helicopter operations. *Human Factors, 49*, 200–213.

Retting, R. A., Ulmer, R. G., & Williams, A. F. (1998). *Prevalence and characteristics of red light running crashes in the United States.* Arlington, VA: Insurance Institute for Highway Safety.

Retting, R. A., Weinstein, H. B., Williams, A. F., & Preusser, D. F. (2001). A simple method for identifying and correcting crash problems on urban arterial streets. *Accident Analysis and Prevention, 33,* 723–734.

Retting, R. A., Williams, A. F., Preusser, D. F., & Weinstein, H. B. (1995). Classifying urban crashes for countermeasure development. *Accident Analysis and Prevention, 27,* 283–294.

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and compensation hypothesis. *Current Directions in Psychological Science, 17*, 177–182.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21,* 25–34.

Rice, S., & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *The Journal of General Psychology, 137*, 362–375.
Riley, J. M., Endsley, M. R., Boldstad, C. A., & Cuevas, H. M. (2006). Collaborative planning and situation awareness in Army command and control. *Ergonomics, 49*, 1139–1153.
Riva, G., Davide, F., & Ijsselsteijn, W. A. (Eds.). (2003). *Being there: Concepts effects and measurements of user presence in synthetic environments.* Amsterdam, Netherlands Antilles: IOS Press.
Rizzo, A. A., Neumann, U., Enciso, R., Fidaleo, D., & Noh, J. Y. (2001). Performance-driven facial animation: Basic research on human judgments of emotional state in facial avatars. *Cyberpsychology and Behavior, 4,* 471–487.
Robertson, C. C., & Morgan, B. J. T. (1990). Aspects of the design and analysis of signal detection experiments. *British Journal of Mathematical and Statistical Psychology, 42,* 7–14.
Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35*, 747–758.
Robertson, L. S., & Keeve, J. P. (1983). Worker injuries: The effects of workers' compensation and OSHA inspections. *Journal of Health Policy Law, 8*, 581–597.
Robson, L. S., Clarke, J. A., Cullen, K., Bielecky, A., Severin, C., Bigelow, P. L., Irwin, E., Culyer, A., & Mahood, Q. (2007). The effectiveness of occupational health and safety management system interventions: A systematic review. *Safety Science, 45*, 329–353.
Roebuck, J. A., Jr. (1995). *Anthropometric methods: Designing to fit the human body.* Santa Monica, CA: Human Factors and Ergonomics Society.
Roebuck, J. A., Jr., Kroemer, K., & Thomson, W. (1975). *Engineering anthropometry methods.* New York: Wiley Interscience.
Roscoe, S. N. (2006). The adolescence of engineering psychology. In J. Stuster (Ed.), *The Human Factors and Ergonomics Society: Stories from the first 50 years* (pp. 12–14). Santa Monica, CA: Human Factors and Ergonomics Society.
Rosenbaum, D. A. (2005). The Cinderella of psychology: The neglect of motor control in the science of mental life and behavior. *American Psychologist, 60,* 308–317.
Rosenbaum, D. A., Slotta, J. D., Vaughn, J., & Plamondon, R. (1991). Optimal movement selection. *Psycholoical Science, 2*, 92–101.
Rothman, A. L., & Weintraub, M. I. (1995). The sick building syndrome and mass hysteria. *Neurologic Clinics, 13,* 405–412.
Rotter, J. B. (1966). Generalized expectancies for the internal versus external locus of control of reinforcement. *Psychological Monographs, 80.*
Rowe, W. D. (1977). *An anatomy of a risk.* New York: Wiley.
Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*, 1059–1069.
Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 763–797.
Russell, S. M., Funke, G. J., Knott, B. A., & Strang, A. J. (2012). Recurrence quantification analysis used to assess team communication in simulated air battle management. *Proceedings of the Human Factors and Ergonomics Society, 56*, 468–472.
Saal, F. E., & Knight, P. A. (1988). *Industrial/organizational psychology: Science and practice*. Belmont, CA: Wadsworth.
Saarela, K. L. (1990). An intervention program utilizing small groups: A comparative study. *Journal of Safety Research, 21*, 149–156.
Sadeghipour, A., Yaghoubzadeh, R., Rüter, A., & Kopp, S. (2009). Social motorics: Towards an embedded basis of social human–robot interaction. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Human centered robot systems: Cognition, interaction, technology* (pp. 193–203). New York: Springer.
Sakei, K. (2001). *Nonlinear dynamics and chaos in agricultural systems.* Amsterdam, the Netherlands: Elsevier.

Salminen, S., & Heiskanen, M. (1997). Correlations between traffic, occupational, sports, and home accidents. *Accident Analysis and Prevention, 29,* 33–36.

Salmon, P. M., Stanton, N. A., Walker, G. H., Barber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*, 297–323.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9*, 185–211.

Salzer, Y., Oron-Gilad, T., Ronen, A., & Parmet, Y. (2011). Vibrotactile "on-high" alerting system in the cockpit. *Human Factors, 53*, 118–131.

Samuelson, L. (1997). *Evolutionary games and equilibrium selection.* Cambridge, MA: MIT Press.

Sandal, G. M. (2000). Coping in Antarctica: Is it possible to generalize results across settings? *Aviation, Space, and Environmental Medicine, 71,* A37–A43.

