McGRAW-HILL ENGINEERING AND TECHNOLOGY MANAGEMENT SERIES

2 DISKS Included

TECHNOLOGICAL FORECASTING FOR DECISION MAKING

THIRD EDITION

JOSEPH P. MARTINO

Cnam
Chaire de Prospective

Technological Forecasting for Decision Making

McGraw-Hill Engineering and Technology Management Series
Michael K. Badawy, Ph.D., Editor in Chief

ENGLERT • *Winning at Technological Innovation*

ETTLIE, STOLL • *Managing the Design-Manufacturing Process*

GAYNOR • *Achieving the Competitive Edge Through Integrated Technology Management*

RAHEJA • *Assurance Technologies: Principles and Practices*

RITZ • *Total Engineering Project Management*

STEELE • *Managing Technology*

WHEELER • *Computers and Engineering Management*

WHEELER • *Professional Workstations*

Technological Forecasting for Decision Making

Joseph P. Martino
Research Institute
University of Dayton
Dayton, Ohio

Third Edition

McGraw-Hill, Inc.
New York St. Louis San Francisco Auckland Bogotá
Caracas Lisbon London Madrid Mexico Milan
Montreal New Delhi Paris San Juan São Paulo
Singapore Sydney Tokyo Toronto

Library of Congress Cataloging-in-Publication Data

Martino, Joseph Paul.
Technological forecasting for decision making / Joseph P. Martino. —3rd ed.
p. cm.
Includes index.
ISBN 0-07-040777-0
1. Technological forecasting. 2. Decision-making. I. Title.
T174.M38 1993
658.4'0355—dc20 92-27833
CIP

The previous editions were published by American Elsevier Publishing Co. and Elsevier Science Publishing Co. in 1973 and 1983.

1 2 3 4 5 6 7 8 9 0 DOH/DOH 9 8 7 6 5 4 3 2

ISBN 0-07-040777-0

The sponsoring editor for this book was Robert W. Hauserman, the editing supervisor was Paul R. Sobel, and the production supervisor was Suzanne W. Babeuf. It was set in Century Schoolbook by McGraw-Hill's Professional Book Group composition unit.

Printed and bound by R. R. Donnelley & Sons Company.

In loving memory of Mary, whose support and encouragement meant so much for so long

Contents

Series Introduction

Technology is a key resource of profound importance for corporate profitability and growth. It also has enormous significance for the well-being of national economies as well as international competitiveness. Effective management of technology links engineering, science, and management disciplines to address the issues involved in the planning, development, and implementation of technological capabilities to shape and accomplish the strategic and operational objectives of an organization.

Management of technology involves the handling of technical activities in a broad spectrum of functional areas including basic research; applied research; development; design; construction, manufacturing, or operations; testing; maintenance; and technology transfer. In this sense, the concept of technology management is quite broad, since it covers not only R&D but also the management of product and process technologies. Viewed from that perspective, the management of technology is actually the practice of integrating technology strategy with business strategy in the company. This integration requires the deliberate coordination of the research, production, and service functions with the marketing, finance, and human resource functions of the firm.

That task calls for new managerial skills, techniques, styles, and ways of thinking. Providing executives, managers, and technical professionals with a systematic source of information to enable them to develop their knowledge and skills in managing technology is the challenge undertaken by this book series. The series will embody concise and practical treatments of specific topics within the broad area of engineering and technology management. The primary aim of the series is to provide a set of principles, concepts, tools, and techniques for those who wish to enhance their managerial skills and realize their potentials.

The series will provide readers with the information they must have and the skills they must acquire in order to sharpen their managerial performance and advance their careers. Authors contributing to the series are carefully selected for their expertise and experience. Although the series books will vary in subject matter as well as approach, one major feature will be common to all of them: a blend of practical applications and hands-on techniques supported by sound research and relevant theory.

The target audience for the series is quite broad. It includes engineers, scientists, and other technical professionals making the transition to management; entrepreneurs; technical managers and supervisors; upper-level executives; directors of engineering; people in R&D and other technology-related activities; corporate technical development managers and executives; continuing management education specialists; and students in engineering and technology management programs and related fields.

We hope that this series will become a primary source of information on the management of technology for practitioners, researchers, consultants, and students, and that it will help them become better managers and pursue the most rewarding professional careers.

DR. MICHAEL K. BADAWY
Professor of Management of Technology
The R. B. Pamplin College of Business
Virginia Polytechnic Institute and State University
Falls Church, Virginia

Preface

Since the second edition of this book was published there have been significant advances in the art of technological forecasting. These include refinements and improvements of older techniques, as well as some completely new techniques. In addition, the widespread availability of personal computers has changed the way in which technological forecasters work.

The changes in technological forecasting methods are reflected in every chapter. Improvements in methods have been incorporated into the descriptions. A completely new chapter, on probabilistic forecasting methods, has been added. The availability of personal computers is reflected by the inclusion with this edition of a set of computer programs to carry out some of the computations necessary to prepare forecasts. Also, the data tables have been updated where appropriate, and additional tables included.

This third edition, as with the second, benefited from considerable feedback from users, both individual readers and those who have taught from it as a text. These users are too numerous to mention individually, but their help is gratefully acknowledged.

Like the first two editions, this book remains a "how-to" book: how to do technological forecasting; how to apply it in specific decision situations; how to avoid some of the more common errors and difficulties in the preparation of individual forecasts. It is intended for use as a graduate or senior-level text, and as a guide to technological forecasting methods for practicing engineers, scientists, and technical administrators.

For the student of forecasting as well as the professional forecaster, this book presents a detailed discussion of all the important techniques currently in use. For each technique, the discussion includes the rationale behind it, a description of the methodology, and one or more historical examples. In addition, the proper circumstances for use of each technique are described, and the strengths and weaknesses of each technique are discussed. Each chapter concludes with problems intended to give the user an opportunity to practice the methods covered in the chapter.

For practicing scientists and engineers, the book provides information they will need to understand technological forecasts. There are three reasons why they will need this information. First, scientists and engineers will more and more be called upon to participate in preparing forecasts in their technical areas. They should understand the methods which will be used by forecasting professionals, with whom they will be asked to collaborate. Second, technological forecasting will increasingly be used in planning scientific and technological activity. It is advantageous to them, therefore, to understand the bases upon which decisions will be made about their work. Third, technological forecasting can be useful in expanding the application of their work. Regardless of which of these reasons is most important in a specific case, practicing scientists and engineers will find the information presented here to be valuable in both making and using forecasts.

Administrators of scientific and engineering activities, and other decision makers whose work will be influenced by technological change, will not usually be making detailed technological forecasts. They will have forecasts prepared for them by others. Nevertheless, it is important for these decision makers to be aware of the methods commonly used, to know when these methods are appropriate and when they are not, and particularly to know the strengths and weaknesses of the various methods. This book can provide them with an adequate background to make fully effective use of the forecasts they will receive.

The material in this book is divided into three parts. The first part, Chaps. 2 through 12, presents forecasting methods; the second part, Chaps. 13 through 17, describes applications of technological forecasting to various decision-making situations; the third part, Chaps. 18 through 20, provides instruction and guidance for the preparation and presentation of forecasts to be used by decision makers. The student of forecasting and the professional forecaster will find the first and third parts most useful; they will draw on the second part as the occasion demands. Other users will probably make most frequent use of the second part; they will consult the first and third parts for specific information about forecasting methods or for evaluating forecasts they have been given.

This book stresses the idea that technological forecasts are not made for their own sake but as inputs to decisions. The book emphasizes that technological forecasting is an important element in the decision-making process. However, the distinction between the role of the forecaster who *advises* a decision maker and that of the decision maker is repeatedly emphasized.

Joseph P. Martino

Chapter

1

Introduction

1.1 What Is Technological Forecasting?

This book is about technological forecasting: How to do it, how to apply it in specific situations, and how to prepare forecasts so that they can be useful from the decision-making standpoint.

Before discussing technological forecasting we must first define "technology." The *American Heritage Dictionary of the English Language* defines it as: "The entire body of methods and materials used to achieve [industrial or commercial] objectives." This definition is adequate for our purposes since both goods and services are included in the objectives to be achieved. Thus, technology here will mean the tools, techniques, and procedures used to accomplish some desired human purpose, i.e., technology is not restricted to hardware only, but may include "know-how" and "software."

The term "technology" is frequently used in the narrower sense of applied science or science-based knowledge. In actuality, however, most of the technology in the world is not science based; it is largely empirical in nature. Before the growth of science in the seventeenth century, virtually all technology was empirical rather than based on scientific theories.

The intent here is not to disparage science nor its use in the advancement of technology. Instead, the point is simply to recognize that much of technology is still based on practical experience and not on scientific theory. Even this purely empirical technology, however, is the province of the technological forecaster. Thus, we are concerned not solely with science-based technology but with all means used to provide the tools, techniques, and procedures necessary for human sustenance and comfort.

The *American Heritage Dictionary* defines "forecast" as "To estimate or calculate in advance...[To make] a conjecture concerning the future." Combining the ideas of technology and forecasting, then, we define a *technological forecast* as a prediction of the future characteristics of useful machines, procedures, or techniques.

There are three important points contained in this definition. First, a technological forecast deals with characteristics, such as levels of performance (e.g., speed, power, or temperature). It does not have to state how these characteristics will be achieved, i.e., the forecaster is not required to invent the technology being forecast. Even though the forecaster may predict characteristics that exceed the limitations of current technical approaches, the forecast need not state how these will be achieved. The forecaster's obligation is fulfilled by warning that these limitations will be surpassed.

Second, technological forecasting deals with useful machines, procedures, or techniques. In particular, this is intended to exclude from the domain of technological forecasting those items intended for luxury or amusement, which depend more on popular tastes than on technological capability. It does not seem possible to predict these rationally; however, the forecaster might be concerned with the means by which popular tastes will be formed, such as advertising or propaganda.

Having defined technological forecasting in a general sense, we now want to characterize it more precisely. A technological forecast (just as any other forecast) has four elements: the technology being forecast; the time of the forecast; a statement of the characteristics of the technology; and a statement of the probability associated with the forecast. We now examine these elements in more detail.

The *technology being forecast,* the first element, may be stated in either of two ways, depending upon the intent and nature of the forecast. Before specifying these ways, we need to define two terms.

The first term is "technical approach," which is a specific technical means of solving a problem or performing a particular function. For instance, piston engines and jet engines are two different technical approaches to the general function of powering aircraft; incandescent lamps, fluorescent lamps, and arc lights are three different technical approaches to the function of providing illumination. Sometimes a technical approach can be further subdivided. Jet engines can be divided into turbojets and turbofans. For some purposes the forecaster will consider jet engines as a single technical approach; in other cases he or she will consider turbojets and turbofans as alternative technical approaches to the function of powering aircraft.

The second term is "technology." Here we mean not the definition given above, but more narrowly, a family or series of technical approaches that have some major characteristic in common or that perform the same function. For instance, when the forecaster wishes to distinguish between turbojets and turbofans as technical approaches, the entire class of jet engines is then a technology to be distinguished from the technology of piston engines. When the forecaster wishes to distinguish between incandescent and fluorescent lamps, the entire class of electric lights is a technology to be distinguished from other technologies such as gas lamps. On the basis of context, this use of the word "technology" will be easily distinguished from the broader use defined above.

Armed with these definitions, we can be precise about the technology element of a technological forecast. The forecast must state whether it is for a single technical approach or for a more general technology. If it is for a technical approach, the forecast must be clear about how that approach is to be distinguished from other approaches in the same general technology. If the forecast is for a technology, it must be clear about how that technology is to be distinguished from others used for the same function.

The second element of the forecast is the *time* when it is to be realized. This may be a single point in time or a time span. In either case the time of the forecast should be stated clearly.

The third element of the forecast is the characteristics of the technology, given in terms of *functional capability.* Technology is intended to perform some function, and functional capability is a quantitative measure of its ability to carry out the function.

At this point, it is useful to distinguish between *technical* and *functional* parameters (more will be said about these in Chap. 6). A functional parameter is one that directly measures the extent to which the technology satisfies the user's needs. Put another way, a functional parameter enters into the user's utility function. For instance, a jet engine is intended to provide thrust and to do so economically. Thus functional parameters for a jet engine might include total thrust, specific fuel consumption, and thrust-to-weight ratio (total engine thrust divided by engine weight). These parameters are of direct interest to the user. However, the engine designer does not work directly with these parameters. The engine designer works with parameters such as turbine inlet temperature and compression ratio. By adjusting these technical parameters, the designer achieves the combination of functional parameters desired by the engine user.

A technological forecast may specify values of either technical or functional parameters as measures of functional capability. It is important to distinguish between the two, however, and doubly important not to mix functional with technical parameters. Mixing the two can give rise to double-counting of some technical parameters, if they are included explicitly and again implicitly as underlying a functional parameter.

In either case, functional capability is a quantitative measure of the technology's ability to carry out some function. The specific measures of functional capability chosen to characterize the given technology at the stated future time should be selected to satisfy the decision-making needs of the forecast user.

The fourth element of the forecast is the *probability* associated with it, which may be stated in several ways. The forecast may give the probability of achieving at all a given level of functional capability; it may state the probability of achieving a given level by a certain time; or it may state the probability distribution over the levels that might be achieved by a specific time. When the probability is not stated, it is assumed to be 100 percent.

The amount of information contained in a forecast is fixed by the data on which it is based. However, that information may be distributed among the four elements of the forecast in different ways. If the forecast is precise about the date, it must be less precise about the level of functional capability or the probability distribution. If the forecast is precise about the level of functional capability, it may have to be less precise about the date or the probability. Increased precision in one element of the forecast always means less precision in the other elements, given the fixed amount of information available to the forecaster.

How should the information be distributed among the elements of the forecast? Which should be given higher precision and which lower precision? This depends upon how we use the forecast. In all cases the forecast should be tailored to the needs of the user, providing greater precision where required, at the cost of lesser precision in other elements of the forecast.

1.2 Why Forecast Technology?

The question, "Why forecast technology?" has a false implication. It implies that there is a choice between forecasting and not forecasting. But forecasting technology is no more avoidable than is forecasting the weather. All people implicitly forecast the weather by their choice of whether to wear a raincoat, carry an umbrella, and so on. Any individual, organization, or nation that can be affected by technological change inevitably engages in forecasting technology with every decision that allocates resources to particular purposes. A change in technology may completely invalidate a particular decision about allocating resources. Every decision, then, carries within itself the forecast that technology either will not change at all or will change in such a way as to make the decision a good one.

Given that technological forecasting is inevitable, however, there is still the issue of what specific reasons people have for making technological forecasts. In actuality, people make technological forecasts for the same reasons they make other forecasts:

1. To maximize gain from events external to the organization
2. To maximize gain from events that are the result of actions taken by the organization
3. To minimize loss associated with uncontrollable events external to the organization
4. To offset the actions of competitive or hostile organizations
5. To forecast demand for purposes of production and/or inventory control
6. To forecast demand for facilities and for capital planning
7. To forecast demand to assure adequate staffing
8. To develop administrative plans and policy internal to an organization (e.g., personnel or budget)
9. To develop policies that apply to people who are not part of the organization

Most of the items on this list boil down to the idea of maximizing gain or minimizing loss from future conditions. Each item could be a reason for technological forecasting as well as for economic, business, political, or weather forecasting.

Throughout this book, the emphasis will be on forecasting for decision making, i.e., the entire justification for producing forecasts will be their use in making decisions. The implication of this is that using forecasts helps make better decisions. In particular, the forecasts play specific roles in improving the quality of decision making. Ralph Lenz, a pioneer in technological forecasting, has identified the following specific roles that technological forecasts play in improving the quality of decisions:

1. It identifies limits beyond which it is not possible to go.
2. It establishes feasible rates of progress, so that the plan can be made to take full advantage of such rates; the plan does not demand an impossible rate of progress.
3. It describes the alternatives that can be chosen.
4. It indicates possibilities that might be achieved if desired.
5. It provides a reference standard for the plan. The plan can thus be compared with the forecast at any later time to determine whether it can still be fulfilled or whether, because of changes in the forecast, the plan must be revised.
6. It furnishes warning signals, which can alert the decision maker that it will not be possible to continue the present activities.

In playing these roles, the forecast provides specific pieces of information needed by the decision maker. We have so far tacitly assumed that the improved quality of the decision more than offsets the cost of the forecast. However, this is not at all certain. Sometimes forecasts cost more than they are worth. This is a topic we will take up in Chap. 19. For the moment, we simply note that forecasts do provide specific information that can improve the quality of decisions.

1.3 Alternatives to Forecasting

The definition of a forecast, as estimating or calculating in advance, implies some degree of rationality and analysis of data. However, there are many "alternatives" to rational and analytic forecasting. Most of these alternatives are widely used for the same purposes as forecasts, and, therefore, it is worthwhile to examine them briefly.

No forecast. This alternative means facing the future blindfolded. If taken literally, it means that no attempt is made to determine what the future will be like and that decisions are made without any regard whatsoever to their future consequences, be they favorable or unfavorable. It should be clear that any organization operating on this basis will not survive. Even if the environment is unchanging, most decisions will be wrong, since they will not even take into account a forecast of the constancy of the environment. If the environment is changing rapidly, disaster may come even more quickly, since a decision that is right for a short time may be rendered inappropriate in the longer run. In most cases, however, the concept of "no forecast" is not meant literally but is really intended to mean a forecast of constancy or negligible rate of change. Thus, when a decision maker claims not to believe in or use technological forecasts, what the decision maker really means is that he or she has assumed an unchanging technology. The decisions are made on the basis of a forecast that the technology in existence at the time during which these decisions have their impact will be the same as the technology of today. Thus, this is really not an alternative to forecasting but a very specific, though implicit, forecast.

Anything can happen. This represents the attitude that the future is a complete gamble, that nothing can be done to influence it in a desired direction, and that there is no point, therefore, in attempting to anticipate it. It is doubtful if there is any decision maker who runs his or her personal life this way. Even decision makers who claim to have this attitude are still likely to take a raincoat on cloudy days. Therefore, if they pretend to adopt this attitude toward their professional decisions, it really amounts to a cover for something else—perhaps an attempt to avoid the effort of thinking through the implications of a forecast. Obviously, decision makers who really act on the basis of this attitude are headed for trouble. In particular they may find their organizations unable to withstand the competition from other organizations that do attempt to anticipate the future through rational means. An organization run on this basis can only be short-lived.

The glorious past. This represents an attitude that looks to the past and ignores the future. Many organizations can point to significant achievements at some time or the other in the past. Their very survival over an extended period indicates that they have done the right things. Unfortunately, when conditions change, it is very unlikely that the policies and decisions that led to success in the past will continue to be suitable. Stubbornly clinging to visions of the glorious past, under the assumption that the glorious past guarantees a glorious future, is a certain road to disaster. In short, an organization that concentrates on its past instead of the future can end up only in becoming a museum piece.

Window-blind forecasting. This involves the attitude that technology moves on a fixed track, like an old-fashioned roller window blind, and that the only direction is up. This attitude is encapsulated in expressions such as "higher, faster, and farther," or "bigger and better"; it assumes that the future will be like the past, only more so. While this attitude does at least recognize that changes do take place and is, therefore, somewhat better than the preceding alternatives, it fails to recognize that there are other directions besides up. A particular technical approach, for instance, may come to a halt or move sideways if another technical approach supersedes it. An organization that depends on window-blind forecasting will sooner or later be taken by surprise as some unanticipated technological change brings an end to the track the organization was following.

Crisis action. This can best be described as "pushing the panic button." It consists in waiting until the problem or crisis has arrived and then taking some immediate action to attempt to alleviate the impact of the crisis. Over the long term, crisis action means that the organization is not making any net progress toward its goals. As a result of expedient responses to crises, it may only be zigzagging instead of proceeding directly toward an objective. Furthermore, this alternative is based on the assumption that there will be time to respond effectively after a crisis has arrived. If this assumption proves false in a specific crisis, the organization fails to survive. Finally, this alternative ignores the fact that had a proper forecast been used, the crisis might have been avoided completely. Hence, although this approach is sometimes used, this is not really an acceptable alternative to proper forecasting.

Genius forecasting. This is not really an alternative to forecasting, since it does involve the preparation of a forecast. However, it is an alternative to the use of rational and explicit methods for obtaining forecasts. This method consists in finding a "genius" and asking him or her for an intuitive forecast. It must be recognized that many "genius forecasts" made in the past have been successful. Unfortunately, there have also been many such forecasts so wide of the mark as to be useless. Ralph Lenz has described the shortcomings of genius forecasting as follows: It is impossible to teach, expensive to learn, and allows no opportunity for review by others. Obviously, there is no surefire way of obtaining a genius, and even if a genius has been located, his or her forecast cannot be checked by anyone else, even by another genius. It must be taken on faith. It may be that in some cases there is no alternative to a genius forecast; however, it should be clear that where rational and explicit methods are available, they are much to be preferred. In fact, rational and explicit methods relieve the decision maker of the burden of ferreting out geniuses.

The whole purpose of this recitation of alternatives, of course, is to show that there really is no alternative to forecasting. If a decision maker has several options available, he or she will choose the one that provides the most desirable outcome. Thus, every decision is inevitably based on a forecast. Hence, the decision maker does not have a choice about whether to make a forecast: There is no such thing as not forecasting. The only choice is whether the forecast is obtained by rational and explicit methods, or by intuitive means from the depths of someone's subconscious.

The virtues of using rational methods are that they are teachable and learnable; they can be described and explained; and they provide a procedure that can be followed by anyone who has been able to absorb the necessary training. In some cases, in fact, the methods are even guaranteed to provide the same forecast, regardless of who uses them.

Another virtue of the use of explicit methods is that they can be reviewed by others. In particular, the forecast can be reviewed by several people before being accepted by the decision maker. It can be checked to see if any mistakes have been made in the application of the method, in the calculations, or in the data itself. Furthermore, the forecast can be reviewed at any subsequent time to see if it is still acceptable. If conditions have changed enough to invalidate the forecast, the plans that were based on it can be altered accordingly. If the forecast is not rational and explicit, it cannot be subjected to subsequent review to assure that it is still acceptable. This creates a risk that the plans based in it will be left unchanged although they are no longer appropriate.

Even though decision makers may agree that they have no alternative to forecasting and that they recognize the virtues of rational and explicit methods, decision makers may confront the technological forecaster with yet another issue.

1.4 Will It Come True?

It might appear that a good forecast is one that comes true. After all, what good is a forecast that is wrong? A person seeking a weather forecast, for instance, wants to know what the weather will be in order to prepare for it. A forecast that turns out to be wrong is totally useless. The user may have made unnecessary preparations for bad weather when the weather turns out to be good or may have failed to make routine preparations for bad weather because the forecast called for good weather. Clearly, a weather forecast has to be correct if it is to be useful.

Many people apply the same criterion to technological forecasts (as well as to economic and political forecasts). However, there are two things wrong with this criterion. The first is that it cannot be applied before the fact. The second is that it fails to take into account self-altering forecasts.

A self-altering forecast is one that, by virtue of having been made, alters the outcome of the situation. Suppose someone forecasts an undesirable situation. Then suppose a decision maker accepts the forecast and acts to prevent the undesirable situation. Clearly the forecast did not come true. Was it then a bad forecast? On the contrary, it was highly useful. Because of the forecast, someone obtained a better outcome. The same argument can be made for a forecast of a favorable outcome that comes true only because the forecast is accepted and acted upon. Clearly, the forecast was useful, because it led someone to take an action that made things better. However, the value of the forecast was in its usefulness, not in its coming true.

It is important that the forecaster not get trapped into evaluating forecasts by whether they came true. It is even more important that forecasters educate forecast users to the idea that the goodness of a forecast lies in its utility for making better decisions and not in whether it eventually comes true.

1.5 Stages of Innovation

Next, we introduce a concept that will be extremely useful in our later discussions of technological change, i.e., "stages of innovation." No technological device springs directly from the fertile mind of an inventor to immediate widespread use. It passes through a number of stages along the way, with successive stages representing greater degrees of practicality or usefulness.

Various writers have given lists of the stages through which an innovation passes. The following list is an expanded version of one originated by James R. Bright.

1. Scientific findings
2. Laboratory feasibility
3. Operating prototype
4. Commercial introduction or operational use
5. Widespread adoption
6. Diffusion to other areas
7. Social and economic impact

Each of these stages is discussed below.

Scientific findings. At this stage the innovation exists in the form of the scientific understanding of some phenomenon, the properties of some material, the behavior of some force or substance, and so on. It is not in any sense capable of being utilized to solve a problem or carry out a function. Scientific findings, in general, merely represent a knowledge base from which solutions to specific problems can be drawn.

Laboratory feasibility. At this stage a specific solution to a problem has been identified and a laboratory model has been put together. It is clear that no natural or physical laws are violated by the device and that it is capable of performing the desired function or solving the problem of concern, but only under laboratory conditions. The innovation at this stage may be described by terms such as "breadboard model" or "brassboard model." It certainly would not function well outside the laboratory nor without the constant attention of a skilled technician.

Operating prototype. At this stage a device has been built that is intended to function satisfactorily in the operational environment. That is, the device is expected to be rugged enough, reliable enough, and easy enough to operate and maintain that it can perform its intended function in the hands of a typical user. It should be noted that frequently the purpose of building operating prototypes is to verify that the design will function as intended in the intended situation or environment. An operating prototype of a consumer appliance might be use-tested by a panel of "typical" buyers. An operating prototype of a military device might be tested in maneuvers with "typical" troops, and so on. At this stage, the innovation is called upon to demonstrate the adequacy of its design.

Commercial introduction or operational use. This stage represents not only technical and design adequacy, but economic feasibility. The innovation is supposed to demonstrate that people will want it badly enough to give up other things in order to obtain it. The "first production model" is often chosen as representing the point in time at which an innovation has reached this stage.

Widespread adoption. By this stage the innovation has demonstrated that it is technically and economically superior to whatever else was used in the past to carry out the same function. It bids fair to replace those prior devices, techniques, or procedures on a wide scale. There is no precise definition of when this stage is reached. It might be expressed in terms of a percentage of the total potential use, the gross national product, the sales of all the items performing the same function, the items in use for the same purpose, and so forth. While the forecaster has some freedom to define precisely what is meant by "widespread adoption" in a particular case, once this is done, the definition must be followed consistently.

Diffusion to other areas. This is the stage when the innovation has not only dominated the application area in which it was adopted but has been adopted in other application areas as well. If the innovation supplanted some earlier device or technique, at this stage the innovation has been adopted for purposes to which the earlier device was never applied. The transistor, for instance, has reached this stage. Not only has it virtually replaced the vacuum tube in much of conventional electronics, but it now appears in applications for which vacuum tubes were never used, such as automobile ignitions and automatic exposure control in cameras.

Social and economic impact. When an innovation reaches this stage, it has in some way changed the behavior of society or has reached the point where a significant portion of the economy is somehow involved with it. Television, for instance, has certainly had a major impact on the U.S. society. The automobile has not only had a major impact on U.S. society, but it has reached the point where more than 10 percent of the U.S. gross national product is involved in its manufacture, sale, operation, and maintenance.

Not every innovation goes through all these stages. Some reach a certain stage and go no farther. Others may cover the entire list but combine two or more stages. Some, especially innovations that are based on empiricism rather than theory, may skip the early stages. Nevertheless, this list is useful for avoiding confusion about the degree of development, or extent of use, of some device or technique. The concept of stages of innovation will be useful in subsequent chapters.

1.6 Methods of Forecasting

It would be a mistake to think that the methods used by technological forecasters are unique to that application area and are somehow different from the methods used in other application areas. In reality, there is a great deal of similarity among the forecasting methods used in all application areas. There are four basic methods of forecasting that are used by all forecasters. Regardless of the area of application, all forecasters use variations and combinations of the following four basic methods:

1. *Extrapolation:* In this method, the forecaster extends or projects a pattern that has been found in the past. The forecaster starts with a time series of past data about the entity to be forecast. For instance, a technological forecaster who was attempting to forecast aircraft speed would obtain a time series of aircraft speed records. The forecaster then attempts to find a pattern in this historical data. The typical kinds of patterns found may be trends (technological and economic data) or cycles (weather data). Once a pattern is found, it is extended to the future to obtain the forecast. The basic assumption behind forecasting by extrapolation is that the past of a time series contains all the information needed to forecast the future of that same series.

2. *Leading indicators:* In this method, the forecaster uses one time series to obtain information about the future behavior of another time series. For instance, a falling barometer often precedes rain. A weather forecaster, thus, uses a "turning point" in the time series of barometric pressure to forecast a future turning point in the amount of precipitation. The basic assumption behind forecasting by leading indicators is that the time series of interest shows the same behavior as another time series, the leading indicator, but with a known time lag. Thus, what the leading indicator is doing today will be matched by the time series of interest at a specific time in the future.

3. *Causal models:* Both extrapolation and leading indicators require only a finding of correlation between past and future. It is not necessary to know anything about the causal factors that make the future follow the pattern revealed by the past. For instance, the forecaster using a leading indicator does not need to know *why* the leader and follower behave in that fashion. Causal models, as implied by the name, incorporate information about cause and effect. A forecast of a solar eclipse, for instance, is based on a causal relationship, involving some fundamental laws of physics. The assumption behind use of causal models for forecasting is that the cause-effect linkages in the topic of interest are known and can be expressed mathematically or in some similar fashion (e.g., a mechanical model).

4. *Probabilistic methods:* In the preceding three methods of forecasting, a given set of data (past time series, leading indicator series, and equations of the model) produce a point estimate of future conditions. Given past and present conditions, there is exactly one possible value for the future of the parameter to be forecast. Probabilistic methods differ from the preceding three in that, instead of producing a single-valued forecast, they produce a probability distribution over a range of possible values. Some weather forecasts are now given in this fashion. The probability of rain tomorrow may be stated as, for instance, 30 percent. This means that over the range of possible outcomes, rain and no-rain, the associated probabilities are 30 and 70 percent, respectively.

In reality, these four methods are often used in combination. The methods presented later in this book will include both "pure" methods and combinations. The point that must be remembered, however, is that there are only four basic methods of forecasting. The differences found in different application areas arise primarily from differences in the type of data available, uses to which the forecasts will be put, and so on; the outcome is that the basic methods are used differently or are combined differently. For instance, economic data such as prices come at regular intervals (weekly, monthly, etc.). Technological data do not come at regular intervals. Hence, although both economists and technological forecasters use extrapolation, the specific techniques they use to extrapolate their data differ because of this fundamental difference in the nature of the data.

1.7 Computer Programs

Much of the activity of forecasting, such as computing the extrapolation of a pattern, is purely "mechanical" in nature. It is a matter of carrying out a specified procedure on the data that pertain to the topic of interest. Much of this mechanical activity can be carried out by a computer. Included with this book is a disk containing two programs, TEKFOR and MAXIM. These are described in Apps. C and D, respectively. Their use is presented in the various chapters that deal with the techniques embodied in these programs. Inclusion of these computer programs is intended to relieve the forecaster of some of the mechanical aspects of forecasting and allow concentration on the intellectual and judgmental aspects.

1.8 Remainder of the Book

The emphasis of this entire book is on the *use* of technological forecasts for making decisions. The forecast is viewed solely as an input to the decision-making process. The book is divided into three major sections, each devoted to a particular aspect of the preparation and use of technological forecasts.

The first section, consisting of Chaps. 2 through 12, is devoted to *methods* of technological forecasting. Various methods are described and illustrated by example. The strengths and weaknesses of each method are described.

The second section, consisting of Chaps. 13 through 17, is devoted to *applications* of technological forecasting. Planning and decision making are discussed in general at first and then in detail for each of several major application areas. These applications are illustrated by practical and historical examples.

The third section, consisting of Chaps. 18 through 20, is devoted to *presentation* of the forecast, to show the forecaster how to present the results effectively so they will be useful for making decisions. The various chapters discuss good practice, bad practice, and pitfalls.

References

Bright, James R., "The Manager and Technological Forecasting," in James R. Bright (ed.), *Techological Forecasting for Industry and Government*, Prentice-Hall, Englewood Cliffs, NJ, 1968.

Problems

1 Which of the following items or activities represent or exemplify technology as it is defined in the text?

Farming	A legislature
Money	Fixed-wing aircraft
Cattle raising	The limited-liability corporation
Roads	Smelting metals from ores
Bridges	The violin

2 Which of the items in Prob. 1 was based from its inception on a scientific understanding of the principles underlying its operation or activity?

3 Trace the history of radio communication, identifying the dates at which it passed through the stages of innovation given in the text.

4 Consider a technology for which there is a steady demand and that receives broad-based support from a wide variety of sources, each seeking competitive advantage by achieving an advance in the level of functional capability. Why would one of the groups involved in advancing this technology need a forecast of its probable future progress?

5 Consider a technology for which the demand fluctuates about some norm or regular pattern and that is advanced by a small number of corporations or other groups, each of which is interested in seeing that its products or capabilities do not fall behind those of the others. Why would one of these groups need a forecast of the probable future course of this technology?

6 Consider a technology that is utilized by one group, whose advance is supported by that group alone and whose rate of advance is geared to the needs of the group (an example would be instrumentation on a military missile-testing range). Why would this group need a forecast of this monopolistically controlled technology?

Chapter

2

Delphi

"Two heads are better than one." JOHN HEYWOOD
Proverbes Part I, Chapter IX
(1546)

2.1 Introduction

Formal forecasting methods are intended to replace subjective opinion with objective data and replicable methods. However, there are three conditions under which expert opinion will always be needed. (Note that the choice of an "objective" method may itself require expert judgment and involve implicit assumptions. The importance of these assumptions will be discussed in Chap. 18.)

The first condition is when no historical data exist. In technological forecasting, this usually involves new technologies. Despite the lack of historical data, a forecast may be required. Expert opinion then is the only possible source of a forecast.

The second condition is when the impact of external factors is more important than the factors that governed the previous development of the technology. These external factors may include decisions of sponsors and opponents of the technology, and changes in public opinion. In such a case, data about the past may be irrelevant. Expert opinion about the effects of these external factors may be the only possible source of a forecast. Recent historical examples of technologies in which external factors were dominant are space technology (President Kennedy's decision to make the moon a goal) and nuclear power (opposed on safety grounds).

The third condition is when ethical or moral considerations may dominate the economic and technical considerations that usually govern the development of technology. These issues are inherently subjective, and expert opinion may be the only possible source of a forecast. An example, current as of this writing, is research on fetal tissue obtained via abortions. The technology is straightforward. The use of that technology depends on the outcome of a debate on the ethics of such use. Forecasting the future of the technology can be done only by using expert opinion about the outcome of the debate.

Given that expert opinion is needed in a specific case, how is it to be obtained? The problems of "genius forecasting" were discussed in Chap. 1. Can they be overcome? One way to overcome them is use of several experts—two heads are better than one. In a group of experts, individual biases may be canceled out, and the knowledge of one member may compensate for another's lack of it.

Nevertheless, we must recall the adage, "a camel is a horse designed by a committee." A forecast designed by a committee may be equally grotesque. What is needed is some way to obtain the benefits of a committee while minimizing the disadvantages.

2.2 Advantages of Committees

The first major advantage of a committee is that the sum of the information available to a group is at least as great as that available to any individual member. Adding members to a group does not destroy information. Even if one member knows more than the rest put together, the other members do not reduce the total information available to the group. They may still make useful contributions. If the group has been chosen to contain only people who are experts in the subject, the total information available to the group is probably many times that possessed by any single member.

The second major advantage is that the number of factors that can be considered by a group is at least as great as the number that can be considered by a single member. This point is at least as important as the first. Studies of forecasts that have gone wrong show that one very common cause of failure is the failure to take into account important factors outside the technology being forecast, which in the long run turned out to be more important than those internal to the technology. This advantage of a group is therefore very important.

These two advantages are the usual reasons for using a committee instead of a single expert. The widespread use of committees implies that these advantages are significant.

2.3 Disadvantages of Committees

The first major disadvantage of a group is that there is at least as much misinformation available to the group as there is to any single member. One reason for using a group is the hope that the misinformation held by one member may be canceled out by valid information held by another. However, there is no guarantee this will take place.

The second major disadvantage is the social pressure a group places on its members—pressure to agree with the majority even when the individual believes that the majority is wrong. This is especially true in the production of group forecasts. One member may well give up presenting certain relevant factors if the remainder of the group persists in taking a contrary view.

The third major disadvantage is that a group often takes on a life of its own. Reaching agreement becomes a goal in itself, of greater importance than producing a well thought-out and useful forecast. Group forecasts may, thus, be only a watered-down least common denominator that offends no one, even though no one agrees strongly either.

A fourth major disadvantage is the influence that the repetition of arguments can have. Experiments with small groups show that often it is not the validity of a position but the number of comments for or against it that carries the day. A strong vocal minority may overwhelm the majority by pushing its views vigorously, even though their arguments may have little objective merit.

A fifth major disadvantage of groups is their vulnerability to the influence of dominant individuals. One individual, by active participation in debate, by putting ideas forth with a great deal of vigor, or through a persuasive personality, may have an undue influence on the group's deliberations. Such an individual may get his or her way simply by wearing down the opposition with persistent arguments.

A sixth disadvantage of groups is that the entire group may share a common bias. This often arises from a common culture shared by the members—especially a subculture peculiar to the technology in which the members are experts. The presence of a common bias nullifies the advantage of a group in canceling biases.

These disadvantages of committees offset to some extent the advantages listed in the previous section. In many cases these disadvantages are not sufficient to justify abandoning the advantages of committees. Nevertheless, it would be better if the advantages could be obtained without the disadvantages.

2.4 The Delphi Procedure

Delphi is intended to gain the advantages of groups while overcoming their disadvantages. It was originally developed at the Rand Corporation as a means of extracting opinion from a group of experts. Its first public presentation in a Rand report dealt with a series of technological forecasts, which led to the misunderstanding that Delphi was primarily a forecasting method: It is not. It can be used for any purpose for which a committee can be used. While the emphasis in this book is on technological forecasting, other uses of Delphi are discussed in Linstone and Turoff (1975).

Delphi has three characteristics that distinguish it from conventional *face-to-face* (FTF) group interaction. These are anonymity, iteration with controlled feedback, and statistical response.

Anonymity. During a Delphi sequence the group members do not know which members of the group contributed particular statements or opinions. They may not even know who else is in the group. The interaction among members takes place in a completely anonymous manner through the use of written questionnaires. This anonymity has several beneficial effects. First, it avoids the possibility that a panel member may be influenced by the (good or bad) reputation of a particular member who presents a fact or argument. Instead, the members are influenced only by the cogency of the arguments or the perceived validity of the statements. Second, it means that members may change their opinions, in the face of contrary arguments or evidence, without anyone else knowing they have done so. It thus eliminates the problem of reluctance to change a publicly stated position for fear of "losing face."

Iteration with controlled feedback. The group interaction is carried out through responses to questionnaires. The panel moderator extracts from the questionnaires only those pieces of information that are relevant to the issue and presents these to the group. The individual panel members are then informed only of the current state of the collective opinion of the panel and the arguments for and against each point of view. The panel members are not subjected to a harangue or to endless restatements of the same arguments. Both the majority and minority can have their views presented to the panel, but not in such a way as to overwhelm the opposition simply by weight of repetition. The primary effect of this controlled feedback is to prevent the group from taking on its own goals and objectives. The controlled feedback permits the group to concentrate on its original objectives, without being distracted by self-chosen goals such as winning an argument or reaching agreement for the sake of agreement.

Statistical group response. Typically, a FTF group will produce a forecast that contains only a majority opinion. This will contain only those views on which a majority of the group could agree. At most, there may be a minority report. There is unlikely to be any indication of the difference of opinion that existed within the group. Delphi, by contrast, presents instead a statistical response that includes the opinions of the entire group. On a single item, for instance, group responses are presented in statistics that describe both the "center" of the group opinion and the degree of spread about that center.

These three characteristics of Delphi distinguish it from more conventional uses of groups, such as FTF committee meetings. They allow Delphi to retain the advantages of groups while reducing or eliminating their disadvantages. However, Delphi retains an important characteristic of groups, namely interaction among the participants. Delphi is not like a public opinion poll, wherein the objective is to measure opinion without affecting it. Delphi, like a more conventional committee, is intended to allow change of opinion among the participants through interaction. The clash of argument, the presenting of alternative opinions, the marshaling of facts, are just as possible in Delphi as in a FTF group. In fact, the structure of Delphi is intended to foster these interactions rather than inhibit them.

There is another aspect in which Delphi is like a committee rather than like a public opinion poll. Delphi is intended to obtain the opinions of the "expert community" on some subject. The participants in a Delphi are, therefore, selected to be representative of the expert community rather than of the public at large. Instead of containing a random statistical sample of experts, a Delphi may be designed to include extreme opinions, specific schools of thought, and minority views as well as majority views. A Delphi panel, then, may be designed to overrepresent these minority positions. More will be said in Sec. 2.8 about selection of panel members. The point here is simply that the selection methods appropriate for a public opinion poll are totally inappropriate for a Delphi. Instead, a Delphi panel should be selected in the same manner as a committee of experts would be selected.

2.5 Conducting a Delphi Sequence

The "classical" Delphi, as originated at Rand, is carried out as described below. Departures from this classic form will be discussed later.

Before describing the mechanics of Delphi, however, some definitions are required. A Delphi sequence is carried out by interrogating a group of experts with a series of questionnaires. Each successive questionnaire is a "round." However, the term "questionnaire" may be misleading. The questionnaires not only ask questions, but provide information to the group members about the degree of group consensus and the arguments presented by the group members for and against various positions. The questionnaire is the medium for group interaction. The set of experts taking part in the Delphi is usually referred to as a "panel." The person responsible for collecting the panel responses and preparing the questionnaires is called the "moderator."

Delphi will be described in terms of rounds. Each round calls for somewhat different activities on the part of the panelists and moderator. Before the first round there must be preliminary activities such as clarifying the subject and explaining the methods. After these preliminaries, the first round can begin.

Round one. The first questionnaire is completely unstructured. The panelists are asked to forecast events or trends in the area for which the panel was assembled. This has some disadvantages, which will be discussed later, but it also has some significant advantages. The panelists have been selected because of their expertise in the area to be forecast. They should know much more than the moderator does about that area. If the first questionnaire were too structured, it might prevent the panelists from forecasting some important events of which the moderator might not be aware.

The questionnaires are returned to the moderator, who then consolidates the forecasts into a single set. Similar items must be combined; items of lesser importance must be dropped to keep the list at a reasonable length; and each event must be stated as clearly as possible. The list of events then becomes the questionnaire for the second round.

Round two. The panelists receive the consolidated list of events. They are asked to estimate the time of occurrence for each event. The estimate may be a date; it may be "never" if they think that the event will never occur; or it may be "later" if some time horizon has been specified for forecasts and they believe that it will occur later than that horizon.

The moderator collects the forecasts from the panel and prepares a statistical summary of the forecasts for each event. This usually consists of the median date and the upper and lower quartile dates for each event. The median date is a date such that half the forecasts are earlier and half later. It, thus, represents the "middle" of the panel responses. The quartiles divide the upper and lower halves of the responses in the same way that the median divides the entire set of responses. One-fourth of the responses are above the upper quartile, and three-quarters are below it. The converse is true for the lower quartile. The quartiles are thus a measure of the spread of panel responses about the median. This spread can be quantified as the "interquartile range" (upper quartile minus the lower quartile). The third questionnaire consists of the set of events and the statistical summary of the forecasts (median and quartiles for each event).