Sandal, G. M., Endresen, I. M., Vaernes, R., & Ursin, H. (1999). Personality and coping strategies during submarine missions. *Military Psychology, 11,* 381–404.

Sanders, A. F. (1991). Simulation as a tool in the measurement of human performance. *Ergonomics, 34,* 995–1025.

Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering design.* New York: McGraw-Hill.

Sarason, I. G., Sarason, B. R., Potter, E. H., & Antoni, M. H. (1985). Life events, social support, and illness. *Psychosomatic Medicine, 47,* 156–163.

Sarfi, R. J., Tao, M. K., & Gemoets, L. (2011). Making the smart grid work for community energy delivery. *Information Polity, 16*, 267–281.

Sarter, N. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors, 50*, 506–510.

Sartory, G., Heine, A., Muleeler, B. W., & Elvermann-Hallner, A. (2002). Event- and motor-related potentials during the continuous performance task in attention-deficit/hyperactivity disorder. *Journal of Psychophysiology, 16,* 97–106.

Sauter, S. L., & Murphy, L. R. (Eds.). (1995). *Organizational risk factors for job stress*. Washington, DC: American Psychological Association.

Sawyer, R. K. (2005). *Social emergence: Societies as complex systems*. New York: Cambridge University Press.

Scerbo, M. W. (2001). Adaptive automation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1077–1079). London: Taylor & Francis.

Schels, M., Thiel, C., Schwenker, F., & Palm G. (2009). Classifier fusion applied to facial expression recognition: An experimental comparison. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Human centered robot systems: Cognition, interaction, technology* (pp. 121–130). New York: Springer.

Scherer, S., Schwenker, F., Campbell, N., & Palm, G. (2009). Multimodal laughter detection in natural discourses. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), Human centered robot systems: Cognition, interaction, technology (pp. 111–120). New York: Springer.

Schiflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (Eds.). (2004). *Scaled worlds: Development, validation, and applications*. Burlington, VT: Ashgate.

Schifrman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications.* New York: Academic Press.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T., Jr. (1979). Motor output variability: A theory for the accuracy of rapid motor acts. *Psychological Review, 86,* 415–451.

Schmidt, W. C. (2001). Presentation accuracy of web animation methods. *Behavioral Research Methods, Instrumentation, and Computers, 33,* 187–200.

Schmorrow, D. D., & Stanney, K. M. (2008). *Augmented cognition: A practitioner's guide*. Santa Monica, CA: Human Factors and Ergonomics Society.

Schneider, B., & Reichers, A. E. (1983). On the etiology of climates. *Personnel Psychology, 36*, 19–39.

Schneider-Garces, N. J., Gordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., Maclin, E. L., Gratton, G., & Fabiani, M. (2009). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience, 22*, 655–669.

Schraagen, J. M., Chipman, S. F., & Shute, V. J. (2000). State-of-the-art review of cognitive task analysis techniques. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 467–487). Mahwah, NJ: Lawrence Erlbaum Associates.

Schutte, N. S., Malouf, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheirn, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*, 167–177.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*, 163–196.

Seligman, M. P. (1975). *Helplessness: On depression, development and death.* San Francisco: Freeman.

Sellnow, T. L., Seeger, M. W., & Ulmer, R. R. (2002). Chaos theory, informational needs, and natural disasters. *Journal of Applied Communications Research, 30*, 269–292.

Selye, H. (1976). *The stress of life* (2nd ed.). New York: McGraw-Hill.

Seminara, J. L., Gonzales, W. R., & Parsons, S. O. (1977). *Human factors review of nuclear power plant control room design* (Tech. Rep. No. EPRINP-309). Sunnyvale, CA: Lockheed Missiles and Space.

Semovski, S. V. (2001). Self-organization in fish school. In W. Sulis & I. Trofimova (Eds.), *Nonlinear dynamics in the life and social sciences* (pp. 398–406). Amsterdam, the Netherlands: IOS Press.

Sengupta, S., & O'Brien, K. J. (September 21, 2012). Facebook can ID faces, but using them grows tricky. Wall Street Journal, p. A1. Retrieved from http://www.nytimes.com/2012/09/22/technology/facebook-backs-down-on-face-recognition

Seo, M.-G., Putnam, L. L., & Bartunek, J. M. (2004). Dualities and tensions of planned organizational change. In M. S. Poole & A. H. Van de Ven (Eds.), *Handbook of organizational change and innovation* (pp. 73–107). New York: Oxford University Press.

Serfaty, D., Entin, E., & Deckert, J. (1993). Team adaptation to stress in decision making and coordination with implications for CIC team training (Report No. TR-564). Burlington, MA: ALPHATECH.

Sesto, M. E., Irwin, C. B., Chen, K. B., Chourasia, A. O., & Weigman, D. A. (2012). Effect of touch screen button size and spacing on touch characteristics of users with and without disabilities. *Human Factors, 54*, 425–436.

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors, 52*, 234–245.

Shakioye, S. O., & Haight, J. M. (2010). Modeling using dynamics variables: An approach for the design of loss prevention programs. *Safety Science, 48*, 46–53.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana, IL: University of Illinois Press.

Shannon, H. S., & Guastello, S. J. (1996, February). *What do we know really works in safety?* Paper presented at the third International Conference in Injury Prevention and Control, Melbourne, Australia.