Round three. The panelists receive the questionnaire with events, medians, and quartiles. They are asked to prepare new forecasts for each event, either sticking with their previous forecast or making a new one. If their third-round forecast falls in either the upper or lower quartile (that is, if they are "outliers"), they must present reasons why they believe they are correct and three-quarters of the panel are incorrect. Their reasons may include reference to specific factors the other panelists may be overlooking, facts the other panelists may not be considering, and so on. The panelists are as free to advance arguments and objections as they would be in a FTF group. The only difference is that their arguments are written and anonymous.

When the third-round responses are received, the moderator prepares a statistical summary of the forecasts, as well as a consolidated summary of the panel's reasons for advancing or delaying the forecasts. Similar arguments are combined, and lengthy arguments are summarized. (Fortunately the need to write the arguments often forces the panelists to be concise.) The questionnaire for the fourth round consists of the list of events, the medians and quartiles for the third round, and the summary of arguments for changing the forecasts for each event.

Round four. The panelists receive the events and dates, and the reasons given for changing their estimates. They are asked to take the reasons into account and make new forecasts for each event. Depending upon the needs of the moderator, they may be asked to justify their position if their forecasts fall in the upper or lower quartile. In addition, the moderator may invite comments from the panelists on the arguments given during the third round. If the moderator asks for comments, then any panelist may respond, regardless of whether their responses fall "in the middle" or in one of the extreme quartiles.

Upon receiving the forecasts from the panelists, the moderator again computes medians and quartiles. If comments were requested, the moderator consolidates and summarizes them. (If the moderator does not plan to analyze the arguments, there is no point in asking for them.) In some cases, when the panel has not been able to reach a consensus, the moderator may well be interested in the arguments for both sides. In such cases, the moderator should ask for comments and be prepared to analyze them.

The final results of the Delphi sequence are forecasts (fourth-round median dates for each event), measures of panel disagreement (fourth-round interquartile ranges for each event), and summaries of critical issues related to each event (the arguments for advancing or delaying the forecasts). This is much more information than is usually available from a conventional FTF committee.

One of the advantages of Delphi is that the statistical response includes the views of "outliers" as well as "centrists." Moreover, the moderator can determine the degree of disagreement within the panel by the upper and lower quartiles. A narrow interquartile range implies considerable agreement; a wide one implies considerable disagreement. The ratio of interquartile range to length of forecast (span from present date to median forecast date) is a useful measure of panel disagreement. It has been found that for a given panel this ratio tends to be relatively constant for forecasts of different lengths. This means that the relative uncertainty of the panel is fairly constant, but the absolute uncertainty grows with length of the forecast. If for a particular event the ratio of interquartile range to length of forecast is much larger (smaller) than the average over all forecasts, this means the degree of panel agreement on that event is much less (more) than for the remainder of the events.

The comments on each event provide a summary of those factors that the panelists believe are important and that may affect the forecast. Moreover, the structure of Delphi focuses this information on the topics of interest to the moderator and organizes it in a readily understandable manner.

Ordinary committees are judged as successes if they reach agreement or consensus. Indeed, committee action is designed to achieve consensus and may actually force a false consensus. Delphi is intended to display disagreement where it exists and to search for the causes of disagreement. Delphi sequences are judged as successful when they reach stability, that is, when there is no further change from round to round, with the reasons for divergence clearly displayed.

Experience has shown that Delphi panels often do reach consensus on many events. The panel members often have widely varying estimates on each event on the second round. However, as the panelists offer their reasons for shifting the estimates, the subsequent estimates tend to cluster near preferred dates. This convergence results (at least in part) from actual transfer of information and interaction among the panel members. It is this interaction and shifting of responses that makes Delphi different from public opinion polls.

Panel members do not always shift their opinions under the influence of the arguments of other panelists. Delphi panelists have just as much opportunity to stick with their original views as do members of a FTF committee. An advantage of Delphi is that panel members can shift positions without losing face when they see convincing reasons from other panel members for a shift of opinion.

2.6 Variations on Delphi

The four-round structure described above is the original format of Delphi. There have been numerous variations from this format. Some of the more important are the following.

Providing an initial list of events. Classical Delphi has been described as "starting with a blank piece of paper." While this has advantages, it may also bother some panelists, who find themselves confused by the unstructured situation. Some users of Delphi have started with an initial list of events generated by some process before the start of the Delphi. The panelists may be asked to make forecasts for these events, effectively going immediately to round two. Alternatively, they may be asked to suggest additional events. The augmented event set then becomes round two.

Beginning with a context. The exact course of the development of a technology will depend upon external political and economic conditions. When these are important, the forecast will depend upon the assumptions made about these external conditions. If the panel is composed of technology experts, they should not be expected to forecast these economic and political conditions as well as the technology. After all, they are not experts on these issues. Hence, it may be desirable to obtain a political and economic forecast and present this to the panelists prior to round one. This provides the panelists with a common context for their forecasts of technology. If the economic and political forecasts are in error, the resulting technological forecasts will also be in error. However, this problem cannot be avoided by failing to provide a context. Doing so simply means that the technology experts will make their own political and economic forecasts, which may all be different. Providing a context can be especially useful in industrial applications of Delphi when a panel of experts has been chosen from the company's technical staff. A context provided by the company's sales, marketing, and top management personnel can provide a helpful guide to the technical experts on the panel.

Number of rounds. Classical Delphi utilizes four rounds. Some Delphis have taken five or more rounds. Experience indicates that four rounds are usually sufficient. Round four can be deleted if the moderator sees no need to obtain rebuttals to the arguments presented in round three. Round one can be omitted if the panel is started off with a list of events. Thus, in some cases two rounds may be sufficient. Since Delphi provides advantages over FTF groups, it should be used if possible, even when a full four rounds cannot be used. Even two rounds may be better than the use of a single expert or a FTF panel.

Some researchers have looked into the issue of when to terminate Delphi studies. The articles by Chaffin and Talley (1980) and Dajani and Sincoff (1979) report the use of the chi-squared test to determine whether stability has been reached. If it has, there is nothing to be gained by continuing the Delphi, and it can be terminated. Thus, the original design for the Delphi can be open-ended in terms of number of rounds, with termination whenever the panel responses have achieved stability.

Multiple dates. In the classical Delphi each panelists provides one forecast for the date of an event. In some cases this is specified as the date by which the event is 50 percent likely to have happened. In other applications of Delphi, however, panelists may be asked to provide three dates. In addition to the 50-percent date, they may be asked to provide "barely possible" and "virtually certain" dates. These may be quantified as 10-, 50-, and 90-percent probability estimates, or some other suitably chosen probabilities. The statistical center of the group response is then obtained by taking the median date for the 50-percent estimates. The degree of disagreement in the panel is represented by the spread between the median dates for the low-probability and high-probability dates.

Computerization. In some Delphis, written questionnaires have been replaced by interactive computer programs. The panel members log on to a central computer that keeps track of the current status of each event and the last estimate made by each panelist. The participants are provided with the current median and quartiles for each event, and any comments that have been added since they last logged on. The computer then reminds the panelists of their most recent estimate and asks if they wish to change it. This approach does away with the round structure. Panelists may log on as often as they choose. Some will do so more frequently than others. Some panelists will change their estimates frequently, while others will permit theirs to stand for a longer time. This "real-time, on-line Delphi" can allow participants to reach stability much more rapidly than via written questionnaires sent through the mail.

Delphi with partial anonymity. Delphi is sometimes used in FTF situations. Arguments are made publicly, while estimates are still made anonymously through secret voting. The panelists discuss an event and then make their forecasts. This may go through several rounds as panelists offer reasons why the others should change their forecasts. The secret voting may be done by using a computer rather than by using pencil and paper. A common approach is to provide each panel member with an input device such as a rotary switch that can be set to a number from 1 to 10 (each number may be assigned to a range of years). The panel member can set the switch without anyone else being aware of the setting. The responses are read by the computer and displayed immediately as medians and quartiles. Participants can quickly see how much consensus has been reached and when stability has been achieved.

They can then decide whether further discussion is worthwhile. The discussion is public, but the responses are anonymous, giving some of the advantages of the full Delphi.

2.7 Guidelines for Conducting a Delphi Sequence

It is a mistake to say, "We don't have time to get a forecast by any other methods, so let's do a Delphi." Probably more people have had bad experiences with Delphi for this reason than for any other reason. Delphi cannot be done "on the cheap." Delphi takes as much time, effort, and expense as does preparing an equivalent forecast by other means.

Even though Delphi is neither cheap nor easy, it can be done with reasonable cost and effort if the more common mistakes are avoided. The following guidelines can help the user avoid such mistakes. They should not be taken as an indication that Delphi is either cheap or easy.

Obtain agreement to serve on the panel. If questionnaires are simply sent out to a list of names, without making sure that these people are willing to serve on the panel, the moderator runs the risk of not getting enough answers to be meaningful, especially if the list of names is a short one. A few attempts to run Delphi sequences have begun by sending the first questionnaire to 200 or 300 names. Response rates typically run 50 percent or less, and 6 to 8 weeks are sometimes required to get even that many responses. In addition to the delay involved, there is no assurance that the same people will respond to every round. The moderator may well be putting in a lot of effort and not gaining any of the advantages of Delphi. The effort may simply amount to running a poll by mail of a very poorly selected group. Choosing the panel members is the most important task of the moderator. However, choosing them is not enough. The moderator must make sure they will actually serve on the panel.

The panel selected should be slightly larger than the moderator thinks will be necessary. Delphis have occurred during which a panel member died, reducing panel size. In addition, if the panel includes the best people available, the moderator must expect that from time to time some of the panelists will have to miss a round because of higher-priority demands on their time. If the original panel is just big enough, any losses such as these may seriously reduce the utility of the resulting forecast.

Explain the Delphi procedure completely. Delphi has been in use for over a quarter of a century. Nevertheless, it is not yet so well known that the moderator can be certain that all the experts selected for a panel are familiar with it or have even heard of it. Even if they are aware of it, they may have only a distorted picture of what is involved and what will be expected of them. It is especially important that they understand the iterative nature of the sequence. Several Delphi sequences have run into problems because some of the panelists did not understand the purpose of the successive questionnaires.

Make the questionnaire easy. In Sec. 2.9, we will discuss the Delphi questions themselves. The point here is to make the mechanics of filling out the questionnaire easy. The format of the questionnaire should be designed to help, not hinder, the panelist. Respondents should be thinking about the forecast, not wrestling with a complicated or confusing questionnaire.

One good approach is to make use of "check the block" or "fill in the blank" questions. This is not always possible, especially if there is considerable controversy over whether the event will occur at all. However, it should be done whenever possible. In addition, the arguments for and against each event should be summarized and presented in a compact form that makes it easy for the panelists to follow the arguments and connect them with the questions. There should be ample room on the questionnaire for the panelists to write in their comments and arguments. Another helpful practice is to include the panelist's response on the preceding round for each event. This saves the panelist from having to check a copy of the preceding questionnaire. It also means, of course, that each questionnaire must be printed individually. In short, the questionnaires should be designed for the convenience of the panelists. Efforts in making the questionnaire easy to answer will directly improve the quality of the response.

Number of questions. There is a practical upper limit to the number of questions to which a panelist can give adequate consideration. This number will vary with the type of question. If each question is fairly simple, requiring only a single number in response to a simple event statement, the limit will be higher. If each question requires considerable thought, with the weighing of conflicting arguments and the balancing of opposing trends, the limit will be lower. As a rule of thumb, 25 questions should be considered a practical upper limit. In special circumstances the number of questions may be higher. However, if the number of questions rises to 50, the moderator should screen them carefully to be sure they all deal with important issues. It is a mistake to dilute the efforts of the panel with minor issues.

Contradictory forecasts. It is entirely possible that contradictory forecasts will appear from the set of questions generated by the panelists during the first round. These might be, for instance, pairs of events that are both possible but are mutually exclusive. In principle, there is no reason why both such events should not be included in the questionnaire. This is particularly true if both are of interest to the moderator. However, it should be made clear to the panelists that the contradictory forecasts arose from their own round-one responses rather than from any attempt by the moderator to "catch them" in an inconsistency.

Injection of the moderator's opinions. The case of deliberate manipulation by the moderator will be discussed below. Here we are concerned with a different issue. From time to time during a Delphi sequence, it may appear to the moderator that the two sides in a debate on some event are not effectively meeting each other's arguments or that there is some obvious (to the moderator) argument or fact that both sides are overlooking. Under these circumstances the moderator may be tempted to include his or her own opinions in the feedback on the next round. This temptation must be resisted without fail. Under no circumstances should a moderator inject personal opinions into the feedback. This advice may seem harsh, but there is no alternative. Once the moderator has violated this rule, there is no recognizable place to draw the line. If a little meddling is helpful, why not a little more? This can continue until the entire forecast is distorted to conform to the views of the moderator. If the moderator's own opinions are injected into the feedback, there is a risk of converting the Delphi sequence into an elaborate and expensive means of fooling the moderator, or fooling the client who is impressed by the names of the panelists.

The moderator has gone to considerable trouble picking a panel of experts, people who presumably know much more about the subject than does anyone else. Their deliberations should not be meddled with. If a moderator becomes convinced that the panelists are overlooking some significant elements of the problem, this means that in some way the panel is unqualified. The only solution is to discard the forecast and repeat the work with another panel.

Payment to panelists. In the early days of Delphi, most panelists were unpaid. It was considered almost an honor to be asked to participate. However, those days are long gone. The moderator of a Delphi panel is asking for time and expert advice from the panelists and should be prepared to pay for these valuable commodities at market rates. The forecast is presumed to be valuable to the organization asking for it; a bad forecast may cost much more than just the cost of preparing it. Thus, the panelists should be paid at customary consulting rates.

Professional societies and charitable institutions may still be able to obtain unpaid Delphi panelists. Experts may be as willing to lend their time and knowledge to these organizations as they are to donate money or other kinds of effort. Nevertheless, moderators for such Delphis should remember that they are asking for something valuable and are depending on the good will of the panelists, which should not be abused.

Workload involved in a Delphi sequence. During the Delphi, the main task of the moderator is to receive and analyze responses from the panelists and prepare the questions for the next round. Experience shows that this will require about 2 person-hours per panelist per round. The clerical workload in preparing the questionnaire is about the same, but the timing is different from the moderator's workload.

For large panels, computerizing the analysis is almost essential. Even for panels of 50, the manual-processing workload is so heavy that there is no time for adequate analysis, and the turnaround time becomes excessive. Even for small panels, computerizing the computation of medians and quartiles is often worth the effort.

Turnaround time between questionnaires. Delphis that are transacted through the mail usually take about a month between successive questionnaires. When Delphis are carried out within organizations located in a small area (plant, laboratory, university campus, etc.), turnaround times can be much shorter. For panels of 10 to 15 members, using interoffice mail or couriers, 2 weeks has often been sufficient to carry out 4 full rounds. However, the panelists must be motivated to respond promptly, or the advantages of internal communication can be lost.

2.8 Selecting Delphi Panel Members

In recent years there has been some debate in the forecasting community about whether "experts" are any better at forecasting than "nonexperts." Many of the experiments that allegedly found no better performance among experts, however, suffered from the defect that the term "expert" was often not defined carefully. The alleged experts often really knew no more about the questions asked than did the nonexperts with whom they were compared. See Goldschmidt (1975) for an analysis of this criticism.

This point becomes important in the selection of experts for Delphi panels. The panelists should be experts in the sense that they know more about the topic to be forecast than do most people. On all other topics, however, the panelists may know even less than most people. A panel member should be selected for expertise *with regard to the topic to be forecast.* Expertise in other areas is irrelevant and is certainly not implied by selection for the panel.

How can the panel moderator identify an expert? Here we will focus on identifying experts in technology. Delphis run for other purposes may require other methods for identifying experts in their subject matter.

There are two aspects to selecting experts for a Delphi panel. First, how does one identify an expert? Second, of the experts identified, which should be selected for the panel? A related issue is whether to select experts from inside or outside the organization.

The question of whether to use inside or outside experts depends primarily on the type of forecast needed and, in some cases, the uses to be made of the results. If the preparation of the forecast requires intimate knowledge of the organization, its history, policies, and so on, then there is little alternative to the use of experts from within the organization. If, however, the forecast does not depend on knowledge of the organization but more on familiarity with some area of technology, then it is probably better to obtain the best people available, and in general, these will come from outside the organization. Except for organizations like large universities, most organizations cannot afford to have on their own staffs more than one or two people of the caliber desired for this type of Delphi panel.

If the forecast is intended to be used in some manner that requires that it remain secret to be effective, then again there is little choice but to use experts from within the organization. The federal government, when obtaining a forecast in an area touching on national security, probably would have little difficulty in maintaining the desired degree of secrecy, even if outsiders were employed to help prepare the forecast. A business firm, however, which hopes to gain a competitive advantage through the effective use of a forecast, probably cannot count too heavily on maintaining a proprietary status for the forecast if really high-caliber outsiders are to serve on the Delphi panel. Some of the desired people may not be willing to serve if the results are to be maintained in a proprietary status. In such cases, the firm is most likely better off using its own people. The employees of the firm may well make up for their lack of expertise, as compared with the best experts available anywhere, with their knowledge of the firm's interests, strengths, and weaknesses.

If the decision is made to use experts from within the organization, the identification of such experts is very much simplified. This is especially true if part of the required expertise is knowledge of the organization itself. The panel moderator will look for people in responsible technical or managerial positions who have been with the organization long enough to have acquired the desired knowledge of its special or unusual features. Evaluation of an employee's level of technical expertise can usually be obtained from supervisors and from records of merit promotions and pay increases. In some cases, the organization chart will be a sufficient guide.

Once the experts within the organization have been identified, there remains the problem of selecting among them. The biggest problem in this regard is that the desirable panelists are busy people. This will be more true the higher they are placed within the management structure. This means that they may not have time to give the Delphi questionnaire adequate attention. In practice, a tradeoff must usually be made between getting panelists whose organizational position gives them a sufficiently broad view and getting panelists who will be able to spend adequate time filling out the questionnaire. There is always a temptation for the panelist to make an estimate coincide with the panel median simply to avoid the problem of justifying a different viewpoint. If the panelist is a busy executive, trying to fill out the questionnaire in his or her spare time, the temptation may be overwhelming, despite a sincere desire to provide a responsive and useful answer. The hasty opinion of a vice president is probably not worth as much as the considered opinion of someone two or three levels lower in the organization.

One problem often overlooked in using inside experts at lower levels in the organization is making sure that their supervisors support their participation in the Delphi. The supervisor is usually responsible for certain results. In principle, the expert is already fully committed to helping achieve those results. If part of the expert's time is diverted to taking part in a Delphi, the supervisor may be held responsible for failure to achieve the prescribed results. It is important, then, to be sure that the Delphi has the support of top management. Diversion of the experts' time from their normal duties will then not be held against their supervisors.

If the decision is made to use outside experts, then the problem of identification is much more difficult. Peer judgment is usually the best criterion for identifying an expert. If the organization has on its own staff one or more specialists in the desired field, they can be asked to nominate outside experts. The first group of outside experts that are selected can themselves be asked to nominate additional experts (this is called the "snowball technique" for forming a panel). A good rule of thumb is to select those who have been nominated by at least two other people. In addition to these nominations, there are other selection criteria that have at least the appearance of being objective and that are in any case useful aids to judgment: honors by professional societies, the number of papers published, the number and importance of patents held, citation rates of published papers, and other signs of professional eminence such as holding office in a professional society.

With outside experts, assuring that they will have adequate time to answer questionnaires is usually not a problem. Outside experts are usually chosen from among people who have considerable control over their own time, such as university faculty members and private consultants. Their agreement to serve on a Delphi panel can be construed as a commitment to devote adequate time to preparing the forecast. The most serious problem is finding a panel who will not only agree to serve but also be available for the full sequence of questionnaires. University faculty members, for instance, tend to do a great deal of traveling during the summer. If the panel is to be staffed mainly with university faculty members, the sequence should be timed so that it can be completed during the academic year.

Given that a set of experts has been identified, which of them should be asked to serve on the panel? Or, viewing it from a more practical standpoint, which should be asked first in the hope that they will agree to serve and that it will not be necessary to contact others? How can the panel moderator establish a priority ranking among the potential panelists? Degree of expertise, as determined during the initial search, is probably the most important single consideration. The forecast should represent the best opinion available; hence, the panel should be composed of the most knowledgeable experts available. After that, considerations such as likely availability and probable willingness to serve can be taken into account.

There is another factor that must be given consideration during the selection of the panel. As pointed out earlier, one of the difficulties with any forecast prepared by a group is the problem of common or cultural bias. If the members of the panel share some set of biases, these will almost inevitably show up in the forecast. The panelists themselves are unlikely to be aware of them. There is no absolute guarantee that this problem can be eliminated. It can only be minimized by selecting representatives of every major school of thought in the subject area. If there are people within the organization who are sufficiently familiar with the field, they may be asked to identify the major schools of thought and to indicate which experts belong to which schools. The panel moderator can also make use of the various "Who's Who" publications, rosters of professional societies, and so on, to determine the background of each expert. Facts such as previous employers, schools attended, and identity of their thesis advisor(s), can be used to help assure that a panel is not inadvertently chosen that has a one-sided outlook. If this kind of information is not readily available, then the panel should be chosen to include members with widely varying ages and representing a variety of institutions with as wide a geographical spread as possible. As pointed out earlier, the Delphi panel is not supposed to be a statistical sample of the "expert community" in question. Instead, it should include representatives of all views. Inevitably, then, the minority viewpoints will be overrepresented.

It cannot be emphasized too strongly that choosing the panel is the most important decision the panel moderator will make. Considerable effort in making a good selection is fully justified.

2.9 Constructing Delphi Event Statements

A Delphi questionnaire is neither a public opinion poll nor a psychological test. Many critics have failed to understand this and have complained that Delphis do not follow the rules developed for questionnaires in these fields. There is no reason Delphis should follow those rules. However, there are some rules that must be followed if Delphi questionnaires are to obtain the information the moderator wants. Some of the most important rules are the following.

Avoid compound events. If the event statement contains one part with which a panelist agrees and another part with which he or she disagrees, there can be no meaningful response. Consider the following event statement: "A plant generating electric power from nuclear fusion using deuterium from sea water will begin commercial operation in ________." The panelist who thinks that nuclear fusion will be based on the use of tritium cannot respond to this event. If he or she responds with the date when the panelist thinks that tritium-based power will be available, this may be interpreted as a "vote" for deuterium from sea water. If the response is "never," it may be interpreted as doubting that fusion power will ever become commercial. In general, it is best to avoid event statements of the form "Capability A will be achieved by method B in the year ________." Replace the compound question with two or more simple questions.

The moderator can never be certain to have eliminated all compound events. Despite one's best efforts, some panelists may find two distinct parts to what was intended to be a single event. In such a case, the feedback between rounds can help the moderator improve the question. Clarifying an event statement on the basis of feedback may be as important as the forecast itself if it uncovers alternatives that were not apparent at first.

Avoid ambiguous statements of events. Ambiguity can arise from the use of technical jargon or from terms that "everyone knows." Most ambiguity comes from the use of terms that are not well defined. Consider the following event statement: "Terminals linked to central computers will be common in private homes by the year ________." The question is ambiguous. How common is common? Ten percent of all homes? Fifty percent? Ninety percent? If 70 percent of all homes with incomes over $40,000 have such links but only 10 percent of homes where the income is below $40,000, is that "common"? Descriptive terms such as "common," "widely used," "normal," "in general use," "will become a reality," "a significant segment of," and so on are ambiguous and should not be used. Ambiguity can often be eliminated by using quantitative statements of events. However, consider the statement, "By 19 ___ the per capita electric power consumption in Africa will be 25 percent of the U.S. per capita power consumption." Does this mean 25 percent of today's U.S. consumption, or 25 percent of the U.S. consumption in the same year? Even though the use of a percentage makes the statement appear quantitative, it is still not clear. Consider the statement, "By 19 ___ a majority of all foods sold in supermarkets will be radiation sterilized and will not require refrigeration." Does this mean over 50 percent of each kind of food? Or does it mean over 50 percent of the total, but some foods not at all? In either case, is the 50 percent by weight, volume, or dollar sales? It is important to use terms that will not be interpreted differently by different panel members.

Avoid too little or too much information in event statements. Research by Salancik et al. (1971) shows that it is just as bad for a statement to have too much information as too little. The researchers compared the degree of consensus in the forecast with the complexity of the statement. They measured complexity by the number of words, which is a crude but objective measure of complexity. They also used a measure of consensus more sophisticated than the interquartile range. They borrowed a concept from information theory, where the information content of a message is measured in "bits." One bit is the information contained in a single "yes" or "no" when both are equally likely. That is, one bit of information is just enough to answer a binary question. So the degree of consensus was measured in bits by comparing the actual distribution of forecasts with a condition of complete uncertainty, i.e., with a uniform distribution of forecasts over the entire possible range of years to the time horizon.

Complete consensus on their questionnaire would have provided 2.58 bits. However, the actual degree of consensus provided an average of only 0.6 bits per event, about one-fourth of the maximum possible. On the average, the greatest consensus was achieved for event statements about 25 words long, which provided about 0.85 bits. The degree of consensus declined for either longer or shorter statements. The number of bits provided was only about 0.45 for both 10- and 35-word statements.

Some of the event statements dealt with technology, while others dealt with applications. For the technology events, the shorter the event statement, the higher the degree of consensus. For application statements, the reverse was true. The researchers tested the possibility that the panelists were more familiar with the technology than with the applications, which would mean that fewer words were needed to define the technology events adequately. They divided the applications statements into three categories on the basis of degree of use, from common to unusual. For the more common applications, the most consensus was reached for the shortest statements. For the unusual applications, the most consensus was reached for the longest statements.

The panelists were asked to rate their expertise on each question. For the nonexperts, the longer a statement, the greater the consensus reached. For the experts, the most consensus was reached at intermediate-length statements, with consensus declining for both longer and shorter statements.

The conclusion is that if an event is unfamiliar to the panelists, the more description given, the greater the degree of understanding and the greater the degree of consensus. If an event is familiar, the more description given, the more confusing the statement appears and the less the degree of consensus. Event statements should, therefore, be chosen to provide neither too much nor too little information. If the panel has trouble reaching consensus, the problem may be an event statement that provides either too much or too little information. The moderator should attempt to clarify this and reword the statement.

2.10 Manipulation of Delphi Results

Clearly the moderator of the Delphi panel plays an important role in the outcome, even though he or she ordinarily does not respond to any of the questions. One of the issues that has been of interest to Delphi researchers is the extent to which the moderator can manipulate or distort the results of the Delphi by falsifying the feedback presented to the panelists.

If the moderator is also the intended user of the forecast, it is difficult to see what he or she is accomplishing by such an act. The user is fooling only himself or herself and can do that more cheaply by not doing the Delphi at all. However, if the moderator is preparing a forecast for someone else, then manipulation of the results amounts to misleading the client.

Nelson (1978) found that the results could be distorted by false feedback. In his experiments, he falsified the feedback on every round. The end result was that for almost every event, the median estimates on the final round converged to the predetermined result he was trying to achieve or to a value close to it. This result is not surprising. Since Delphi panelists are interacting with each other, false feedback would distort the responses of the panelists.

Scheide et al. (1975) found a somewhat different result. In their experiments, feedback was falsified on only one round. After that, correct feedback was provided to the panelists. They found that panel responses followed the falsified feedback for one round, then returned to the original values. That is, the panel responses "recovered" from a one-round falsification of the feedback.

The fact that Delphi results can be manipulated is not grounds for rejecting Delphi. The results of almost any other forecasting method can be manipulated by an unscrupulous forecaster. These findings, do, however, emphasize that the integrity of the Delphi moderator is important in getting valid results from Delphi.

2.11 Using TEKFOR

Analyzing Delphi results, including summarizing arguments and interpreting responses, involves a great deal of intellectual effort on the part of the moderator. However, computing medians and quartiles is a purely mechanical exercise. It amounts to nothing more than sorting the responses to each question, and counting down one-fourth, one-half, and three-fourths of the way through the sorted list. This can be computerized. The program TEKFOR allows the user to type in the responses in any order. It then sorts them and prints out the median and quartiles. It, thus, relieves the Delphi moderator of the "intellectual coolie labor" involved in preparing the statistical feedback required for the next round.

2.12 Summary

In the quarter century since Gordon and Helmer (1964) brought Delphi to public notice, thousands of Delphi sequences have been run by hundreds of organizations and groups, for a wide variety of purposes. Descriptions of many of these sequences have been published in report form, as well as in the form of articles in journals devoted to management, planning, forecasting, and education.

A recent attempt to summarize research findings on Delphi (Rowe et al., 1991) found that in all too many cases, different researchers used different forms of "Delphi" for their studies. It is, therefore, difficult to reach strong conclusions from the Delphi research literature. However, while recommending further research on specific issues, the authors do not reject Delphi but conclude that the three characteristics of Delphi do provide face plausibility for its value. (This study contains an extensive bibliography of research papers on Delphi.)

Despite the lack of comparability among many of the studies of Delphi, some conclusions can be drawn that should be of value to those considering the use of Delphi.

Delphi does permit an effective interaction between members of the panel, even though this interaction is highly filtered by the moderator's summarizing of the arguments. Several experiments in which the panelists were asked to give reasons why they changed their estimates showed that the panelists were, in fact, reacting to the views of the rest of the panel. This is particularly evident in those experiments involving deliberately falsified feedback. However, this cannot be viewed as weakness of will on the part of panelists. In one such experiment, one of the panelists claimed that it made him even more "stubborn" to know that "only I had the right answer." Panelists do shift their estimates when the arguments of their fellow panelists are convincing; otherwise, they will hold tenaciously to their differing opinions.

At the same time, however, there is ample evidence from a number of experiments that if the panelists feel that the questionnaire is an imposition on them or if they feel rushed and do not have time to give adequate thought to the questions, they will agree with the majority simply to avoid having to explain their differences. In this respect, therefore, the Delphi procedure is not an absolute guarantee against the degrading influences of the "bandwagon effect" and fatigue. However, in a Delphi, these problems are to some extent under the control of the moderator, while they are virtually uncontrollable in a FTF committee or problem-solving group.

The Delphi procedure is, thus, a feasible and effective method of obtaining the benefits of group participation in the preparation of a forecast while at the same time minimizing or eliminating most of the problems of committee action. It can take longer to complete a Delphi than a FTF committee meeting, especially if the deliberations are carried out by mail. Since it is unlikely that a long-range forecast will need to be prepared in a hurry, this delay need not be a disadvantage and can be dealt with through adequate advanced planning. Moreover, it may be more feasible to prepare a forecast using Delphi than to find a time when a dozen busy people can meet together. Thus, whenever adequate time is available, Delphi should be considered as a practical approach to obtaining the required forecast.

References

Chaffin, W. W. and W. K. Talley, "Individual Stability in Delphi Studies," *Technological Forecasting & Social Change,* **16(1)**:67–73 (1980).

Dajani, J. S. and M. Z. Sincoff, "Stability and Agreement Criteria for the Termination of Delphi Studies," *Technological Forecasting & Social Change,* **13(1)**:83–90 (1979).

Goldschmidt, P. C., "Scientific Inquiry or Political Critique?" *Technological Forecasting & Social Change,* **7(2)**:195–213 (1975).

Gordon, T. J. and O. Helmer, "Report on Long Range Forecasting Study," Report P-2982, The RAND Corporation, Santa Monica, CA, 1964.

Linstone, H. A. and M. Turoff, *The Delphi Method,* Reading, MA: Addison-Wesley, 1975.

Nelson, B. W., "Statistical Manipulation of Delphi Statements: Its Success and Effect on Convergence and Stability," *Technological Forecasting & Social Change,* **12(1)**:41–60 (1978).

Rowe, G., G. Wright, and F. Bolger, "Delphi: A Reevaluation of Research and Theory," *Technological Forecasting & Social Change,* **39(3)**:235–251 (1991).

Salancik, J. R., W. Wenger, and E. Helfer, "The Construction of Delphi Event Statements," *Technological Forecasting & Social Change,* **3(3)**: 65–73 (1971).

Scheide, M., M. Skutsch, and J. Shofer, "Experiments in Delphi Methodology," in H. A. Linstone and M. Turoff (eds.), *The Delphi Method,* Addison-Wesley, Reading, MA, 1975, pp. 262–287.

Problems

1 Which of the following items are likely to require expert judgment to determine, and which would be better obtained by some objective means?

- *a* The likelihood of the Supreme Court upholding a patent on a technological innovation
- *b* The profitability of a new device as compared with the device it will replace
- *c* The likelihood of new technology being rejected on moral or ethical grounds
- *d* The willingness of the public to accept a specific alternative to the automobile for personal transportation
- *e* The federal government's probable response to a new technological advance

2 Assume that the laboratory feasibility of a radically new technological device has just been demonstrated. This device is based on new principles and is largely the work of one expert, who therefore knows much more about it than anyone else in the world.

You want to obtain a forecast of its future level of functional capability and degree of acceptance by potential users. Is it worth supplementing the judgment of the inventor by organizing a committee, with the inventor as a member, to prepare a forecast? If your answer is yes, what characteristics would you look for in selecting other members of the committee?

3 If a forecast is being prepared by a committee, would you insist that the committee forecast only those things on which a majority of the members agree? Would you insist on unanimous agreement? Is insistence on agreement likely to produce a better forecast?

4 What are the relative advantages and disadvantages of asking the panel to suggest events whose times of occurrence are to be forecast in subsequent rounds?

5 When would it be desirable to provide a panel with the economic or political contexts for the technological forecast it is to produce?

6 Your company wants a forecast of technological advances that may supplant the current products or provide it with new products. It is decided to obtain the forecast using a Delphi panel. What are the relative advantages and disadvantages of the following panel types?

a A panel of experts from outside the company
b A panel of experts from within the company
c A panel combining both company experts and outside experts
d Two panels, one of company experts and one of outside experts

7 You are an official of a charitable organization that has in the past supplied funds for much of the medical research on a particular class of diseases. Your research program has been successful. Cures or satisfactory preventatives for these diseases are expected to be available within the next few years. You need to determine the avenues of medical research toward which your organization should shift its support. What kind of a panel (or panels) would you select to provide forecasts useful in this situation?

8 Your company manufactures a type of device that, traditionally, has been bought and installed by the federal government for widespread public use. Technological progress in the field has been rapid, with successive devices being rendered obsolete by improvements every few years. As a guide to your company's long-range planning, you wish to obtain a forecast of likely technological progress in the field over the next 20 years. What type of members would you include on a Delphi panel?

9 What is wrong with each of the following Delphi event statements? How might they be improved?

a Computer-controlled education for self-teaching will be available in the home by the year ______.
b The teaching-machine market will be a significant part of the total market for educational materials and equipment by the year ______.
c Power from nuclear fusion will be a reality by the year ______.
d Electric automobiles will be common as "second cars" by the year ______.
e A majority of office clerical operations now handled manually will be done by computer by the year ______.

Chapter

3

Forecasting By Analogy

"I have but one lamp by which my feet are guided, and that is the lamp of experience; I know of no way of judging the future but by the past." PATRICK HENRY
Speech in the Virginia Convention
March 23, 1775

3.1 Introduction

New technological projects are sometimes compared to older projects in terms such as, "This is even bigger than the moon-landing project was for its time," or "This isn't as big as the Manhattan Project was for its time." The idea is to convey the relative difficulty of the project with respect to the conditions of the time. This is an analogy.

The use of analogies in forecasting builds on this notion. Forecasting by analogy attempts to answer such questions as, "What would it take today for a project to be as difficult a technological challenge as the Polaris missile project was for its time?", "Economically, was the Aswan High Dam project on the Nile River more or less expensive than the Egyptian pyramids were for their time?", or "What would it take for a new technology to have the same potential for impact and application as the airplane did in 1903?" By answering questions like these, the forecaster can roughly estimate the chances for successfully forecasting a current project.

It must be recognized at the outset that forecasting by analogy is essentially a qualitative method. It cannot produce numbers. Nevertheless, it is a conscious and deliberate attempt to draw upon historical experience.

Forecasting by analogy involves a systematic comparison of the technology to be forecast with some earlier technology that is believed to have been similar in all or most important respects. But what does it mean to be "similar," and which respects are "important"? Answering these questions is the whole point behind the idea of systematic comparisons. This chapter presents a method for identifying those comparisons that are important and estimating their degree of similarity.

3.2 Problems with Analogies

The use of analogies is subject to several problems. These must be understood so they may be overcome. These problems are: lack of inherent necessity, historical uniqueness, historically conditioned awareness, and casual analogies.

Lack of inherent necessity simply means that the outcome of a historical situation is not completely determined by physical factors. The fall of an object, for instance, is completely determined by factors such as gravity, air resistance, initial velocity, and so on. Knowing these factors, we can "forecast" very accurately what the trajectory of the object will be. There are no similar "laws" governing the outcome of historical situations.

A forecaster attempting to forecast the outcome of a current situation may identify a "model" historical situation, which he or she then compares with the current situation. If the two are sufficiently similar, the forecast would be that the current situation will turn out as the model situation did. However, there is no guarantee of this. The current situation will not necessarily follow the pattern of the model situation; only in Greek tragedy is the outcome inevitable. Moreover, a study of historical records will show that the actual outcome of most historical situations turned out to be a surprise to at least some of the participants, who at the time expected something else. Some historical outcomes appear implausible even in retrospect. Even though we know what happened, we find it hard to believe that people acted as they did. Analogies are based on the assumption that there is a "normal" way for people to behave and that given similar situations, they will act in similar ways. However, there is no guarantee that people today will act as people did in the model situation. Hence, the forecast is at most probable, never certain.

Historical uniqueness simply means that no two historical situations are ever identical in all respects. Thus, it is important to be able to say which comparisons are important and which can be ignored. An analogy will be strengthened if there are several historical cases with parallel outcomes that can be compared to the present case to be forecast. However, since each of these historical cases is unique in itself, it is important to be able to determine whether they are "similar enough" to each other to be considered analogous. Thus, a systematic means for comparing model situations with each other and with the current situation is essential.

The problem of *historically conditioned awareness* arises because people may be aware of what happened "the last time." Even though a historical situation may be judged to be sufficiently similar to the present situation to be called analogous, people may be aware of the previous outcome. If they do not like the way the previous situation turned out, they may deliberately act differently this time in order to secure a more preferred outcome. Historically conditioned awareness, then, violates the assumption that there is a "normal" way for people to behave and that they always behave that way. Thus, despite a forecaster's best efforts to check for analogous situations, the forecast may be invalidated by people's awareness of prior outcomes.

A *casual analogy* arises when we observe that two things are alike in a few respects, and assume without evidence that they are alike in most other respects. It is possible, of course, that such an analogy is valid. However, in most cases it will be in error. It is not sufficient to observe that two things are alike on one or two characteristics. It is necessary to make a systematic check of all the observable characteristics before you can risk concluding that the two are alike on as of yet unobservable characteristics.

Despite these problems, analogies can be a very useful method for forecasting technological change. The problems cannot be solved completely, but they can be minimized by using a systematic method for establishing analogies.

3.3 Dimensions of Analogies

In forecasting by analogy, we wish to compare a historical or model situation with a current situation in order to determine whether the two are "analogous." Since we are primarily concerned with technological change, it is important to compare the two situations on the basis of factors that affect technological change—the invention of some device or procedure, the adoption of the invention, and the widespread diffusion of the invention. The following list of dimensions is based on studies of factors that have affected technological change and therefore provides a suitable basis for comparing situations for a possible analogy.

1. Technological
2. Economic
3. Managerial
4. Political
5. Social
6. Cultural
7. Intellectual
8. Religious-ethical
9. Ecological

In the following section, we discuss briefly how each of these dimensions has affected technological change in the past, and why, therefore, it is important to compare two possibly analogous situations with respect to these dimensions.

The technological dimension. Technologies do not exist in isolation. Every technology exists to perform some function desired by people. There are usually alternate ways of performing the same function that compete with the technology to be forecast. The technology must also draw on supporting technologies. Finally, the technology must mesh with complementary technologies that perform other functions. A comparison of two situations with respect to the technological dimension requires that each of these aspects be compared.

The technology used as the "model" had to be superior to its competing technologies or else it would not have succeeded. The forecaster must determine what were the competing technologies in the model situation and identify the competing technologies in the current situation to be forecast and compare them. What are the weaknesses of the competing technologies then and now? Are they comparable? What are the strengths of the model technology and the technology to be forecast? Are they comparable? Does the technology to be forecast provide the same degree of advantage over competing technologies as did the model technology?

A technology must draw on other technologies for support in such things as production, maintenance, energy supply, and transportation. For instance, the Newcomen steam engine, invented in 1706, was limited in its power output by the size of the cylinder. Machine tool technology was not good enough to allow machinists to bore large cylinders that were round enough to have tight-fitting pistons. In 1744, Wilkinson built the first precision-boring lathe, which was used to bore cylinders up to 50 in in diameter for Watt's steam engine and did it with unprecedented accuracy. This supporting technology was critical to the success of Watt's steam engine. It is necessary to identify the supporting technologies in the current and the model cases. Is the degree of support provided comparable in the two cases?

A technology must be compatible with complementary technology, usually in terms of input and output. This compatibility may be as simple as a common truck bed height or railroad gauge, or it may be as sophisticated as the pitch and shape of screw threads or the frequency of an alternating-current power supply. The steam engine provides another example of this need for complementarity. During the first half of the eighteenth century, European industry used machines such as grinding and crushing mills driven by wind or water power, which required rotary motion in their motive power. They could not have been driven by a steam engine, because all steam engines of the time produced reciprocating motion (which was, however, ideal for driving pumps). Only after Watt invented the sun-and-planet gear to convert reciprocating into rotary motion was the steam engine used for driving industrial machinery. It is necessary to identify the complementary technologies in the current and the model cases. Is the "fit" the same in the two cases?

In examining the technological dimension, one must look beyond the physical hardware to the theories behind both the model technology and the technology to be forecast. The forecaster must ask whether the predictive and explanatory powers of the theories and the level of understanding of the relevant phenomena are the same in the two cases. The question is not whether one theory is more advanced than the other, since the more recent theory is likely to be more advanced. The question is whether the theories in both cases were equally capable of meeting the demands placed on them.

As an illustration of the technological dimension of a possible analogy, consider the comparison between the Suez Canal and the U.S. attempt to build the Panama Canal presented in Table 3.1. Clearly the Panama Canal was a harder task; the same engineer (deLesseps) who built the Suez Canal attempted to build a canal in Panama using Suez Canal technology and failed. The real question is, though, whether by the time the United States attempted the Panama Canal, the technology had advanced to the point where it was no harder a job than the Suez Canal had been 40 years earlier. The comparison in Table 3.1 shows that while each canal set some technological precedents, in general the two can be considered comparable, given the state of the art of construction technology at the respective times they were attempted. Also note that the only theoretical issue involved in the two projects was the theory of lock design. The Suez Canal did not require locks; Gatun Lock at the Panama Canal was well within the state of the art, using principles already demonstrated in the lock at Sault Ste. Marie. Hence, for both

TABLE 3.1 Comparison of Technological Dimensions of the Suez and Panama Canals

	Suez Canal	Panama Canal
Size of Excavation, $\times 10^6$ m^3	75 15 by hand, 60 by machinery	160 63 by dredger, 97 by steam shovel
Available earth-moving machinery	Adequate for task; 60 dredgers (75 hp) moved 2000 m^3/day	Adequate for task; 100 steam shovels moved 38,000 m^3/day
Topography	Flat	Mountainous
Soil type	Sand and mud	Rock and clay
Size of previous construction projects	Longer than any previous sea level canal	World's largest rock-crushing plant at Colon; Gatun Lock: poured 3000 yd^3/day of concrete; previous record 1700 yd^3/day
Medical problems	Cholera epidemic, 1865, solved by distilling sea water	Knew that malaria and yellow fever were mosquito-borne; solved by drainage and screening
Theory of lock design	Not needed	Biggest canal lock in the world at the time was Sault Ste. Marie, which was in service 20 years; Gatun lock only slightly larger, based on same principles
Complementary technology	Required steam ships; prevailing winds ruled out use of sailing ships	Required steam ships; prevailing winds ruled out use of sailing ships

projects, the existing theory was completely adequate. No new theoretical developments were required.