Shannon, H. S., & Guastello, S. J. (1997, June). *Workplace safety interventions: What can we say about what works?* Paper presented at the International Ergonomics Association Conference, Tempere, Finland.

Sharma, S., & Ivancevic, V. G. (2010). Nonlinear dynamical characteristics of situation awareness. *Theoretical Issues in Ergonomics Science, 11*, 448–460.

Shelton, R. L. (2001). *Driver distractions: Electronic devices in the automobile.* Retrieved from http://www-nrd.nhtsa.dot.gov/PDF/nrd-13/Nov1_ppt.pdf

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika, 27*, 219–246.

Shepard, R. N. (1982). Geometrical approximation to the structure of musical pitch. *Psychological Review, 89*, 305–332.

Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation study. *Psychological Bulletin, 113*, 67–81.

Sheridan, T. B. (2002). *Human and automation: System design and research issues*. New York: Wiley.
Sheridan, T. B. (2008). Risk, human error, and system resilience: Fundamental ideas. *Human Factors, 50*, 418–426.
Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human–computer interaction*. Reading, MA: Addison-Wesley.
Shneiderman, B. (2000). *The eyes have it: User interfaces for information visualization*. Retrieved from www.cs.umd.edu/hcil/pubs/presentations
Shortliffe, E. H. (1976). *Computer-based medical consultations: MYCIN*. New York: Elsevier.
Sieber, J. E. (1980). Well-being and privacy in space: Anticipating conflicts of interest. In T. S. Cheston & D. L. Winter (Eds.), *Human factors of outer space production* (pp. 65–78). Boulder, CO: Westview Press and American Association for the Advancement of Science.
Simon, H. A. (1957). *Administrative behavior* (2nd ed.). Totowa, NJ: Littlefield Adams.
Sival, M., Olson, P. L., & Pastalan, L. A. (1981). Effect of driver's age on night-time legibility of highway signs. *Human Factors, 23*, 59–64.
Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Science, 10,* 161–195.
Slater, M. (2004). Presence and emotions. *Cyberpsychology and Behavior, 7*, 121.
Sleimen-Malkoun, R., Temprado, J. J., Jirsa, V. K., & Berton, E. (2010). New directions offered by the dynamical systems approach to bimanual coordination for therapeutic intervention and research in stroke. *Nonlinear Dynamics, Psychology, and Life Sciences, 14*, 435–462.
Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience, 20*, 458–469.
Smith, C. M. (2001). Human factors in haptic interfaces. Retrieved from www.acm.org/crossroads/xrds.3–3/haptic.html#2
Smith, J. B. (1994). *Collective intelligence in computer-based collaboration*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Smith, K., & Hancock, P. A. (1995). Situational awareness is adaptive, externally directed consciousness. *Human Factors, 37*, 137–148.
Smith, L., & Iskra-Golec, I. (2003). Internal locus of control and shiftwork effects. *Theoretical Issues in Ergonomic Science, 4*, 327–339.
Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes*. Chicago: Rand McNally.
Smith, R. S. (1976). *The Occupational Safety and Health Act: Its goals and its achievements*. Washington, DC: American Enterprise Institute for Public Policy.
Smith, R. S. (1979). The Impact of OSHA inspections on manufacturing injury rates. *Journal of Human Resources, 14*, 145–170.
Smith, S. F. (1977). Running nips. In W. Handley (Ed.), *Industrial safety handbook* (2nd ed., pp. 186–197). New York: McGraw–Hill.
Smith, T. S., & Stevens, G. T. (1997). Biological foundations of social interaction: Computational explorations of nonlinear dynamics. R. A. Eve, S. Horsfall, & M. E. Lee (Eds.), *Chaos, complexity and sociology: Myths, models, and theories* (pp. 197–214). Thousand Oaks, CA: Sage.
Smith, W., & Dowell, J. (2000). A case study of coordinative decision-making in disaster management. *Ergonomics, 43*, 1153–1166.
Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133*, 780–799.
Soderquist, D. R., & Shilling, R. D. (1990). Loudness and the binaural masking level difference. *Bulletin of the Psychonomic Society, 28*, 553–555.
Spangenberg, S., Mikkelson, K. L., Kines, P., Dyreborg, J., & Baarts, C. (2002). The construction of the Oresund link between Denmark and Sweden: The effect of a multifaceted safety campaign. *Safety Science, 40*, 457–465.
Spencer, H. (2004). *Pioneers of modern typography* (2nd ed.). Cambridge, MA: MIT Press.
Spielberg, S. (Dir.). (2002). *Minority report*. Universal City, CA: Dreamworks and 20th Century Fox.
Sprott, J. C. (2003). *Chaos and time-series analysis*. New York: Oxford.

Sprott, J. C. (2013). Is chaos good for learning? *Nonlinear Dynamics, Psychology, and Life Sciences, 17*, 223–233.

Sridharan, N. S. (1985). Evolving systems of knowledge. *A. I. Magazine, 6*, 108–121.

Stachowski, A. A., Kaplan, S. A., & Waller, M. J. (2009). The benefits of flexible team interaction during crises. *Journal of Applied Psychology, 94*, 1536–1543.

Stamovlasis, D. (2006). The nonlinear dynamical hypothesis in science education problem solving: A catastrophe theory approach. *Nonlinear Dynamics, Psychology and Life Science, 10*, 37–70.

Stamovlasis, D. (2011). Nonlinear dynamics and neo-Piagetian theories in problem solving: Perspectives on a new epistemology and theory development. *Nonlinear Dynamics, Psychology and Life Science, 15,* 145–173.

Stamovlasis, D., & Tsaparlis, G. (2012). Applying catastrophe theory to an information-processing model of problem solving in science education. *Science Education, 96*, 392–410.

Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2010). Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science, 11*, 29–40.

Starch, D., & Ash, J. E. (1917). The mental curve of work. *Psychological Bulletin, 24*, 391–402.

Staszewski, J. J. (2004). Models of human expertise as blueprints for cognitive engineering: Applications to landmine detection. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, 48*, 458–462.

Staszewski, J. J. (2006). Spatial thinking and the design of landmine detection training. In G. A. Allen (Ed.), *Applied spatial cognition: From research to cognitive technology* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Staudenmeyer, N., & Lawless, M. W. (1999, November). *Organizational decay*: *How innovation stresses organizations*. Paper presented at INFORMS '99, Philadelphia.

Stergiou, B., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: Is there a connection? *Human Movement Science*, *30*, 869–888.

Stergiou, N., Harbourne, R. T., & Cavanaugh, P. T. (2006). Optimal movement variability: A new theoretical perspective for neurologic physical therapy. *Journal of Neurologic Physical Therapy, 30*, 120–128.

Sterman, J. (1988). Deterministic models of chaos in human behavior: Methodological issues and experimental results. *System Dynamics Review, 4*, 148–178.

Sterman, J. (1989). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*, 301–335.

Sternberg, R. J. (1995). *Psychology: In search of the human mind*. Orlando, FL: Harcourt.

Sternberg, R. J. (2004). *Psychology* (4th ed.). Belmont, CA: Thompson-Wadsworth.

Stevens, A., & Minton, R. (2001). In-vehicle distraction and fatal accidents in England and Wales. *Accident Analysis and Prevention, 33*, 539–545.

Stevens, M., & Campion, M. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management, 20*, 503–540.

Stevens, R., Gorman, J. C., Amazeen, P., Likens, A., & Galloway, T. (2013). The organizational neurodynamics of teams. *Nonlinear Dynamics, Psychology, and Life Sciences, 17*, 67–86.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–50). New York: Wiley.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.

Stone, M., Remington, R., & Dismukes, K. (2001). *Remembering intentions in dynamic environments.* Retrieved from http://olias.nasa.gov/reports/detecting.html

Stout, R. J., Cannon–Bowers, J., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors, 41*, 61–71.

Straub, R. O. (2001). *Faculty guide for use with Scientific American Frontiers video collection for introductory psychology* (2nd ed.). New York: Worth.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). Comparison of the cell phone driver and the drunk driver. *Human Factors, 48*, 381–391.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*, 23–52.

Strang, A. J., Horwood, S., Best, C., Funke, G. J., Knott, B. A., & Russell, S. M. (2012). Examining temporal regularity in categorical team communication using sample entropy. *Proceedings of the Human Factors and Ergonomics Society, 56*, 473–477.

Strogatz, S. (2003). *Sync: The emerging science of spontaneous order*. New York: Hyperion.

Suedfeld, P. (2003). Canadian space psychology: The future may be almost here. *Canadian Psychology, 44*, 85–92.

Suh, N. P. (2007). Ergonomics, axiomatic design and complexity theory. *Theoretical Issues in Ergonomics Science, 8*, 101–122.

Sukel, K. E., Catrambone, R., Essa, I., & Brostow, G. (2003). Presenting movement in a computer–based dance tutor. *International Journal of Human–Computer Interaction*, *15*, 433–452.

Sulis, W. (1997). Fundamental concepts of collective intelligence. *Nonlinear Dynamics, Psychology, and Life Sciences*, *1*, 35–54.

Sulis, W. (2008). Stochastic phase decoupling in dynamical networks. *Nonlinear Dynamics, Psychology, and Life Sciences, 12*, 327–358.

Sulis, W. (2009). Collective intelligence: Observations and models. In S. J. Guastello, M. Koopmans, & D. Pincus. (Eds.), *Chaos and complexity in psychology: Theory of nonlinear dynamical systems* (pp. 41–72). New York: Cambridge University Press.

Sulsky, L., & Smith, C. (2005). *Work stress*. Belmont, CA: Thompson-Wadsworth.

Surry, J. (1969). *Industrial accident research: A human engineering appraisal*. Toronto, Canada: Labor Safety Council, Ontario Ministry of Labor.

Suruda, A., Whitaker, B., Bloswick, D., Philips, P., & Sasek, R. (2002). Impact of the OSHA trench and excavation standard on fatal injury rates among residential construction workers, 1994–1998. *American Journal of Industrial Medicine 45*, 210–217.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26.

Szalma, J. L. (2009). Individual differences: Incorporating human variation into human factors/ergonomics research and practice. *Theoretical Issues in Ergonomics Science*, *10*, 377–380.

Szalma, J. L., & Teo, G. W. L. (2012). Spatial and temporal task characteristics as stress: A test of the dynamic adaptability theory of stress, workload, and performance. *Acta Psychologica, 139*, 471–485.