The economic dimension. Technology is intended to perform some function. It will be used for that function only if people are willing to pay for it, i.e., give up alternate uses for the resources needed to deploy the technology. Therefore, the forecaster must look at the ability and willingness of the relevant people, in both the model and the present cases, to pay for the technology.

There may be two relevant costs to be considered: the cost to deploy the entire system embodying the technology and the cost to the ultimate customer to use the system. These may well be independent issues. It may be possible to raise the money to deploy some technology, but potential users may not be willing to pay to use it. Conversely, it may be that people would be willing to pay their share of the cost to use something, but the resources to deploy it cannot be mobilized.

An example of the former situation is mass transit in the United States. Numerous subway or street railway systems have been built with money raised through taxation. However, people refuse to ride them in numbers sufficient to pay back the cost of construction and operation. The Washington, D.C., subway, for instance, charges less than a dollar a ride, even at rush hour, and still cannot attract people away from the traffic jams on the streets above. The amortized cost of the system at the actual level of ridership, however, including construction and operation, is about $20 per ride. The potential users clearly are not willing to pay the full cost of deploying the system, even though it was possible to mobilize the resources to build it.

An example of the latter situation also arises in the field of public transportation, with the "jitney." These are public transportation vehicles that are small (like taxis) but travel fixed routes (like buses). Because they share riders and maintain schedules, thus attracting regular riders, they can be profitable with rates lower than taxis. Because they carry only a few passengers, they are conventionally scheduled at shorter intervals than buses. Wherever they have been tried, jitneys have proven economically successful. However, they have been prohibited in most U.S. cities, because they attract business away from more conventional mass transit systems and taxis. Thus, here is a system that the public has shown it will pay for but for which financing cannot be raised because of legal restrictions. In general, in a market economy, if people are willing to pay for some service, means will be found to finance its deployment. Failure to deploy a popular system usually results from legal restrictions of one kind or another.

In any case, the question regarding the economic dimension is whether the money can be raised to deploy the system, and then whether users will support it.

In comparing two technologies for a possible analogy, we must examine the cost of each relative to the resources available for deployment and operation. Relevant costs may include research and development, capital investment, manufacturing costs, and the costs of support and operation. Even the costs of training operators and technicians may have to be included. However, the proper comparison is not between absolute costs but between relative costs. The range of costs of the engineering projects listed in Table A.43 (in App. A) covers six orders of magnitude. If we were to develop an analogy between two of them, however, we would be concerned not with the dollar cost of each but with the cost of each as a fraction of the total resources available to those persons supporting the projects.

In addition to comparing the relative costs of two possibly analogous projects, we must look at the financial mechanisms available for mobilizing resources. Can money be raised on the scale necessary? Are risk-spreading mechanisms available so that many segments of society can share in the funding of the project? The essential feature in an analogy is not the similarity of financial mechanisms in the two cases, but their respective abilities to raise the necessary resources.

It is also necessary to compare the "market" in the two cases. That is, to compare the demand for the technology in terms of market size. History is full of cases of two essentially similar technologies, one of which succeeded because it had an adequate market and the other of which failed because a market was lacking. For instance, McCormick is generally credited with the invention of the reaper, yet his machine differed little from any invented in the decades prior to 1841. The reason why McCormick's reaper succeeded where previous ones failed is that McCormick introduced it just when the rolling farmlands of Ohio, Indiana, and Illinois were being opened to wheat farming. The rocky and uneven soil of the older states had been unsuitable for mechanical reaping, and machines invented earlier lacked the market McCormick had.

In addition to the size of resources, it is necessary to compare economic climates. If the historical technology was introduced at a time of economic expansion whereas the present is a time of contraction, an otherwise good analogy may be invalidated.

Finally, it is necessary to compare the two situations in terms of economic theories. Differences in economic theory may spoil an analogy despite agreement on other elements of the economic dimension. As an example, consider Marxist economic theory. Marx held that the owner of capital performed no service and, therefore, deserved no payment for its use. Hence, prior to the 1950s, industrial enterprises in the Soviet Union were not charged interest on the money they borrowed to finance expansion. Thus, capital was costless. Workers, however, had to be paid wages. The result was that Soviet factory managers emphasized automated and highly mechanized factories, which led to a great deal of "hidden unemployment." Under similar conditions, a manager who had to pay for both capital and labor would have employed more workers and less automation. Here is a clear-cut case where economic theory affected the direction of technological change.

As an example of the economic dimension of a possible analogy, consider the Manhattan Project, which developed the atomic bomb during World War II, and the U.S. Air Force's Ballistic Missile Project, which developed the Atlas and Titan ICBMs from 1955 to 1957.

	Manhattan Project	Ballistic Missile Project
Cost	$2.2 Billion	$3 billion
U.S. GNP*	1942–1945: $194.15 billion	1955–1957: $419 billion
U.S. military budget	1942–1944: $48.65 billion	1955–1957: $36.8 billion
Percent of GNP		
Total	1.133	0.72
Annual	0.28	2.71
Percent of military budget		
Total	4.52	8.15
Annual	1.13	2.71

*GNP: gross national product.

The Manhattan Project and the Ballistic Missile Project represented about equal proportions of the GNP at the times they were carried out. The Ballistic Missile Project represented about twice the share of the military budget that the Manhattan Project did. However, the 1942–1945 military budget included the costs of waging a war, while the 1955–1957 budget was a peacetime budget. Hence the relationship to GNP is probably more meaningful. We can, thus, conclude that from the economic standpoint, the Ballistic Missile Project was about the same size as the Manhattan Project, and the two are analogous on this dimension.

The managerial dimension. The introduction of technological change does not just happen; it must be managed. For instance, it has already been suggested that on the technological dimension, the Suez and Panama Canals were analogous. However, on the managerial dimension, they were not. It turned out that the key to the successful completion of the Panama Canal was not digging the dirt but managing its disposal. This required the scheduling and dispatch of trains to each relay of steam shovels, on a system of tracks that was continually being torn up and moved, because the next relay of steam shovels was excavating the ground from under the tracks. This managerial task was carried out by a cadre of engineers who had recently completed the construction of the U.S. railroad system and were no longer required for that task.

In developing an analogy, then, it is necessary to compare the levels of managerial capability in the model and current situations relative to the size and complexity of the task.

Size depends upon the number of people to be managed and their geographic spread. How big was the model task relative to tasks already successfully managed? Was it comparable or even smaller? Was it bigger? Then, how big is the current task relative to tasks already managed successfully? Is it comparable? Is it larger? Is it smaller? If it is larger, is the increase in size no greater than in the model task?

Complexity is measured by the number of different types of activities involved and the number of locations at which these must be carried out. Was the complexity of the model task greater or lesser than the complexity of tasks previously managed successfully? Is the complexity of the current task within the demonstrated capabilities of project managers? If it is greater, how does the increase in complexity compare with that in the model project?

What of the pool of available managerial talent? It is not a question of absolute size but of relative size. Does the current project demand a bigger share of the pool of talent than did the model project? If so, the analogy would be invalid.

Finally, it is necessary to compare managerial techniques. The ability to manage "larger" projects comes from better managerial techniques and procedures—managerial "technology." The comparison then demands that the managerial techniques in both the model and current cases be equally adequate to the task. A difference in the absolute level of managerial technique is to be expected; however, a difference in the relative adequacy in the two cases would invalidate the analogy.

As an example, we consider the managerial dimension of a possible analogy between the Manhattan Project and the Ballistic Missile Project.

Manhattan Project	Ballistic Missile Project
19 prime contractors	7 major contractors
37 installations at 10 major sites	10 sites
Average labor force: 5900 government personnel 37,800 contractor employees	Average labor force: 2900 government personnel 47,000 prime contractor personnel 84,000 subcontractor personnel
U.S. labor force (1944): 66 million	U.S. labor force (1956): 70 million
Percent of total labor force: 0.066	Percent of total labor force: 0.071
Largest single site: Hanford, WA Peak labor force: 45,000 Subcontractors to DuPont: 4000	

The Ballistic Missile Project employed more people and had a slightly larger share of the U.S. labor force than did the Manhattan Project. However, fewer prime contractors were involved, and the project required coordination of activities at fewer sites. These two factors somewhat offset each other. Nevertheless, from the managerial standpoint, the Manhattan Project was probably larger in its time than was the Ballistic Missile Project in its time.

The political dimension. *Webster's Ninth New Collegiate Dictionary* defines politics as "competition between competing interest groups or individuals for power and leadership (as in a government)." Innovation represents a change that will affect the relative power of different groups. The basic questions about the interaction of politics and technological change, therefore, are, "Who benefits?" and "Who gets hurt?"

Those who realize they will benefit from a technological change will try to encourage it; those who realize they will suffer from it will try to stop it, often by legal or political means. Hence, it is necessary to compare the relative political power of the people who benefit and those who get hurt in both the model and current situations.

Those who may suffer from the introduction of a new technology include the producers, suppliers, and operators of the old technology. Some other "losers" are the people involved in related institutions, such as craft unions or government regulatory bodies. Conversely, "winners" include those who will produce, supply, and operate the new technology, as well as their related institutions. Of course, the consumers are also winners, since if the new technology does not benefit them, they will not pay for it anyway. However, there is a fundamental asymmetry between losers and winners. The potential winners, including those who would get jobs involving the new technology and especially the consumers who would prefer it, may not know they are potential winners. They are likely to be unorganized. The potential losers, however, recognize the threat and organize to stop it. Thus, the potential winners are often unable to take part in the political debates about which technology will be used to provide some service. The potential winners are, thus, often the unfortunate victims in political struggles in which they cannot take part. A good example is cable television, which consumers clearly wanted. The television networks feared it, and the Federal Communications Commission hampered its growth for over a decade in order to protect the networks.

In addition to direct winners and losers, there may be people who have nothing against the technology itself but who oppose it because it hinders political goals they support. The use of nuclear explosives for excavating canals and harbors has been halted, not because the technology itself is such a problem, but because growth in that technology might hamper the political goal of nuclear disarmament. Therefore, those opposed to nuclear weapons have also objected to "peaceful uses" of nuclear explosives. The same problem arose with regard to spaceships propelled by the explosions of "nuclear capsules." Deployment of a fleet of such ships would have meant the wide dispersal of devices that were the equivalent of atomic bombs. Successful development of this technology would have hindered nuclear disarmament.

Developing this dimension of an analogy, then, involves determining the relative power of the supporters and the opponents of the technology. What are their numbers? Can they be mobilized? Can they form alliances, as the opponents of nuclear power have formed alliances with environmentalists and with those opposing nuclear weapons? Will it become a "single issue" for a significant number of voters, determining their choice of candidates? And how do all these things compare with the similar factors in the model case?

In addition to comparing the relative power of supporters and opponents in both the model technology and the current case, it is necessary to compare the political theories in the two cases. What are the rights and duties and the privileges and obligations of different groups in the two cases? If a group wishes to oppose or support a technological change, can it do so within the prevailing political theories and laws? For instance, the success of the "Civil Rights" movement in the U.S. South during the 1960s was due in no small measure to the fact that "passive resistance," including "sit-ins" and boycotts, was politically acceptable. The same tactics were attempted in several nations in Eastern Europe at the same time and led to bloody repression. In those societies, the same tactics were politically unacceptable.

In addition to comparing the model and current situations in terms of political power, the forecaster must compare them in terms of the avenues of support or opposition "legitimately" open to the various groups. Even a numerous and motivated group may be powerless to affect the course of a technology, either to favor it or hinder it, if there are no legitimate avenues of political action open to it.

The social dimension. *Webster's Seventh Collegiate Dictionary* defines society as "a community, nation, or broad grouping of people having common traditions, institutions, and collective activities and interests." Every technological change occurs within a society. Moreover, it both acts on and is acted on by that society. Hence, the forecaster must compare a model situation with the present situation in terms of the society into which the innovation was or will be introduced. The comparison must include the people in the society, their institutions, and their traditions and customs.

The people making up a society can be characterized in terms of total population, age distribution, geographic distribution, urban versus rural distribution, income distribution, levels of education, etc. In comparing two possibly analogous situations, the important consideration is not the absolute numbers but the relative sizes. Was the model innovation introduced into a mostly rural society, while the present innovation is being introduced into a mostly urban society? The difference may invalidate the analogy. However, if both the model and the current technologies are both used by the same segment of society, differences in distribution may not be significant. For instance, if both the model and current technologies are intended for use in agriculture, the rural versus urban mix may not be relevant, since in both cases the target population is agricultural. Age and income distributions of the rural populations may be more significant. The comparison must be made on the basis of the relevant portions of society.

Institutions include the family, schools, businesses, government, fraternal and social organizations, charitable organizations, and churches. Each of these can have a favorable or unfavorable impact on technological innovation. For instance, governments may accelerate or retard innovation by their tax treatment of research and development expenditures, or their regulation of technology. In the United States, government support of agricultural technology has accelerated its growth enormously. By contrast, state and local government building codes and zoning laws have retarded the development of housing technology.

The influence also goes the other way. An innovation may be opposed or supported because it has favorable or unfavorable impacts on one of these institutions. For instance, modern computer and communications technologies have been found to have an unfavorable impact on totalitarian governments. Polish dissidents in the early 1980s found that computers could aid them in organizing their resistance to government repression and in publishing antigovernment newsletters. Dissidents in China in 1989 found that facsimile machines made it possible to communicate among themselves and with foreign supporters, despite the government's control of the public media such as newspapers and television. Because of this potential for antigovernment use, use of photocopying machines was closely restricted in the Soviet Union, and the government of Romania even required typewriters to be registered.

A comparison between two possibly analogous technologies must therefore take into account the interactions between society's institutions on the technologies being compared. Was the effect the same in both cases, whether favorable or unfavorable? Was it favorable in one case and unfavorable in the other? The comparison must also include an analysis of the effects of the technologies on the institutions of society and the possible reactions of the supporters of these institutions.

The traditions and customs of a society bind it together and reflect its self-image. A technological change may be supported if it seems to fit in with this image or enhance or extend it. Conversely, a technological change may be opposed if it seems to undermine tradition or leads the members of a society away from its ideal image.

An example of a technological change that was opposed because it altered a society's self-image is provided by Morison (1966). The steam warship *Wampanoag* was built by the U.S. Navy shortly after the Civil War. This was the fastest, most seaworthy steamship built up to that time. It was literally in a class by itself. After several years of successful service, an official board of officers was convened to determine the ship's future. The board decided the ship should be scrapped. The board offered several theoretical reasons for scrapping it, generally to the effect that its design differed from that of other steam ships, and, therefore, it was probably less effective. This in the face of several years' experience to show it was more effective! As Morison shows, however, these objections were only a cover. The real reason the Board recommended scrapping the *Wampanoag*—and the reason the recommendation was accepted—was that, in the view of the Board, stoking furnaces, tending boilers, and operating steam machinery were incompatible with the type of shipboard life practiced in the days of sail. In short, an effective steam warship posed the risk of changing a way of life that was viewed as an ideal.

Thus, in comparing two technologies for a possible analogy, the forecaster must take into account how each fits in with the traditions and customs of the society into which it was or will be introduced.

The cultural dimension. This dimension deals with the values, attitudes, and goals of the society into which the innovation is to be introduced. Rescher (1969) gives the following definition: "A value represents a slogan capable of providing for the rationalization of action by encapsulating a positive attitude toward a purportedly beneficial state of affairs." By "slogan," Rescher means that the value can be named with a catchword; by "rationalization," Rescher means that it relates to the justification, critique, defense, recommendation, etc., of some course of action. People have needs and desires and are capable of using reason. In particular, they are capable of using reason to achieve goals, rather than simply responding to stimuli. Values arise from this fact, since, according to Rescher, "to have a value is to be able to give reasons for motivating goal-oriented behavior in terms of benefits and costs, bringing to bear explicitly a *conception* of what is in a man's interests and what goes against his interests."

To make this idea concrete, Rescher lists some values: health, comfort, physical security, economic security, productiveness, honesty, fairness, charitableness, courtesy, freedom, justice, beauty, clearness of conscience, intelligence, and professional recognition. Different societies and people at different times and places have ranked these values in differing orders of importance.

Essentially, a person's highest values are the things he or she will give up last; the things for which the individual will sacrifice everything else, including other less-important values. The values held in common by a society are those things which the society will sacrifice other things for.

How do values and technology interact? The interaction may come about in either of two ways. First, the values subscribed to by a society may favor or inhibit change in general. Some societies may oppose change simply because it is change; others may favor change and novelty. Secondly, a proposed technological change may threaten or enhance a value that is important to the society.

As an example of a technological change affected by societal values, consider seat belts. These could have been available for automobiles any time after they were developed for aircraft. A long series of scientific investigations into auto accidents had shown, by the early 1950s, that seat belts would save lives and reduce injuries. Yet in 1956, when the Ford Motor Company offered seat belts on its cars, they were a flop. The public simply wasn't interested. Twelve years later, in 1968, seat belts were required in all cars. By the 1980s, the wearing of seat belts was mandatory, with fines for those caught without them. The technology had not changed significantly. Neither had the facts of auto accidents. The change was in the values subscribed to by members of the public.

A comparison between two technologies must take social values into account. The two societies involved in the possible analogy must be compared on the basis of the general effect of their values on the model and the present innovation. In addition, the effects of the innovations on societal values must also be taken into account. In particular, the extent to which the two innovations weaken or reinforce important societal values must be considered.

The intellectual dimension. The preceding two dimensions dealt with society at large. This dimension deals with the intellectual leaders of society, including decision makers for public and private organizations, people who speak on behalf of prestigious institutions or causes, and opinion leaders such as editors, writers, and TV reporters.

The intellectual leaders of a society may support or oppose a technological change on the basis of whether it reinforces or weakens their values and goals. The effect of their opposition or support will, however, depend upon the extent to which their leadership is actually followed by others on the particular issues being analyzed. A comparison between the model case and the current technology must then do two things. First, it must examine the extent to which intellectual leaders have a following. If the intellectual leaderships have significantly different degrees of effectiveness in the two cases, the analogy is invalidated. Second, if the intellectual leaderships have similar degrees of influence in both cases, then it is necessary to examine the direction of that influence. Is the degree of compatibility between the innovations and the values of the opinion leaders the same in both cases? Do the two technologies act in the same way on these values (to reinforce or to weaken)? Is the strength of their action about the same in both cases?

The religious-ethical dimension. Most people judge actions and events on the basis of some standard of right or wrong, and these judgments affect technological change, just as they affect many other things. This dimension has two components: (1) the beliefs that guide ethical judgments and (2) the institutions that promulgate or formulate ethical beliefs.

Beliefs may affect technological change in two ways. First, people may oppose or favor technological change on general grounds. Some religions, for instance, require their members to refrain from using modern technology such as mechanized farm equipment. Second, people may favor specific technological changes because they appear to support particular doctrines, or they may oppose technological changes because they appear to violate particular doctrines.

A recent example of a technological activity opposed on ethical grounds involves environmental standards for phosgene gas. Phosgene is well known as a war gas. It is described as a "suffocating agent." It operates by destroying enzymes in the lungs, which leads to fluid accumulating in the lungs. The victim literally drowns in his or her own lung fluids. However, phosgene is also an important industrial chemical. Currently, the U.S. chemical industry consumes over a billion pounds of phosgene every year. Because phosgene poses a health hazard, the *Environmental Protection Agency* (EPA) developed standards for phosgene emissions from chemical plants. To develop these standards, the EPA wished to know the effects on humans of various concentrations of phosgene. If the effects are underestimated, the standards (allowable levels) will be too high. People may die from accidental phosgene exposure. If the effects are overestimated, the standards will be too restrictive. People also may die from lack of the pesticides and other chemicals made with phosgene.

Hence, setting the standards correctly is important. It turned out that during World War II, the Nazis had conducted experiments with phosgene on prisoners in labor camps. The EPA contractor charged with developing data on the effects of phosgene proposed using this data, which had become part of the court records at the Nuremberg War Crimes Trials. However, many in the scientific community objected to the use of the Nazi data. Their objection was not on scientific grounds. Instead, they argued that the data were inherently tainted by their source. The consensus of the scientific community was that data collected through Nazi barbarity were unacceptable regardless of their scientific quality or the difficulty of setting standards without human exposure data.

Thus, in comparing two technologies for a possible analogy, it is important to determine whether they have the same relationship to ethical principles. Were both compatible with ethical principles? Did both violate ethical principles? Was one compatible with ethical principles and the other incompatible? The direction and strength of the impact of ethical principles on the technologies should be similar in the two cases.

Institutions may act in an organized manner to oppose or favor technological change, just as their individual members may act, and on much the same grounds. A given institution may have either a favorable or unfavorable attitude toward a particular technological change, depending upon whether the change is compatible with their beliefs, how the technologies affect their members, and how the technologies affect the institutions as institutions. As examples of the latter, churches have in general favored the printing press, radio, and television, because these made it easier for them to teach their doctrines.

Other institutions besides churches may have an impact on technological change through the ethical dimension. One class of important institutions is technical professional societies. These often establish standards of ethical practice for their members. Medical societies establish codes of ethics. Engineering societies establish codes of ethical engineering conduct. These codes may either hasten or retard technological change. For instance, some engineering societies have for years attempted to retard the growth of military technology by arguing that it is unethical for engineers to work on technology specifically intended to kill people.

In comparing two technologies for a possible analogy, it is important to determine whether they obtained the same degree of support or opposition from institutions such as churches and professional societies. In particular, it is important to identify any applicable codes of ethics, regardless of their institutional source, that might affect the attitudes the developers, supporters, or users hold regarding a particular technology. The direction and strength of the effect should be similar in both cases for the analogy to be valid.