Tasto, D. L., Colligan, M. J., Skjel, E. W., & Polly, S. J. (1978). *Health consequences of shift work*. Washington, DC: United States Department of Health, Education, and Welfare, and National Institute for Occupational Safety and Health.

Teas, D. C. (1990). *Review of audition literature: Selection of acoustic signals for use in the synthesis of auditory space* (Rep. No. USAFSAM-TR-90-33). Brooks Air Force Base, TX: U.S. Air Force School of Aerospace Medicine.

Tenner, E. (1996). *Why things bite back: Technology and the revenge of unintended consequences*. New York: Alfred A. Knopf.

Thackray, R., & Touchstone, R. M. (1991). Effects of monitoring under high and low taskload on detection of flashing and coloured radar targets. *Ergonomics, 34*, 1065–1081.

Theeuves, J., van der Burg, E., Olivers, C. N. L., & Bronkhorst, A. (2007). Cross-modal interaction between sensory modalities: Implications for the design of multisensory displays. In Kramer, A. F., Wiegman, D. A., & Kirklik, A. (Eds.), *Attention: From theory to practice* (pp. 196–205). New York: Oxford University Press.

Thom, R. (1975). *Structural stability and morphegenesis*. New York: Addison-Wesley.

Thomas, J. C., & Gould, J. D. (1975). A psychological study of query by example. *National Computer Conference: AFIPS Conference Proceedings, nv*, 439–445.

Thompson, H. L. (2010). *The stress effect: Why smart leaders make dumb decisions—and what to do about it.* San Francisco: Jossey-Bass.

Thompson, S. J., Fraser, E. J., & Howarth, C. I. (1985). Driver behaviour in the presence of child and adult pedestrians. *Ergonomics, 10*, 1469–1474.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 33*, 268–278.

Tolman, E. C. (1932). *Purposive behavior in animals and man*. New York: Century.

Torgerson, W. G. (1958). *Theory and methods of scaling*. New York: Wiley.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 30*, 708–719.

Trainni, V. (2008). *Evolutionary swarm robotics: Evolving self-organizing behaviors in groups of autonomous robots*. Berlin, Germany: Springer.

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry, 32*, 425–443.

Travis, C. B., McLean, B. E., & Ribar, C. (Eds.). (1989). *Environmental toxins: Psychological, behavioral, and sociocultural aspects, 1973–1989. Bibliographies in Psychology, No. 5*. Washington, DC: American Psychological Association.

Trist, E. L., Susman, G. I., & Brown, G. R. (1977). An experiment in autonomous working in an American underground coal mine. *Human Relations, 30*, 201–236.

Trofimova, I. (2002). Sociability, diversity and compatibility in developing systems: EVS approach. In J. Nation, I. Trofimova, J. Rand, & W. Sulis (Eds.), *Formal descriptions of developing systems* (pp. 231–248). Dordrecht: Kluwer.

Tufano, D. R. (1997). Automotive HUDs: The overlooked safety issues. *Human Factors, 39*, 303–311.

Turing, A. M. (1963). Computing machinery and intelligence. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 11–35). New York: McGraw-Hill.

Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Wiley.

Turpin, J. A. (1981). Effects of blur and noise on digital imagery interpretability. *Proceedings of the Human Factors Society, 25*, 286–289.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*, 938–953.

Ulam, S., & Schrandt, R. (1986). Some elementary attempts at numerical modeling of problems concerning rates of evolutionary processes. *Physica D, 22*, 4–12.

U.S. Department of Defense. (1991). *Anthropometry of U.S. military personnel (metric)*. Washington, DC: Author. Technical report no. DOD-HDBK-743A. Retrieved from http://www.everyspec.com/DoD/DoD-HDBK/download.php?spec=DOD=HDBK-743A.016856.pdf

U.S. Department of Defense. (2002). *DARPA's total information awareness program homepage*. Retrieved from http://www.darpa.mil/ao/TIASystems.htm

U.S. Department of Defense. (2012). *Department of Defense design criteria standard: Human Engineering*. Washington, DC: Author. Retrieved from http://www.assistdocs.com

U.S. Department of Labor. (1977). *Dictionary of occupational titles*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1988). *The role of labor-management committees in safe-guarding worker safety and health*. Washington, DC: Author.

U.S. Department of Labor. (2004). *Occupational Safety and Health Administration Standards 1910.36–38, 1926.200 (a, b, c)*. Retrieved from www.osha.gov/pls/oshaweb

U.S. Department of Transportation. (2003). *Manual on uniform traffic control devices for streets and highways*. Retrieved from http://mutcd/fhwa.dot.gov.

U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration. (2012). *Control room management*. Retrieved from http://primis.phmsa.dot.gov/crm/index.htm

Vaillancourt, D. E., & Newell, K. M. (2000). The dynamics of resting and postural tremor in Parkinson's disease. *Clinical Neurophysiology, 111*, 2046–2056.

Vale, C. D., Keller, L. S., & Bentz, V. J. (1986). Development and validation of computerized interpretation system for personnel tests. *Personnel Psychology, 39*, 525–542.

Valenza, G., Lanata, A., Scilingo, E. P., & De Rossi, D. (2010). Towards a smart glove: Arousal recognition based on textile electrodermal response. *Proceedings of the 32nd Annual International Conference of the IEEE EMBS, 32*, 3598–3601.

van der Kleij, R., & te Brake, G. (2010). Map-mediated dialogues: Effects of map orientation differences and shared reference points on map location-finding speed and accuracy. *Human Factors, 52*, 526–536.

van der Velde, M., & Class, M. D. (1995). The relationship of role conflict and ambiguity to organizational culture. In S. L. Sauter & L. R. Murphy (Eds.), *Organizational risk factors for stress* (pp. 53–60). Washington, DC: American Psychological Association.

van Heerden, I. I. I. (2007). The failure of the New Orleans levee system following hurricane Katrina and the pathway forward. *Public Administration Review, 67*, 24–35.

Van Nes, F. (1992). Design and evaluation of applications with speech interfaces: Experimental results and practical guidelines. In M. Galer, S. Harker, & J. Ziegler (Eds.), *Methods and tools in user-centered design for information technology* (pp. 281–298). Amsterdam, the Netherlands: North-Holland.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331–350.

Verbaten, M. N., Huyben, M. A., & Kemner, C. (1997). Processing capacity and the frontal P3. *International Journal of Psychophysiology, 25*, 237–248.

Verhaegen, P. (1993). Absenteeism, accidents and risk-taking: A review ten years later. *Safety Science, 16*, 359–366.

Vicente, K. J., Mumaw, R. J., & Roth, E. M. (2004). Operator monitoring in a complex dynamic work environment: A qualitative cognitive model based on field observation. *Theoretical Issues in Ergonomics Science, 5*, 359–384.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences, 2*, 169–194.

Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of a self-organizing PAGAN (parallel, adaptive, generalized accumulator network) model for multichoice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences, 4*, 1–31.

Voiers, W. D., Sharpley, A. D., & Panzer, I. L. (2002). Evaluating the effect of noise on voice communication systems. In G. M. Davis (Ed.), *Noise reduction in speech applications* (pp. 125–152). Boca Raton, FL: CRC Press.

von Helmholtz, H. L. F. (1962). *Treatise on physiological optics* (3rd ed., J. P. C. Southall Ed. and Trans.). New York: Dover (Original work published 1909).

von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Vora, P. W. (2003). Designing friction-free experience for e-commerce sites. In J. Ratner (Ed.), *Human factors and web development* (2nd ed., pp. 225–240). Mahwah, NJ: Lawrence Erlbaum Associates.

VRex, Inc. (2004). *About VRex Inc*. Retrieved from http://vrex.com/about/index/html

Vroom, V. H. (1964). *The motivation to work*. New York: Wiley.

Vroom, V. R., & Jago, A. G. (1988). *The new leadership: Managing participation in organizations*. Englewood Cliffs, NJ: Prentice Hall.

Wade, C., Redfern, M. S., Andres, R. O., & Breloff, S. P. (2010). Joint kinetics and muscle activity while walking on ballast. *Human Factors, 52*, 560–573.

Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of chaos*. New York: Simon & Schuster.

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Dalladino, D. K., Tingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors, 55*, 157–182.

Walker, C. C., & Dooley, K. J. (1999). The stability of self-organized rule-following work teams. *Computational and Mathematical Organization Theory, 5*, 5–30.

Walker, G. H., Stanton, N. A., Jenkins, D. P., Salmon, P. M., & Rafferty, L. (2010). From the 6 Ps of planning to the 4 Ds of digitization: Difficulties, dilemmas, and defective decision making. *International Journal of Human–Computer Interaction, 26*, 173–188.

Walker, J. L., & Cavenar, J. (1982). Vietnam veterans: Their problems continue. *Journal of Nervous Mental Disorder, 170*, 174–179.

Wallace, J. C., Popp, E., & Mondore, S. (2006). Safety climate as a mediator between foundation climates and occupational accidents: A group-level investigation. *Journal of Applied Psychology, 91*, 681–688.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Hoboken, NJ: Wiley.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–45.

Walther, J. B., Anderson, J. F., & Park, D. W. (1994). Interpersonal effects in computer-mediated interactions: A meta-analysis of social and antisocial communication. *Communication Research, 21*, 460–487.

Wandell, B. A. (1982). Measurement of small color differences. *Psychological Bulletin, 89*, 281–302.

Wang, J.-S., Knipling, R. R., & Goodman, N. J. (1996, October). *The role of driver inattention increases, new statistics from the 1995 crashworthiness data system*. Paper presented at the 40th American Association of Automobile Manufacturers conference, Vancouver, Canada.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.

Ward, L. M. (2004). Oscillations and synchrony in cognition. In V. K., Jirsa & J. A. S. Kelso (Eds.), *Coordination dynamics: Issues and trends* (pp. 217–242). New York: Springer.

Ward, L. M., & West, R. L. (1998). Modeling human chaotic behavior: Nonlinear forecasting analysis of logistic iteration. *Nonlinear Dynamics, Psychology, and Life Sciences, 2*, 261–282.

Ware, K. (2011, August). *Observing the emergence of and the control of chaos in the human nervous system*. Paper presented to the 21st Annual International Conference of the Society for Chaos Theory in Psychology & Life Sciences, Orange, CA.