The ecological dimension. In the past, ecological or environmental considerations were often not taken into account. From the environmental devastation of primitive slash-and-burn agriculture, through the effects of excessive logging or animal grazing, to the present-day effects of industrial pollution, environmental effects were often ignored.

This is no longer the case. Environmental effects and ecological considerations are now taken seriously. Hence, the technological forecaster must take this dimension of the environment into account. From the standpoint of the technological forecaster, it is important to recognize today's heightened environmental sensitivity. Hence, an analogy with a past technology might be invalidated simply because of this change in attitude. An environmentally damaging technology might have been accepted in the past, but an equally damaging technology might not be accepted today.

It is important to recognize that technology as such is not inherently damaging to the environment. Some technologies actually improve the environment. Despite the current bad-mouthing of automobiles by some environmentalists, the automobile has eliminated the problem of horse manure from city streets. In 1900, New York City's biggest waste disposal problem was the daily removal of horse manure. The automobile has actually helped clean up our cities.

In considering the ecological dimension, then, it is necessary to consider the detrimental or beneficial effects of the model technology and the technology to be forecast. Do they have the same kinds of effects? To the same degree?

This completes the description of the dimensions of analogies. Using these dimensions, it is possible to make a systematic comparison between two situations to determine whether a valid analogy exists. Not all the dimensions will be important in each case. However, the forecaster should always check each dimension to determine whether or not it is important. The forecaster should not fall into the practice of ignoring certain dimensions simply because they have been found to be unimportant in specific cases. Conversely, the validity of an analogy in one dimension should not be allowed to outweigh the forecaster's judgment that the dimension is not important, particularly if it causes him or her to overlook the failure of the analogy in other dimensions.

3.4 Selecting Analogous Technologies

The success of a forecast by analogy depends upon choosing technologies that are truly analogous to the one being forecast. A technology that is insufficiently similar will give no useful information. It will probably fail to be analogous in one or more of the dimensions of an analogy and, therefore, will be rejected. If the forecaster finds that all candidate analogies are rejected, there will be no basis for a forecast.

The forecaster's first inclination will often be to select a "similar" technology. That is, if a transportation technology is to be forecast, the potential analogies are selected from other transportation technologies. However, this is not always the best approach. Even though a technology performs the same function as the one to be forecast, it may not be similar enough in the important dimensions to serve as an analogy.

Thus, while the forecaster should not overlook candidate technologies that perform the same function, technologies serving other functions should be considered as well. In many cases, a technology that serves some totally different function may turn out to be highly analogous.

The late Willy Ley, a pioneer rocket experimenter in Germany during the 1930s, related an incident that illustrates this point. He described a meeting at which German rocket designers met with pump manufacturers. The rocket designers stated they needed pumps that could deliver a high flow rate of fuel, be brought up to speed quickly, and would operate reliably even though they had spent months in unattended storage. The pump manufacturers stated that these requirements were identical with those for pumps on fire engines, and, therefore, they saw no problem in meeting them. As it turned out, the pump manufacturers were able to meet the design requirements for pumps for rockets. Their experience in a totally different application area turned out to be highly relevant, because the demands placed on the technology were similar.

Krawiec (1983) describes the preparation of a forecast of solar energy technology (heliostats). The forecasting team sought out industrial and commercial components that were similar to those incorporated into heliostat designs. They found numerous items manufactured for automobiles that were sufficiently similar to heliostat components that could serve as analogous technologies. They also found that components of a radio telescope were in many cases similar enough to heliostat components to serve as analogies. Other technologies that they identified as analogous to components of a heliostat included aircraft windshields, laminated mirrors, aircraft landing gear, personal calculators, metal door sash and trim, conduit raceways, drive wheel transmissions, metal household furniture, truck and bus bodies, and aircraft wings.

These examples serve to illustrate that analogous technologies should be sought from a wide range of candidates. A search that examines only technologies performing functions similar to the technology to be forecast may miss some highly analogous candidates. An imaginative and thorough search will yield a great many candidates that at first glance seem to have no connection with the technology to be forecast, but on further examination turn out to be useful analogies.

3.5 Deviations from a Formal Analogy

Suppose that two situations were compared for a formal analogy and the requirements were found to be violated in several dimensions, but always in the same direction. That is, suppose all the deviations were such as to make the current innovation even more likely than the historical model. Even though a formal analogy does not exist, can the forecaster still predict that the current technology will follow the same pattern as the model case, but even more rapidly, since the conditions are more favorable?

Conversely, suppose that two situations are compared and it is found that on some dimensions the situation is more favorable to the current technology than it was to the model technology, but on other dimensions the situation is less favorable. Suppose, in the judgment of the forecaster, the favorable deviations outweigh the unfavorable ones? Can the forecaster use the analogy to predict that the current technology will advance at least as rapidly as the model technology did?

The forecaster cannot make such forecasts on the basis of analogy, since a formal analogy does not exist. If a forecast is made in this manner, the forecaster is actually making use of a crude form of a causal model, a topic we will take up in Chap. 8. The difficulty in this case is that there is neither the precedent of an analogy nor the rigor of a true causal model to support the forecast.

The forecaster cannot simply refuse to forecast. The historical situations, even though not valid analogies, can still be used as starting points for forecasts. Klein (1986) has shown that a combination of historical cases and expert judgment can be used effectively to make forecasts. The procedure is to identify the degree of similarity of each of the dimensions and to estimate the effect that each difference will have on the overall outcome. This approach, of dividing the problem into separate factors and estimating the influence of each factor, has been found to be more effective than trying to make an overall global judgment for the current technology. The dimensions of analogy discussed above provide a framework in which expert judgment can be used. They allow the overall task to be partitioned into smaller and more homogeneous tasks. Once this is done, it is possible to use multiple experts on each dimension. Delphi, for instance, might be used to gain the opinions of the experts on the degree to which the differences on a specific dimension will affect the overall outcome. Moreover, once the task is partitioned this way, it is possible to use different experts on different dimensions. It is not necessary to have one person who is an expert on the whole problem. Indeed, there may not be any one such "global" expert. There may well be many experts on individual pieces of the problem. Partitioning the forecasting task in this way allows the several experts to combine their disparate knowledge.

3.6 Summary

In this chapter we have discussed the use of historical analogies for forecasting technology. In cases where the forecaster wants to make use of something more rigorous than judgment alone, historical analogy may be used. If some previous situation can be found that is similar to the one to be forecast, the forecaster will predict that the outcome of the current situation will be analogous with the outcome of the model historical situation.

The use of analogies is a very common way of thinking and is the basis of inductive inference. However, there is nothing inevitable about the outcome of a situation involving people. There is no guarantee that if the circumstances are repeated in detail, the outcome will be repeated also. Nevertheless, by assuming that there is a normal or natural way for people to behave in a particular situation and that the historical model situation illustrates that natural way of behavior, the forecast can be based on historical analogy.

Since the use of analogies depends on comparing two situations in detail, it is helpful to have some systematic way of making these comparisons. For this purpose we have presented a set of dimensions on which situations can be compared. Because the primary interest here is in forecasting the progress of technology, the dimensions are based on factors that have historically demonstrated a connection with a tendency either to favor or resist innovation. In particular cases, however, other sets of dimensions might be more useful. Thus, while this set has been found to be useful in many technological forecasting tasks, the forecaster should be alert to the possibility that for the case in question, the set of dimensions may need to be modified to suit the problem.

Finally, in those cases where the forecaster cannot find any cases that are formally analogous to the one to be forecast but can find some in which the net effect of the deviations from a formal analogy is in a particular direction, a forecast can still be made. This forecast has neither the support of historical precedent nor the rigor of a causal model. It will be based on judgment about the impact of deviations from a strict analogy. However, the use of the framework provided by the incomplete analogy will allow a forecast that should be better than one obtained through the use of completely unstructured judgment.

References

Horwich, M., *Clipped Wings*, MIT Press, Cambridge, MA, 1982.

Klein, G. A., "Validity of Analogical Predictions," *Technological Forecasting & Social Change*, 30:139–148(1986).

Krawiec, F., "Concepts of Learning and Experience in Developing Solar Thermal Technologies," *Technological Forecasting & Social Change*, 24:207–246 (1983).

McDougall, W. A., *The Heavens and the Earth*, Basic Books, New York, NY, 1985.

Morison, E. E., *Men, Machines, and Modern Times*, MIT Press, Cambridge, MA, 1966.

Rescher, N., *Introduction to Value Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1969.

Problems

1 Explain the advantages and disadvantages of forecasting by the use of analogies.

2 What is "historically conditioned awareness," and what effect may it have on analogical forecasts?

3 Develop the technological dimension of a possible analogy between the Erie Canal (in upstate New York) and the Miami and Erie Canal (Ohio River to Lake Erie). Use the same factors as were utilized in the text for the Suez and Panama Canals comparison. Are the two canals analogous on this dimension?

4 Develop the economic dimension of a possible analogy between the Suez Canal and the Panama Canal. What was the market for each? How many users were there? What was the payoff for the individual user? What were the costs of construction and operation? Are these two projects analogous on this dimension?

5 Develop the managerial dimension of a possible analogy between the construction of Boulder Dam and the Golden Gate Bridge. Are these two projects analogous on the managerial dimension?

6 Develop the political dimension of a possible analogy between the (unsuccessful) development of the Supersonic Transport and the successful development of the Space Shuttle. What were the size and complexity of the two managerial tasks? Are the two projects analogous on this dimension? (See the books by Horwich, 1982, and McDougall, 1985, for information.)

7 Develop the social dimension of a possible analogy between the introduction of railroads into France and into England during the period 1820 to 1850. Are they analogous on this dimension?

8 Develop the cultural dimension of a possible analogy between the introduction of television into the United States and into the Soviet Union. Base the analysis on the cultures of the two countries at the times when television was introduced. Are they analogous on this dimension?

9 Develop the intellectual dimension of a possible analogy between the introduction of motion pictures and of television into the United States. Base the analysis on intellectual values at the times when the two technologies were introduced. Are they analogous on this dimension?

10 Develop the religious and/or ethical dimension of a possible analogy between the Methodists' and the Amish Mennonites' response to the industrial revolution and the development of factories and mass production. Are the two situations analogous?

11 Develop the ecological dimension of an analogy between the introduction of fossil-fueled plants and nuclear-generated electric power plants in the United States. Base the analysis on the 1920s for fossil-fuel plants and the 1950s for nuclear power plants.

Chapter

4

Growth Curves

"Things have their due measure; there are ultimately fixed limits, beyond which, or short of which, something must go wrong."

HORACE

Satires, I, i, 106

4.1 Introduction

The term "growth curve" represents a loose analogy between the growth in performance of a technology and the growth of a living organism. Despite the attempts of some forecasters to make the analogy rigorous, it is at best a casual one. Nevertheless, the term "growth curve" or "S-shaped curve" is descriptive and evokes an image of a pattern of technological change.

Figure 4.1 shows the growth in height and weight of an individual. The growth in height, especially, presents a pattern of rapid initial increase, followed by a slower increase, and finally an approach to a limit. Figure 4.2 illustrates the same behavior for two different technical approaches to illumination technology, the incandescent light and the fluorescent light. The similarity between the curves for organic growth and technological change is the basis for the term "growth curve."

Growth curves are used to forecast the performance of individual technical approaches to solving a problem. A particular technical approach, such as incandescent lighting, will be embodied in a succession of devices all using the same principles and often much the same materials and design. Any single technical approach is limited in its ultimate performance by chemical and physical laws that establish the maximum performance that can be obtained using a given principle of operation. Adoption of a device using a different principle of operation means a transfer to a new growth curve.

Growth curves are therefore used to forecast how and when a given technical approach will reach its upper limit. History seems to show that when a technical approach is new, growth is slow owing to initial difficulties. Once these are overcome, growth in performance is rapid. However, as the ultimate limit is approached, additional increments in performance become more difficult, and growth again becomes slow. It is often important to forecast the tim-

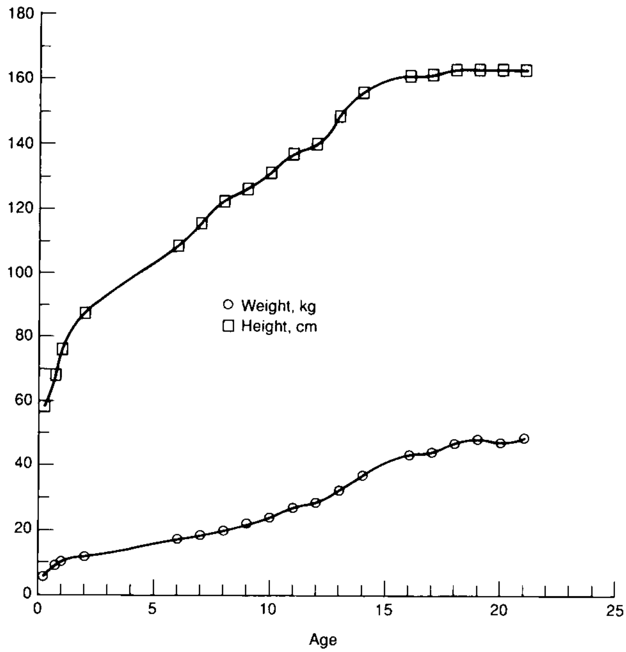

Figure 4.1 Growth in height and weight of an individual. (Data courtesy of Theresa Martino.)

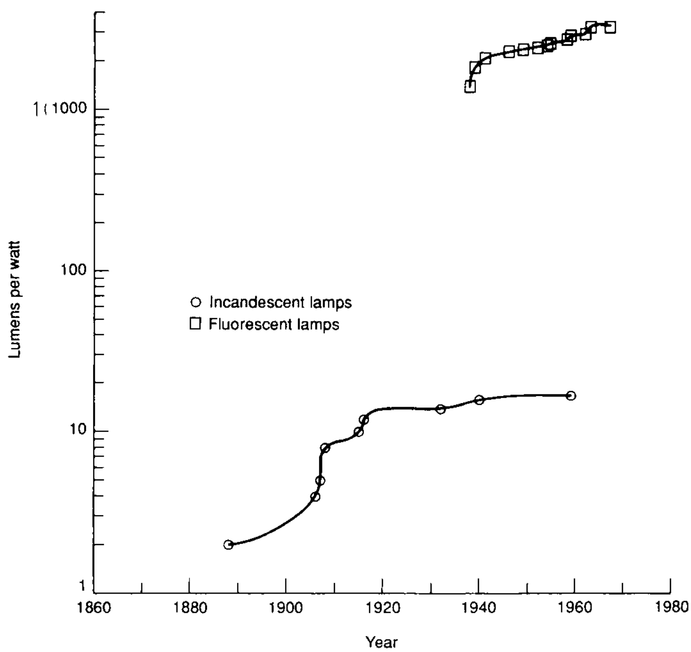

Figure 4.2 Growth in efficiency of incandescent and fluorescent lamps.

ing of these changes in growth rate, both to determine how long a given technical approach will still be competitive and to plan for the conversion to a new technical approach.

4.2 Substitution Curves

Frequently one is interested in forecasting the rate at which a new technology will be substituted for an older technology in a given application. Initially the older technology has the advantage in that it is well understood, its reliability is probably high, users have confidence in it, and both spare parts and technicians are readily available. The new technology is unknown and its reliability is uncertain; spare parts are hard to obtain and skilled technicians are scarce. Hence, the initial rate of substitution of the new technology is low. As the initial problems are solved, the rate of substitution increases. As the substitution becomes complete, however, there will remain a few applications for which the old technology is well suited. The rate of substitution slows as the older technology becomes more and more difficult to replace.

The substitution of new technology for an older one, then, often exhibits a growth curve. Behavior of this type is shown in Fig. 4.3. which shows the substitution of mechanical power for sail power in the U.S. merchant marine.

Growth curves are frequently used to forecast the substitution of one technology for another as well as for forecasting improvements in individual tech-

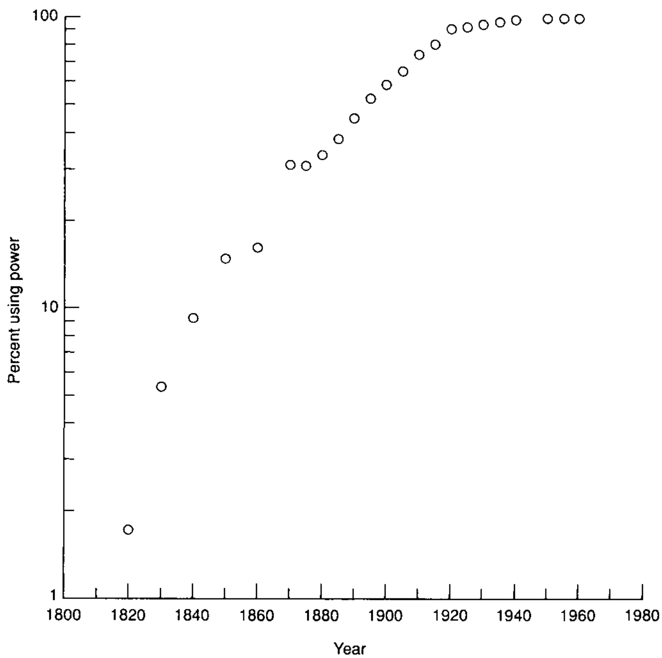

Figure 4.3 Percentage of the U.S. merchant marine using mechanical power.

nical approaches. The usual use is to forecast the share of the market or the share of total installations. If the size of the market is changing, this is usually forecast separately, i.e., growth curves for market substitution are often combined with forecasts of total use to obtain forecasts of actual market size or actual numbers of installations. Here we will focus only on forecasting the rate of substitution. Combining this with forecasts of total installations will be discussed in Chap. 11.

4.3 Using Growth Curves

Forecasting by growth curves involves fitting a growth curve to a set of data on technological performance, then extrapolating the growth curve beyond the range of the data to obtain an estimate of future performance. This activity involves three assumptions:

1. The upper limit to the growth curve is known.
2. The chosen growth curve to be fitted to the historical data is the correct one.
3. The historical data gives the coefficients of the chosen growth curve formula correctly.

The third assumption will not be considered at length. Mathematicians are agreed that in most cases, the method of "least squares" (i.e., minimizing the mean squared error between the fitted curve and the data), also known as "linear regression," is the correct way to fit a curve to a set of data. This is the curve-fitting method employed in the TEKFOR program. Hence, we will utilize linear regression as the fitting procedure and assume that the coefficients obtained by this procedure are the correct ones.

The other two assumptions are critical to use of growth curves for forecasting. We will consider them in more detail.

The first assumption, about the correctness of the upper limit, will be treated in Sec. 4.8. The choice of upper limit is independent of the choice of growth curve, hence this assumption can be dealt with separately.

Here we deal with the correct choice of a growth curve. The growth curves most frequently used by technological forecasters are the Pearl curve and the Gompertz curve. We will consider each in turn. Some other curves have been proposed by various researchers, primarily for forecasting the diffusion of innovations. We will briefly examine several of these. Before we examine specific growth curves, however, we need to take up another issue.

A growth curve is of course an attempt to forecast by extrapolation. How well does extrapolation from early data perform? Are later developments affected by external events that distort the growth curve? Empirically, the answer is that external events do seem to affect growth curves, but not by very much.

Marchetti (1983), for example, has examined the adoption of automobiles in nine different countries. He found that the rate of adoption had been fixed by the time the automobile reached a market penetration of only 1 percent and did not change significantly thereafter (although it was different in each of the countries).

Sanford (1983) studied the adoption of new types of nuclear particle accelerators and of new instrumentation in the field of nuclear physics. He found that the rate of adoption of a new technique (accelerator or instrument) was fixed early in its history. The rate of decline was also fixed once the decline had started, regardless of the number and timing of later innovations that replaced the one being studied.

These studies seem to say that for the time periods under consideration in technological forecasting, the factors that determine the growth of a technology seem to remain fairly well fixed. The result is that early data for a growth curve do seem to predict later developments well, if the right curve has been chosen and the upper limit is estimated correctly.

4.4 The Pearl Curve

This curve is named after the U.S. demographer Raymond Pearl, who popularized its use for population forecasting. It is also known as the "logistic curve." A variant formula for it is known as the "Fisher-Pry curve."

The formula for the Pearl curve is:

$$y = \frac{L}{1 + ae^{-bt}} \tag{4.1}$$

In this equation L is the upper limit to the growth of the variable y, e is the base of the natural logarithms, t is time, and a and b are the coefficients obtained by fitting the curve to the data. The curve has an initial value of zero at time $t = -\infty$ and a value of L at $t = +\infty$. [Note that if the initial value of a technology is not zero, its initial value can be added as a constant to the right-hand side of Eq. (4.1.)] The inflection point of this curve occurs at $t = (\ln a)/b$, when $y = L/2$. The curve is symmetrical about this point, with the upper half being a reflection of the lower half.

The Pearl curve is unusual among growth curves in that the shape (steepness) and location can be controlled independently. Changes in the coefficient a affect the location only; they do not alter the shape. Conversely, changes in the coefficient b affect the shape only; they do not alter the location. This property makes the Pearl curve useful in a variety of mathematical applications besides forecasting.

In fitting a Pearl curve to a set of data, it is customary to "straighten out" the curve first, i.e., the formula is transformed into that of a straight line, and the best straight line is then fitted to the transformed data. Let the transformed variable be Y. Then the transformation is:

$$Y = \ln\left[\frac{y}{L - y}\right] = -\ln a + bt \tag{4.2}$$

where the right-hand side is the equation of a straight line. The transformation is performed by taking the natural logarithm of the data value divided by the difference between the data value and the upper limit. The transformed variable is then a linear function of time t, where the constant term is ln a and the slope is b. The coefficient b is taken to be intrinsically positive, so the line representing the transformed variable Y slopes up and to the right. Once a and b are obtained from a regression analysis, they can be substituted back into the Pearl curve formula. The formula can then be extrapolated to future values of time by substituting the appropriate value for t.

Equation (4.2) involves the use of the base of the natural logarithms. However, the formula for a base-10 Pearl curve can also be given as follows:

$$y = \frac{L}{1 + 10^{A-Bt}} \tag{4.3}$$

Here again, L is the upper limit of the variable y, and t is time. The coefficients A and B have been capitalized to indicate that they have the same roles as a and b in the standard form but have different numerical values. The coefficient A controls the location of the Pearl curve, while B controls the shape.

With an algebraic manipulation that is similar to that applied to the standard version of the Pearl curve, the following straight-line equation is obtained:

$$Y = \log\left[\frac{y}{L - y}\right] = -A + Bt \tag{4.4}$$

The advantage of this formula for the Pearl curve is that the ratio $y/(L - y)$ forms a straight line when plotted against t on semilog paper. Of course, we do not have to express the Pearl curve in this form. All that is necessary is to transform the data as in Eq. (4.4). Not only is it then suited for linear regression, it can also be plotted directly on semilog paper.

Note that the ratio $y/(L - y)$ increases by a factor of ten for every increase in t of $1/B$. This is the "ten-folding time" or "T-time," and it is characteristic of a base-10 Pearl curve with shape coefficient B. It takes four T-times for the ratio $y/(L - y)$ to go from 0.01 to 100 or for the ratio y/L to go from 0.0099 to 0.990099. A base-10 Pearl curve can be characterized in terms of its T-time and the value of t at which the ratio $y/(L - y)$ achieves some specified value such as 1.0.

It takes two T-times for the ratio $y/(L - y)$ to go from 0.1 to 10 or for y to go from approximately 10 percent of L to approximately 90 percent of L. This is sometimes referred to as the "takeover time" for the substitution and is of course equal to $2/B$. A specific substitution is sometimes characterized in terms of its takeover time instead of its T-time.

Fisher and Pry (1971) developed an equation for a growth curve that seemed to fit a great many cases of technological substitution. Their equation was:

$$f = \tfrac{1}{2}[1 + \tanh a(t - t_0)] \tag{4.5}$$

In this equation, f is the fraction of applications in which the new technology has been substituted for the old, t_0 is the time for 50-percent substitution, the coefficient a is the shape coefficient for the curve, and tanh is the hyperbolic tangent.

Fisher and Pry found that by plotting the ratio $f/(1 - f)$ on semilog paper, the substitution curves for some 17 different cases fell very close to straight lines. What was not widely realized at the time was that their formula was simply an alternate version of the Pearl curve. It can be converted into the standard form by algebraic manipulation. In effect, however, Fisher and Pry were the first to use the base-10 Pearl curve as a substitution curve. Thus, the Fisher-Pry curve can be characterized in terms of its T-time and the time for 50-percent substitution. The ratio $f/(1 - f)$ is usually referred to as the Fisher-Pry ratio.

It is now quite common to use the Fisher-Pry transformation to plot substitution curves: The fraction of market share captured divided by the fraction not yet captured is plotted versus time on semilog paper. A straight line is fitted to the plot and projected to forecast the future levels of substitution. An example of this is shown below in Sec. 4.10.

Fisher-Pry curves can be used to illustrate successive substitutions. Marchetti (1977) provides examples showing the successive use of different energy sources.

4.5 The Gompertz Curve

This curve is named after Benjamin Gompertz, an English actuary and mathematician. The equation for the Gompertz curve is:

$$y = Le^{-be^{-kt}} \tag{4.6}$$

where y is the variable representing performance, L is the upper limit, e the base of the natural logarithms, and b and k are coefficients to be obtained from fitting the curve to a set of data.

Like the Pearl curve, the Gompertz curve ranges from zero at $-\infty$ to L at $+\infty$. However, the curve is not symmetrical. The inflection point occurs at $t = (\ln b)/k$, where $y = L/e$. Just as with the Pearl curve, it is necessary to straighten out the Gompertz curve before linear regression can be used to obtain the coefficients b and k. This is done by taking the logarithm twice, obtaining the linear equation:

$$Y = \ln[\ln(L/y)] = \ln b - kt \tag{4.7}$$

When Y is regressed on t, the constant term is $\ln b$ and the slope term is k. As with the Pearl curve, the slope term is taken to be intrinsically positive.

The straight line obtained by this transformation slopes down and to the right. Once b and k are obtained from the regression, they can be substituted in the formula. Future values of t may then be substituted into the formula to obtain forecasts of the variable y.

A convenient way of plotting the Gompertz curve is to take the logarithm of the ratio L/y and plot this logarithm on semilog paper. If the data obey a Gompertz curve, the data points will then fall on a straight line.

4.6 Comparison of the Pearl and Gompertz Curves

As pointed out above, one of the critical assumptions in the use of growth curves for forecasting is that the curve fitted to the data is the correct one. What does this mean? It means that the chosen growth curve must match the dynamics of the growth of the technology, so that the behavior of the curve when extrapolated outside the range of the data will match the future behavior of the technology.

Many computerized curve-fitting packages allow the user to fit several curves and provide a "goodness of fit" measure for each curve. The user is then expected to select the curve that best fits the data. This procedure is appropriate when the purpose of the curve is to represent the data within the range of the data. It is not appropriate when the purpose is to extrapolate beyond the range of the data. In the latter case, we want the curve that will best represent future data. This is not necessarily the one that best represents past data. It is, thus, very bad practice to fit several curves, select the curve that has the best goodness of fit measure, and extrapolate from that one.

Instead, the forecaster should select the growth curve that best matches the underlying dynamics of the process which not only produced the past data but will produce the future behavior of the technology. The Pearl and Gompertz curves have completely different underlying dynamics.

Table 4.1 compares the two curves. Note the slope for each of the curves as given in the table. The slope of the Pearl curve involves both y and $(L - y)$, i.e., distance already come and distance yet to go to the upper limit. On the other hand, for large values of y, the slope of the Gompertz curve involves only $(L - y)$, i.e., the Gompertz curve is a function only of distance to go to the upper limit.

TABLE 4.1 Comparison of Pearl and Gompertz Curves

Curve	Equation		Slope
Pearl	$Y = \dfrac{L}{1 + ae^{-bt}}$		$\dfrac{by(L - y)}{L}$
Gompertz	$y = Le^{-be^{-kt}}$	for all values of y	$-bky\ln(y/L)$
		approximation for $y \geq L/2$	$bk(L - y)$

Consider the growth in performance of a new technical approach. Clearly progress will be harder to achieve the closer the upper limit is approached. But is there any offsetting factor by which progress already achieved makes additional progress easier? If there is such an offsetting factor, then progress is a function of both distance to go and distance already come, and the Pearl curve is the appropriate choice for forecasting future progress. If there is not any such offsetting factor, progress is a function only of distance to go, and the Gompertz curve is the appropriate choice.

An offsetting factor, such that past progress makes future progress easier, usually arises from some past improvement whose full potential has not yet been exhausted. If that past improvement can interact synergistically with further improvements so that the combination produces a greater effect than each improvement individually, then past progress makes future progress easier. In such a case, the dynamics of future change are best described by the Pearl curve.

In the case of a new technology substituting for an old one, a choice must again be made of the appropriate curve for estimating future progress. At the time of introduction of the new technology, there may be few suppliers, few repair facilities, and little information available to would-be adopters. As adoption progresses, however, suppliers and repair facilities become more widespread. Information about the new technology becomes more widely available. Thus, increasing substitution makes further substitution easier. On the other hand, substitution will take place in the easiest applications first. The new technology will be adopted for the more difficult applications only later. Thus, substitution becomes more difficult as substitution progresses. The rate of substitution, then, depends on both the extent of substitution that has already occurred and the extent of the remaining old technology. It involves both distance already come and distance yet to go. That is why the Pearl curve is so successful at forecasting substitutions.

However, in some cases there may not be a beneficial effect from prior adoption of the new technology. In such cases, the Gompertz curve would be appropriate.

Buzbee and Sharp (1985), for example, utilized the Gompertz curve to forecast the performance of supercomputers. Although they did not discuss this point, they evidently concluded that continued progress in speed of supercomputers would be increasingly more difficult to achieve, with no offsetting factors from progress already achieved.

To recapitulate, the correct way to choose a growth curve is to determine the underlying dynamics of the situation. This requires some examination of the factors that hasten or retard growth in the performance or adoption of the technology. When the underlying dynamics have been determined, the appropriate growth curve can be readily selected. The choice of growth curve must be made on the basis of the factors driving the change in technology, not on the basis of mathematical goodness of fit to historical data.

4.7 Other Growth Curves

Numerous researchers have attempted to identify other growth curves that might be more suitable for forecasting than either the Gompertz or Pearl curve. One of the factors that affects the adoption or diffusion of an innovation is "imitation," i.e., one person adopting because he or she observes the success of another person who has already adopted. Some diffusion models explicitly include an imitation coefficient. In most of these models, the imitation coefficient is a constant. However, Easingwood, Mahajan, and Muller (1981) developed a diffusion model that allows the imitation coefficient to change over time. Their reasoning was that the holdouts are less likely to adopt than are those who adopted early; hence, the imitation coefficient should decrease over time. In other cases, however, the later adopters have more opportunity to become informed about the technology; hence, once they learn of it, they may adopt more rapidly than did those who first became informed of it. Thus, the imitation coefficient may increase with time. Their model is of the form:

$$\frac{df}{dt} = bf^d(F - f) \tag{4.8}$$

where f is the number of adopters, F is the maximum number of possible adopters (upper limit), the exponent d is a constant, and b is the time-varying imitation coefficient. In two examples, they found that b varied from about 0.5 at the beginning of the diffusion to near 1.0 at the end. However, they provide no way to forecast the value of b.

Breitschneider and Mahajan (1980) developed another approach to diffusion models with time-varying coefficients. In this approach, they make a series of one-step-ahead forecasts. At the beginning of each forecast period, the imitation coefficient is recomputed on the basis of the forecast error made during the preceding period. Thus, changes in the imitation coefficient need not be monotonic. However, several-period-ahead forecasts are essentially constant-coefficient forecasts, since the imitation coefficient remains fixed for all the periods included in the forecast.

Although there have been numerous efforts to develop more sophisticated growth curves, so far the efforts have borne little fruit. By including several adjustable coefficients, the researcher can get a model that fits past data better than either the Pearl or Gompertz curves do. However, no progress has yet been made toward determining what the coefficients should be in order to forecast future data. Without some way of choosing the coefficients, for instance forecasting the trajectory of a variable imitation coefficient, these more sophisticated growth curves cannot be applied in practice. They remain theoretical exercises.

4.8 Estimating the Upper Limit

Some curve-fitting packages not only allow the user to fit the coefficients of a growth curve to a set of data but allow fitting of the upper limit as well. As indicated earlier, choosing a growth curve on the basis of goodness of fit is bad practice. Extracting an upper limit from historical data is even worse practice. During the early history of a technical approach, the upper limit has very little effect on its growth in performance. Thus, data points from this period contain little information about the upper limit. Values for the upper limit extracted from such data are certain to have a large error component.

Do errors in the estimate of the upper limit matter to a forecast? In fact, they can have significant effects on the forecast produced by fitting a growth curve to a set of data. An example of this is shown in Tables 4.2, 4.3, and 4.4. A Pearl curve was fitted to the data on steam engine efficiencies derived from Table A.13 (see App. A). Five different upper limits for steam engine efficiency were used: 45, 47.5, 50, 52.5, and 55 percent. Table 4.2 shows the variations in a and b that result from the different estimates of the upper limit. The percent variation in a is much larger than the percent variation in the upper limit. The percent variation in b is smaller than the percent variation in the upper limit but is still significant. Table 4.3 shows the year in which the growth curve achieved a value equal to half the upper limit for each of the estimated upper limits. Table 4.4 shows fitted versus actual values for the historical time period and projections beyond the historical data. If the upper limit is underestimated, both a and b are too large; as a result, the curve rises too steeply and reaches the midpoint too soon. If the upper limit is overestimated, the curve rises too slowly and reaches its midpoint too late.

It is very important, then, to estimate the upper limit correctly. Even a small error in the upper limit can result in a fairly significant error in the forecast.

Properly, the upper limit should be based on the physical and chemical limits that are imposed by nature on the particular technical approach being forecast. These natural limits may exist in the form of a breakdown voltage, a maximum efficiency, a limiting mechanical strength, a maximum optical resolution, a minimum detectable concentration of a chemical, a minimum signal-to-noise ratio, or some similar limit. For instance, the upper limit on steam engine efficiency should be determined from the Carnot cycle and the possible boiler and exhaust temperatures. Estimating these natural limits often requires a detailed knowledge of the technical approach in question and is best done by the forecaster in conjunction with a specialist in the field.

TABLE 4.2 Pearl Curve Coefficient versus Several Upper Limits

Upper limit, %	Percent change	a	Percent change	b	Percent change
45		9.640×10^{16}		0.02057	
	5.5		64.0		2.9
47.5		3.468×10^{16}		0.01997	
	5.26		52.9		2.3
50		1.630×10^{16}		0.01951	
	5.0		44.9		1.8
52.5		0.8989×10^{16}		0.01915	
	4.8		43.5		1.9
55		0.5079×10^{16}		0.01879	

TABLE 4.3 Midpoint Year of the Growth Curve for Various Upper Limits

Upper limit, %	Midpoint year
45	1899.89
47.5	1906.93
50	1913.16
52.5	1918.81
55	1924.98

TABLE 4.4 Actual Percent Efficiency of Steam Engines and Fitted Values for Various Upper Limits

		Upper limit, %				
Year	Actual value, %	45	47.5	50	52.5	55
1698	0.5	0.696	0.721	0.740	0.755	0.770
1712	1.3	0.923	0.949	0.968	0.983	0.997
1770	2.8	2.908	2.895	2.884	2.874	2.863
1796	4.1	4.748	4.673	4.614	4.566	4.519
1830	10.0	8.410	8.410	8.242	8.108	7.980
1846	15.0	11.165	10.853	10.620	10.436	10.260
1890	18.0	20.219	19.773	19.444	19.189	18.946
1910	20.0	24.832	24.478	24.228	24.042	23.154
1950	30.0	33.170	33.378	33.617	33.863	34.154
1960		34.875	35.277	35.690	36.095	36.563
1970		36.398	37.000	37.597	38.173	38.832
1980		37.739	38.543	39.326	40.079	40.938

The question to be asked by the forecaster is, "What physical, chemical, or other considerations set limits on the best that can be achieved by *this* specific technical approach?" It is important not to confuse this question with that of what might be achieved by some different technical approach that might be adopted at some other time. Once these considerations are identified, their effects can be calculated and the ultimate limits determined. Calculations of this kind are illustrated in Breedlove and Trammell (1970) and in Keyes (1973). Buzbee and Sharp (1985) estimated an upper limit to single-processor computer speeds of about 3 billion operations per second, based on the physical limit set by the speed of light in transmitting a signal from one end of the computer to the other.

To illustrate the process of computing an upper limit to a growth curve from fundamental physical principles, we will estimate the upper limit on efficiency (lumens/watt) of tungsten-filament incandescent lights.

The lumen, a measure of light output from a light source, is not directly a physical quantity. Instead, it is based on the sensitivity of the human eye, which varies with wavelength. The sensitivity of the eye is depicted in Fig. 4.4. Radiation outside the range of roughly 400 to 700 nanometers is invisible. Even at the edges of the visible range, however, the sensitivity of the eye is

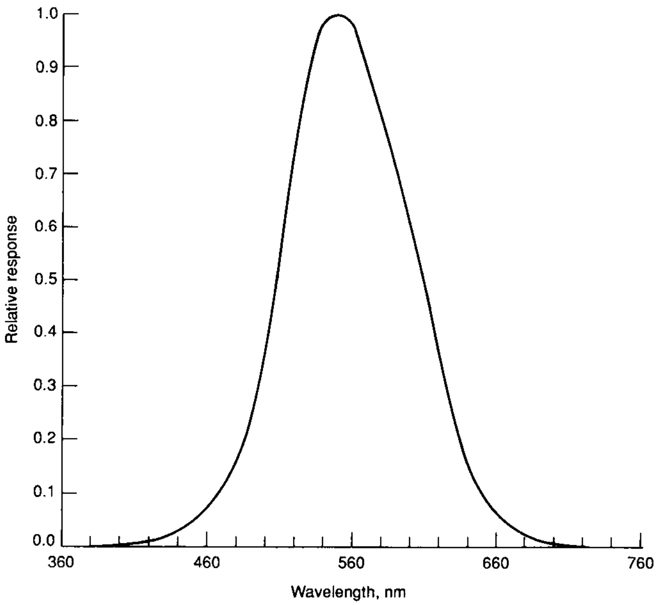

Figure 4.4 Relative response of the human eye to wavelength.

low. Maximum sensitivity of the eye occurs at about 555 nanometers, corresponding to a greenish color. To obtain the lumens output from some light source, it is necessary to determine the output in some physical measure (e.g., watts) as a function of wavelength, then modify this by the relative sensitivity of the human eye. The eye sensitivity curve will be designated as $S(L)$ where L is wavelength in nanometers.

We will treat the filament of an incandescent light as a "black body" that radiates according to Planck's Law:

$$E(L, T) = \frac{n^2 c_1 L^{-5}}{e^{(c_2/LT)}} - 1 \qquad (4.9)$$

E = the energy emitted in watts per square meter of radiating surface, radiated into a hemisphere
T = absolute temperature in Kelvin
n = the index of refraction (3 for tungsten)
L = wavelength in nanometers
c_1 = 3.7418×10^{29} W/m^2
c_2 = 1.4388×10^{7} nanometer · degrees

This expression has a peak value at a wavelength that depends upon the temperature T: The higher the temperature, the shorter is the wavelength at which the peak occurs. From the standpoint of efficiency of illumination, it would be desirable to have the peak occur inside the visible range. Unfortunately, all known materials melt before they reach the necessary temperature to place the peak of the radiation intensity within the visible range.

It might seem that the upper limit on efficiency of a tungsten filament is set by the melting point of tungsten. Clearly a filament of the type conventionally used in incandescent lamps cannot be operated above the melting point of the material of which it is made. However, there is another limitation that must be taken into account. At temperatures near the melting point, the filament material will evaporate rapidly. This has two effects, both undesirable. The first is to shorten lamp life, as the filament "evaporates away." The second is to reduce light output, as the filament material condenses on the (relatively) cool inner surface of the bulb and darkens it. To maintain high light output and reasonable lamp life, tungsten filament temperatures are limited to about 2950 K.

A plot of radiation intensity for tungsten at 2950 K is shown in Fig. 4.5. This plot shows the power radiated into *both* hemispheres, and also takes into account the fact that tungsten has an emissivity of 0.35 (i.e., it radiates only 35 percent as much energy as does a black body at the same temperature). Tungsten at 2950 K has significant radiation from about 100 nanometers to about 2.5×10^6 nanometers. However, the wavelength for peak intensity falls at a wavelength longer than that visible to humans. Most of the radiation from an incandescent bulb occurs in the infrared, with some in the near ultraviolet, and the remainder in the visible range.

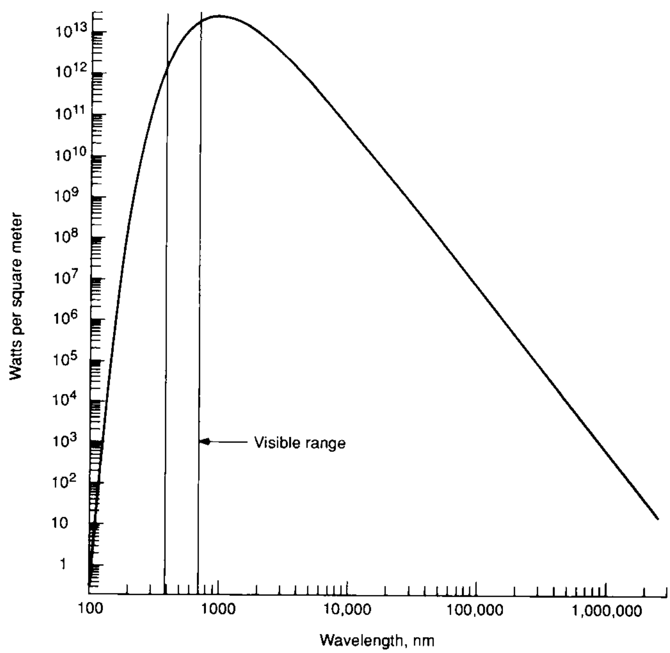

Figure 4.5 Black body radiation at 2950 K.

The total energy input (and also output), $E_{\pm}$, *from tungsten heated to 2950 K is then given by:*

$$E_{\pm} \text{ (watts/meter}^2\text{)} = 0.35 \times 2 \times \int_{100}^{2,500,000} E(L, T)\, dL \tag{4.10}$$

The light output visible to human eyes, Out, is given by:

$$\text{Out (lumens/meter}^2\text{)} = 0.35 \times 2 \times 683 \int_{400}^{700} S(L)E(L, T)\, dL \tag{4.11}$$

where 683 is the conversion factor from watts to lumens at 555 nanometers (wavelength for peak eye sensitivity).

Carrying out these integrations for T = 2950 K and taking the ratio of output to input, we find that the upper limit to efficiency of a tungsten-filament incandescent lamp is 18.915 lm/W or (rounded) 19 lm/W. This efficiency can be increased only by a new technical approach, such as fluorescent lamps (but see Prob. 2 regarding halogen-filled incandescent lamps).

This example illustrates the general concept that the upper limit to performance of a specific technical approach should be computed from fundamental physical and chemical principles. The exact nature of the computations must be specific to the physical and chemical principles that form the basis for operation of the devices. However, the general idea remains the same in all cases.