Wattenberger, B. L. (1980). A look at the VDT issue. *Proceedings of the International Symposium on Human Factors in Telecommunications, 10*, 173–180.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics for "small-world" networks. *Nature, 393*, 440–442.

Wears, R. L. (2010). Exploring the dynamics of resilience. *Proceedings of the Human Factors and Ergonomics Society, 54*, 394–398.

Weber, E. H. (1978). *The sense of touch* (D. J. Murray, Trans.). New York: Academic Press. (Original work published 1846).

Weick, K. E. (2005). Managing the unexpected: Complexity as distributed sensemaking. In R. R. McDaniel, Jr., & D. J. Driebe (Eds.). *Uncertainty and surprise in complex systems* (pp. 51–65). New York: Springer-Verlag.

Weick, K. E., & Gilfillan, D. P. (1971). Fate of arbitrary traditions in a laboratory microculture. *Journal of Personality and Social Psychology, 17*, 179–191.

Weinland, J. P. (1927). Variability of performance in the curve of work. *Archives of Psychology, 30*.

Welford, A. T. (1980). Relationships between reaction time, fatigue, stress, age, and sex. In A. T. Welford (Ed.), *Reaction time* (pp. 321–354). New York: Academic Press.

Wellman, B. (1997). An electronic group is virtually a social network: In S. Keisler (Ed.), *Culture of the Internet* (pp. 179–205). Mahwah, NJ: Lawrence Erlbaum Associates.

Wellman, B., & Berkowitz, S. D. (Eds.). (1988). *Social structures: A network approach*. New York: Cambridge University Press.

Wenzel, E. M. (1994). Spatial sound and sonification. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 127–150). New York: Oxford University Press.

Wenzel, E. M., Wightman, F. L., & Foster, S. H. (1988). Development of a three-dimensional auditory display system. *SIGCHI Bulletin, 20*(2), 52–56.

Wesler, M. (2004). *Tactile displays* (Tech. Rep., Wright State University College of Engineering and Computer Science). Retrieved from www.cs.wright.edu/~jgalli/Tactile.ppt

West, B. J. (2006). *Where medicine went wrong: Rediscovering the path to complexity*. Singapore: World Scientific.

West, B. J., & Scafetta, N. (2003). Nonlinear model of human gait. *Physics Review E, 67*, doi:051917-1.

Whitaker, L. A. (1998). Human navigation. In C. Forsythe, E. Grose, & J. Ratner (Eds.), *Human factors and web development* (pp. 63–71). Mahwah, NJ: Lawrence Erlbaum Associates.

White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban and land-use development. *Environment and Planning B: Planning and Design, 24*, 323–343.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159–177.

Wickens, C. D. (2008a). Multiple resources and mental workload. *Human Factors, 50*, 449–455.

Wickens, C. D. (2008b). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors, 50*, 397–403.

Wiederhold, B. K., & Wiederhold, M. D. (2003). Three-year follow-up for virtual reality exposure for fear of flying. *Cyberpsychology and Behavior, 6*, 441–445.
Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23*, 995–1011.
Wiese, E. E., & Lee, J. D. (2007). Attention grounding: A new approach to in-vehicle information system implementation. *Theoretical Issues in Ergonomics Science, 8*, 255–276.
Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening: II. Psychophysical validation. *Journal of the Acoustical Society of America, 85*, 868–878.
Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. A., & Van Orden, G. C. (2009). 1/f scaling in movement time changes with practice in precision aiming. *Nonlinear Dynamics, Psychology, and Life Sciences, 13*, 79–98.
Wilde, G. J. S. (1988). Risk homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissension in recent reactions. *Ergonomics, 31*, 441–468.
Wilkins, I., & Dragos, B. (2013). Destructive destruction? An ecological study of high frequency trading. Retrieved from http://www.metamute.org/
Williams, A. M., Vickers, J., & Rodrigues, S. (2002). The effects of anxiety on visual search movement kinematics, and performance in table tennis: A test of Eysenck and Calvo's processing efficiency model. *Journal of Sport and Exercise Psychology, 24*, 428–455.
Williams, S. M. (1994). Perceptual principles in sound grouping. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 95–126). New York: Oxford University Press.
Wilpert, B. (2001, September). *Summary remarks to the congress*. Paper presented at the World Congress on Safety in Modern Technical Systems, Saarbrücken, Germany.
Wilson, G. T., & Davison, G. C. (1971). Processes of fear reduction in systematic desensitization. *Psychological Bulletin, 76*, 1–14.
Wisneski, K. J., & Johnson, M. J. (2007). Quantifying kinematics of purposeful movements to real, imagined, or absent functional objects: Implications for modeling trajectories for robot-mediated ADL Tasks. *Journal for Neuroengineering and Rehabilitation, 7,* 10.1186/1743-0003-4-7.
Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A Manual for the Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press.
Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7*, 275–278.
Wockutch, R. E., & VanSandt, C. V. (2000). OHS management in the United States and Japan: The DuPont and Toyota models. In K. Frick, P. Jensen, M. Quinlan, & T. Wilthagen (Eds.), *Systematic occupational health and safety management: Prospectives on an international development* (pp. 367–390). Oxford, UK: Pergamon Press.
Wojtezak-Jaroszawa, J. (1978). *Physiological and psychological aspects of night and shift work*. Washington, DC: U.S. Department of Health, Education, and Welfare, and National Institute for Occupational Safety and Health.
Wolfram, S. (Ed.). (1986). *Theory and applications of cellular automata*. Singapore: World Scientific.
Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.
Woodhill, D., Crutchfield, N., & James, F. S. (1987, September). Successful loss control programs require management's time, money. *Occupational Health and Safety, X*, *nv*, 61–65.
Woods, D. D., & Wreathall, J. (2008). Stress-strain plots as a basis for assessing system resilience. In E. Hollnagel, C. P. Nemeth, & S. W. A. Dekker (Eds.), *Resilience engineering: Remaining sensitive to the possibility of failure* (pp. 143–158). Aldershot, UK: Ashgate.
Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychological Review Monograph, 3*, 1–14.
Xiao, Y., Patey, R., & Mackenzie, C. F. (1995). *A study of team coordination and its breakdowns in a real-life stressful environment examined through video analysis* (Tech. Rep.). Baltimore: University of Maryland, School of Medicine, Department of Anesthesiology.

Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down: The cost of imprecision, unreliability, and visual clutter on cur effectiveness for display signaling. *Human Factors, 45*, 390–407.

Yerkes, R. M., & Dodson, J. D. (1908). The relationship of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science, 3*, 178–194.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

Zadeh, L. A. (1981). Possibility theory and soft data analysis. In L. Cobb & R. M. Thrall (Eds.), *Mathematical frontiers of the social and policy sciences* (pp. 69–129). Boulder, CO: Westview Press.

Zagare, F. C. (1984). *Game theory: Concepts and applications* (Quantitative Applications in the Social Sciences Paper Series, No. 41). Newbury Park, CA: Sage.

Zaidel, D. M., & Noy, Y. I. (1997). Automatic versus interactive vehicle navigation aids. In Y. I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces: Human factors in transportation* (pp. 287–307). Mahwah, NJ: Lawrence Erlbaum Associates.

Zander, A. (1994). *Making groups effective* (2nd ed.). San Francisco: Jossey-Bass.

Zeeman, E. C. (1977). *Catastrophe theory: Selected papers, 1972–1977*. Reading, MA: Addison-Wesley.

Zeggelink, E. P. H., Stokman, F. N., & van de Bunt, G. G. (1997). The emergence of groups in the evolution of friendship networks. In P. Doreian & F. B. Stokman (Eds.), Emotion of social networks (pp. 45–71). Amsterdam, the Netherlands: Gordon & Breach.

Zenger, B., & Fahle, M. (1997). Missed targets are more frequent than false alarms: A model for error rates in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1783–1791.

Zinn, C. (2001). Hospital infection control as part of an integrated safety concept. In J. Althoff (Ed.), *Safety of modern technical systems* (pp. 643–644). Saarbrücken, Germany: TÜV-Saarland Foundation.

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology, 65*, 96–102.

Zohar, D. (2003). Safety climate: Conceptual and measurement issues. In J. C. Quick and L. E. Tetrick (Eds.), *Handbook of occupational health psychology* (pp. 123–142). Washington, DC: American Psychological Association.

Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: Cross-level relationships between organization and group-level climate. *Journal of Applied Psychology, 90*, 616–628.

Zwerling, C., Daltroy, L. H., Fine, L. J., Johnston, J. J., Melius, J., & Silverstein, B. A. (1997). Design and conduct of occupational injury intervention studies: A review of evaluation strategies. *American Journal of Industrial Medicine, 32*, 164–179.

ERGONOMICS AND HUMAN FACTORS

Second Edition

HUMAN FACTORS ENGINEERING AND ERGONOMICS

A Systems Approach

Read the Reviews

"… covers a large number of theoretical, empirical and conceptual themes … The broad focus is supported by the introduction of novel approaches derived from complexity and dynamic systems theory to describing the behavior of human operators and sociotechnical systems thereby increasing the toolbox available for ergonomics and human factors researchers."

—Kjell Ivar Øvergård, Department of Maritime Technology and Innovation, Vestfold University College, Norway

" … a well-organized book, including the most up-to-date research results and … summarized a lot of comprehensive qualitative and quantitative research results."

—Dongyuan Debbie Wang, University of North Florida

"… a thorough and well-organized textbook. The focus and approach are relevant to current scholarship and research in the field."

—Robert W. Proctor, Purdue University

"An addition of a separate chapter on 'stress and work performance' is important, particularly given our current times, when so many people are experiencing stress at work, and unfortunately, are unaware of the problem. I also like the idea of a chapter on environmental design. We have always had a focus on microenvironments within the purview of human factors courses, but rarely do you see any mention of macroenvironments."

—Leah Newman, Pennsylvania State University

See What's New in the Second Edition

- Developments in working memory, degrees of freedom in cognitive processes, subjective workload, decision-making, and situation awareness
- Updated information on cognitive workload and fatigue
- Additional principles for HFE, networks, multiple person–machine systems, and human–robot swarms
- Accident analysis and prevention include resilience, new developments in safety climate, and an update to the inventory of accident prevention techniques and their relative effectiveness
- Problems in "big data" mining
- Psychomotor control and its relevance to human–robot systems
- Navigation in real-world environment
- Trust in automation and augmented cognition

K15876

ISBN: 978-1-4665-6009-3
90000
9 781466 560093

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

www.crcpress.com