Even though upper limits should be computed on the basis of fundamental physical and chemical limits to a specific technical approach, this may not always be possible. Figure 4.6 shows the historic progress in power density (horsepower per cubic inch of piston displacement) for reciprocating aircraft engines. After rapid growth for about three decades, the performance level seems to have reached a plateau in the late 1930s. It is not just that these engines were displaced by jet engines in the 1950s. Progress on this particular measure of performance had essentially reached a halt two decades earlier. How could this upper limit be computed? Clearly it involves such things as strength of materials, ability to cool the engine, energy content of aviation gasoline, and similar factors. However, it also involves economic considerations such as engine life. The power density of an engine can always be increased at the expense of engine life, simply by operating the engine at a higher rotation rate. For instance, racing motorcycle engines typically operate at speeds over 20,000 r/min. However, these engines have a much shorter life than engines in motorcycles that are sold commercially. Hence, the plateau apparent in the figure involves not just technical considerations, but a deliberate design choice that includes economic considerations. In situations such as this, the technology may need to be characterized by multiple parameters rather than by a single one, and growth curves as discussed in this chapter cannot be used. This topic will be addressed in more detail in Chap. 6.

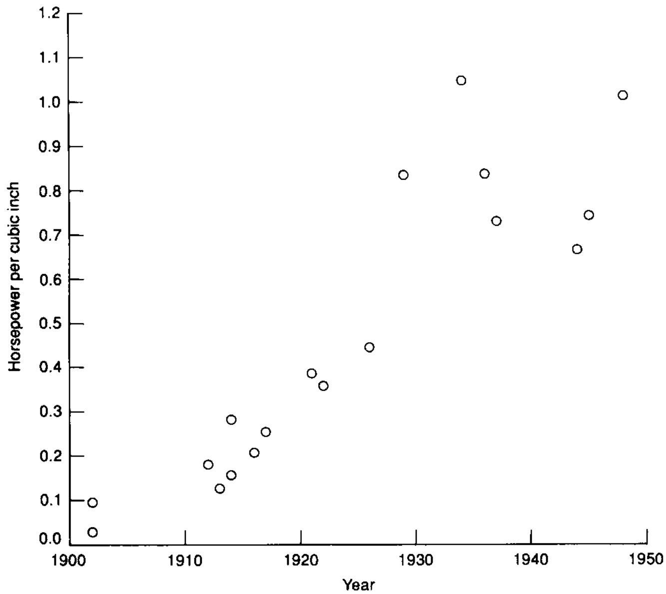

Figure 4.6 Power density of reciprocating aircraft engines.

4.9 Selecting Variables for Substitution Curves

When one technology is replacing another, the upper limit to the substitution is usually 100 percent. In some cases, there is a fraction of the applications for which the new technology is obviously unsuitable, and substitution will not take place. In these cases, the upper limit is clearly 100 percent of those applications for which the technology is suitable. For instance, Fisher and Pry (1971) examined the substitution of synthetic detergents for soap. It was clear that this substitution was impractical in applications such as hand or bath soap. Hence, they took as the proper applications only laundry and dish soaps and examined the history of substitution in only those uses. (Note, however, that in recent years synthetic hand and face soaps have appeared on the market.)

While determining the upper limit for a substitution may be comparatively easy, measuring the degree of substitution is often difficult. For instance, if cost is what is driving the substitution, dollar sales should not be the measure. This would underestimate the degree of substitution of the cheaper new technology for the more expensive old one. If one technology is "denser" than the other, the measure of substitution should not be weight (pounds or tons). For instance, in forecasting the substitution of nuclear-generated for fossil-fueled electric power plants, we would not use pounds of uranium substituted for pounds of coal as the measure of substitution. Instead, the proper measure would be in some units of output or capacity, such as the number of kilowatt · hours generated or the kilowatts of installed capacity. (Each of these would be appropriate for specific purposes.)

The basic principle is to identify some function that is common to the two technologies and choose a variable that measures the amount of this function that each technology performs. For instance, Fisher and Pry (1971) examined the substitution of plastic tile flooring for wood flooring. Since plastic is much lighter than wood, weight would not have been appropriate as a measure of substitution. Instead, they used the number of square feet of each type of flooring installed annually. Choosing a variable in this manner provides a meaningful measure of the degree of substitution, making the resulting substitution forecast more useful.

Even in choosing the measure of substitution, however, the forecaster can encounter data problems. One current example is the substitution of plastics for metal in automobiles. At least one reason for the substitution is lighter weight (another is greater corrosion resistance). The proper measure of the substitution is the number of pounds of metal in an automobile that have been replaced by plastic. Unfortunately, this information is not available. No automobile designer first designs the car in metal, then redesigns it in plastic to see how much metal can be replaced. Instead, the car is designed with plastic components right from the outset. Hence, it is fairly easy to obtain data on the number of pounds of plastic utilized in the average automobile, but impossible to obtain data on the pounds of metal that *would have* been used had the car not been designed to use plastic. Thus, the data on pounds of plastic per automobile underestimate the degree of substitution. However, this is the only data available. Correcting for the relative densities of plastic and metal will overestimate the degree of substitution, since plastic parts tend to be larger (although lighter) than the metal parts they replace. The forecaster can, thus, put upper and lower bounds on the degree of substitution in this case but cannot get a precise measure of the actual substitution.

4.10 An Example of a Forecast

Let us forecast the adoption of *cable television* (CATV) in U.S. households. Table A.42 presents the percentage of U.S. homes subscribing to cable TV. Adoption of this technology amounts to substituting it for over-the-air TV, hence a growth curve is the appropriate forecasting means.

What is the upper limit of adoption of this technology? Clearly homes without TV cannot adopt cable. However, virtually every U.S. household has at least one TV. Indeed, more households have TV than have coffee-makers or indoor plumbing. Hence, for all practical purposes, the upper limit to adoption is 100 percent.

What is the appropriate form of the growth curve, Pearl or Gompertz? We must determine whether the rate of adoption of cable TV depends only on the portion of households not yet reached, or whether it also depends upon the portion already reached.

Cable TV was first adopted in those areas where it gave the greatest advantage: fringe reception areas, small towns, and suburbs where only a limited number of TV stations could be received, or where reception was poor.

Households that could receive several stations clearly over-the-air had less incentive to adopt cable TV. Hence, clearly the rate of adoption will be influenced negatively since the remaining "holdouts" are those for whom it has least value.

However, the more widely adopted cable TV is, the more attractive it will be. More and more potential customers will be within its reach (i.e., their neighbors already have it). The larger market will in turn attract additional programming, further increasing its attractiveness. The more people who adopt cable TV, the more the holdouts will be missing. Hence, the adoption rate is influenced positively by the number of households that have already adopted it.

The Pearl curve is thus the appropriate forecasting method. Hence, we will fit a Pearl curve to the data, and extrapolate it to produce a forecast. We use TEKFOR to carry out the forecast. The first step in using TEKFOR to fit a curve is to create a data file. The data file for cable TV data is included on the TEKFOR disk (CATV.DTA). Next the curve-fitting option is chosen, and the Pearl curve is selected. The upper limit of 100 is entered when requested, as are upper and lower confidence bounds. The fit is then carried out by the program.

The results are shown in detail in Table 4.5, which presents the TEKFOR output showing the upper and lower 50 percent confidence limits as well as the numerical values of the fitted curve for various years. The results are also plotted in Fig. 4.7 as a Fisher-Pry plot. The central line represents the Pearl curve fitted to the data and projected to the year 1989. The lower and upper lines are the confidence bounds. According to the projection, CATV will have been adopted by 69 percent of U.S. homes by 1989. This is actually an over-

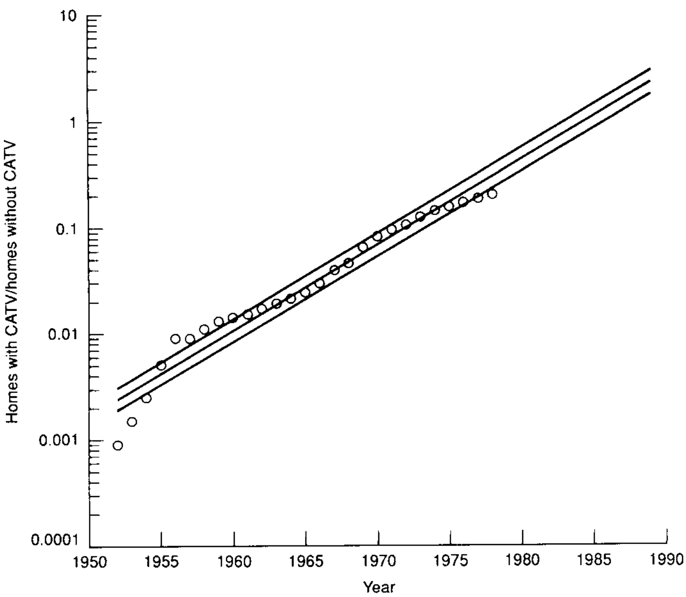

Figure 4.7 Fisher-Pry plot of adoption of CATV in U.S. homes.

TABLE 4.5 TEKFOR Output for Growth of Cable TV

RESULTS OF PEARL CURVE FIT TO CATV.DTA
A = 4.0272839506105E + 159
B = .18518873496506
CORRELATION COEFFICIENT = -.97542235565543
STANDARD ERROR OF B = .0083666570115895

X	Y	BEST FIT	LOWER CONF	UPPER CONF
1952	.09	.24341723803	.19017181960	.31152408976
1953	.15	.29279439404	.22915241696	.37404524709
1954	.25	.35215234174	.27606714344	.44911248526
1955	.5	.42349276249	.33251326544	.53923062941
1956	.9	.50921175565	.40040347683	.64739611960
1957	.9	.61217442831	.48202585970	.77718910750
1958	1.1	.73580219346	.58011429127	.93288070692
1959	1.3	.88417430534	.69793070456	1.1195568996
1960	1.4	1.0621450334	.83936060761	1.3432602381
1961	1.5	1.2754775523	1.0090231498	1.6111497975
1962	1.7	1.5309950032	1.2123967318	1.9316786551
1963	1.9	1.8367481300	1.4559605995	2.3147863230
1964	2.1	2.2021972335	1.7473519261	2.7721007710
1965	2.4	2.6384036959	2.0955364124	3.3171406644
1966	2.9	3.1582227501	2.5109882374	3.9655028782
1967	3.8	3.7764842107	3.0058720331	4.7350129635
1968	4.4	4.5101413040	3.5942152014	5.6458068817
1969	6.1	5.3783593721	4.2920530969	6.7203011732
1970	7.7	6.4025062243	5.1175222457	7.9829965769
1971	8.8	7.6059948768	6.0908679155	9.4600486953
1972	9.7	9.0139187912	7.2343225044	11.178531750
1973	11.3	10.652412094	8.5718014942	13.165322606
1974	12.7	12.547666684	10.128356248	15.445548561
1975	13.8	14.724550193	11.929321066	18.040581134
1976	14.8	17.204799827	13.999100269	20.965625290
1977	16.1	20.004822995	16.359565276	24.227050047
1978	17.1	23.133218180	19.028076843	27.819723496
1980		30.355484447	25.322401144	35.907923770
1981		34.406344150	28.939650254	40.319763845
1982		38.697413420	32.843860320	44.896756442
1983		43.171423055	36.997970166	49.564581304
1984		47.759705828	41.351358106	54.242661812
1985		52.386109353	45.841666426	58.849621858
1986		56.971921825	50.398020273	63.308864669
1987		61.441155638	54.945340760	67.553478415
1988		65.725450279	59.409219480	71.529832228
1989		69.767944791	63.720695880	75.199518787

UPPER LIMIT = 100

estimate, as can be seen by comparing Table 4.5 with the actual data in Table A.42. Part of the problem is that the slope of the early data may be misleadingly high. A better forecast might be obtained by fitting a Fisher-Pry curve to data for 1956 through 1978 (see Prob. 12).

This example has briefly illustrated the use of growth curves in forecasting. Fitting the growth curve to the data is actually the smallest part of the task and is readily computerized. The main task of the forecaster is to assure that the right data have been collected and that they have been processed properly. This means determining what variable is the best measure of the technology to be forecast, determining the appropriate upper limit, and selecting the most suitable form of growth curve. The forecaster's major contributions to the quality of the forecast have all been completed before computerized curve fitting is ever undertaken.

4.11 Summary

There is an inherent upper limit to the performance of any technical approach. Likewise, there is an upper limit to the degree of substitution of a new technology for an old. The proper tool for forecasting either performance or substitution is a growth curve. The two most common growth curves used for this purpose are the Pearl and Gompertz curves. The choice of the proper curve must be made on the basis of the dynamics governing improvement in performance or substitution. The upper limit to a performance growth curve must be based on the physical and chemical limits of the technical approach. The upper limit to a substitution curve must be based on the share of the market open to substitution by the new technology.

References

Breedlove, J. R. and G. T. Trammell, "Molecular Microscopy: Fundamental Limitations," *Science,* **170**:1310–1312 (1970).

Breitschneider, S. I. and V. Mahajan, "Adaptive Technological Substitution Models," *Technological Forecasting & Social Change,* **18**:129–139 (1980).

Buzbee, B. L. and D. H. Sharp, "Perspectives on Supercomputing," *Science,* **8**:591–597 (February 1985).

Easingwood, C., V. Mahajan, and E. Muller, "A Nonsymmetric Responding Logistic Model for Forecasting Technological Substitution," *Technological Forecasting & Social Change,* **20**:199–213 (1981).

Fisher, J. C. and R. Pry, "A Simple Substitution Model of Technological Change," *Technological Forecasting & Social Change,* **3**:75–88 (1971).

Keyes, R. W., "Physical Limitations in Digital Electronic Circuits," *IBM Report RC4572,* IBM Corp. Armonk, NY, October 11, 1973.

Marchetti, C., "Primary Energy Substitution Models: On the Interaction Between Energy and Society," *Technological Forecasting & Social Change,* **10**:345–356 (1977).

Marchetti, C., "The Automobile in a System Context," *Technological Forecasting & Social Change,* **23**:3–23 (1983).

Sanford, T. W. L., "Trends in Experimental High-Energy Physics," *Technological Forecasting & Social Change,* **23**:25–40 (1983).

Problems

1 Plot a Gompertz and Pearl curve on the same graph. Select their coefficients so that they have the same upper limit and their inflection points occur at the same value of time. What are the major differences between the two curves?

2 An improvement on the incandescent lamp is to fill the bulb with a halogen gas. This gas combines with the evaporated tungsten and returns the tungsten atoms to the filament, where they are freed from the halogen by the filament heat. It is, thus, possible to operate the filament at a higher temperature, while still retaining acceptable life and reducing darkening of the bulb. The higher filament temperature increases the lumens/watt output, since the peak of the intensity curve is moved closer to the visible range. Compute the theoretical upper limit on lumens/watt from halogen lamps.

For the following problems, fit a Pearl or a Gompertz curve to data from the appropriate tables in App. A. Explain how you computed the upper limit you used. Explain why you selected the particular growth curve (Pearl or Gompertz) you used.

3 Substitution of surface mining for deep mining of coal.

4 Growth in power plant efficiency (kilowatt · hours per pound of coal).

5 Transition from sail to mechanical power in the U.S. merchant marine.

6 Growth in the percentage of dwellings with electric power.

7 Number of telephones per 1000 population.

8 Substitution of coal for wood as an energy source.

9 Substitution of oil and gas for coal as an energy source.

10 Number of automobiles per capita.

11 Growth in the reliability of space launches (percent successful).

12 Redo the forecast of CATV in the text, fitting a Pearl curve to the data from 1956 through 1978 and forecasting to 1989. Does your forecast agree better with actual data than the forecast in the text?

Chapter

5

Trend Extrapolation

"If you can look into the seeds of time and say which grain will grow and which will not, speak then to me." WILLIAM SHAKESPEARE
Macbeth, Act I, Scene III

5.1 Introduction

A specific technical approach to solving a problem will be limited to a maximum level of performance that it cannot exceed. This upper limit is set by physical and chemical laws that govern the phenomena utilized in the technical approach. However, the upper limit to the performance of a specific technical approach does not present an absolute barrier to progress. When a technical approach is reaching its limit, a new one may be found that utilizes a different set of physical or chemical phenomena. This new approach will therefore be subject to its own ultimate limit, but this can be higher than the limit of the approach it replaces.

Such behavior is illustrated in Fig. 5.1, which shows the growth curve for the speed of propeller-driven aircraft reaching its limit, while the growth curve for jet-propelled aircraft is just beginning its progress toward a higher limit. Note that the initial jet aircraft had lower speeds than did contemporary propeller-driven aircraft. This is often seen in a "successor" technical approach. While it is still in the experimental stage, the performance of the successor approach is often less than that of the current standard approach. The important point is not that "it hasn't caught up yet," but that it has an inherently higher upper limit so that it will ultimately surpass the current approach.

If we are to forecast far into the future, some method is needed that can project progress beyond the upper limit of the current technical approach. Growth curves are not suited for this purpose.

In projecting beyond the limits of the current technical approach, however, the forecaster must distinguish between predicting a successor technical approach and inventing one. The forecaster must be able to estimate the performance of devices available at some future time without knowing how those devices will operate. This chapter presents methods for making such long-term forecasts.

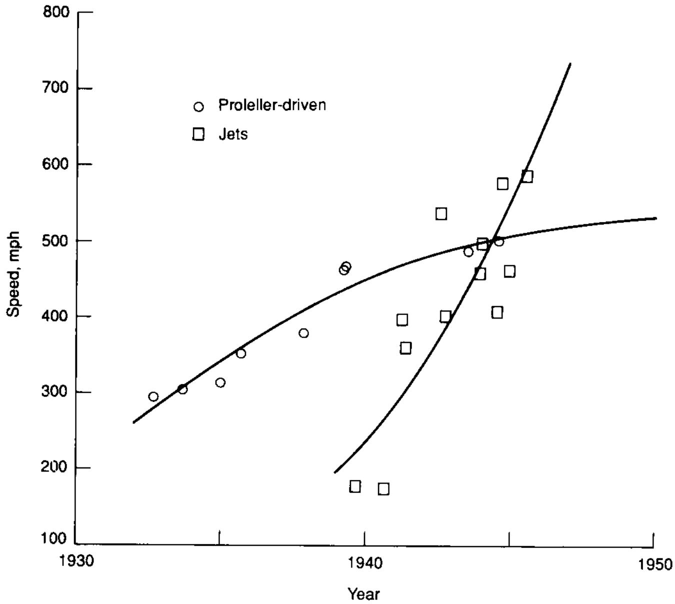

Figure 5.1 Speeds of jet and propeller-driven aircraft, from 1930 to 1950.

5.2 Exponential Trends

Figure 5.2 shows official speed records set by aircraft from the first decade of the century to the mid-1960s. During this period, many aircraft design innovations were introduced, including the enclosed cockpit, the all-metal fuselage, retractable landing gear, and other "streamlining" features, not to mention such innovations as the supercharged engine and the pressurized cabin. Each of these innovations allowed the aircraft designer to overcome some limitation of the prior technical approach. Nevertheless, the speed trend shown in Fig. 5.2 appears to show a fairly smooth, steady progression. The intersecting growth curves so apparent in Fig. 5.1 are completely invisible in Fig. 5.2. There is no hint from the long-term trend that any major changes in the technology have taken place.

This is actually a very common situation, and several of the problems at the end of this chapter display it clearly. The history of some technologies will be filled with a succession of technical approaches. Each successive approach will utilize a different set of phenomena to solve a specific, perennial problem or perform some function for which there is a continuing demand. A plot of performance against time, however, will show a smooth trend, with none of the wiggles and jumps that might be expected as technical approaches with higher performance displace the older approaches. Martino (1971) and Petroski (1973) provide many examples of this type of behavior.

It is the existence of these long-term, smooth trends that makes it possible to forecast beyond the upper limit of the current technical approach. More-

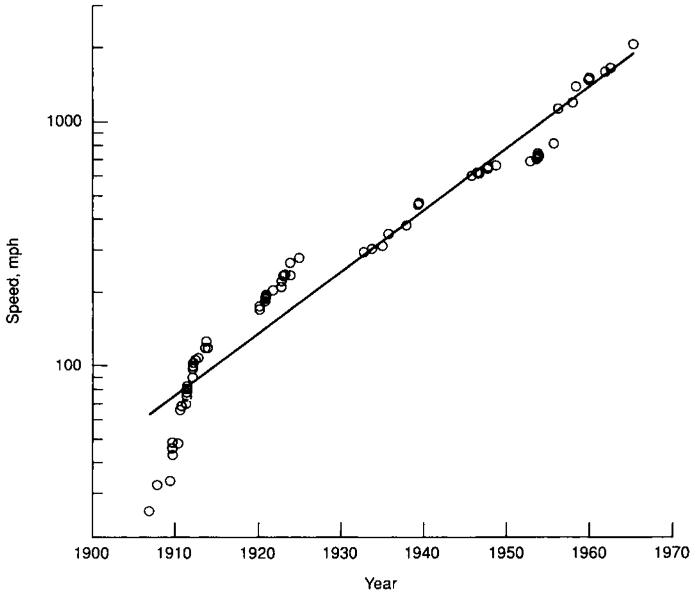

Figure 5.2 Aircraft speed records.

over, the trends make it possible to forecast a successor approach without knowing what that approach will be. In fact, the forecaster's function has been fulfilled by providing warning that something new is coming; it is asking too much to demand that the forecaster invent it as well. Moreover, the forecaster should not shrink from the implications of a long-term trend. If the past history of a technology has seen a succession of technical approaches that have cumulatively produced a trend, that trend must be projected into the future unless the forecaster can find some good reason to believe that the present is a point of discontinuity and the trend will not be continued. In particular, the forecaster should not shrink from projecting a trend simply because he or she does not see how it can be continued. It is not the forecaster's job to *see how* it can be continued, but rather to *foresee* that it *will* be continued.

The most common long-term trend is exponential growth. This is growth by a constant percentage per unit time, i.e., growth is proportional to the value already reached. This is expressed mathematically as:

$$\frac{dy}{dt} = ky \tag{5.1}$$

where k is the constant of proportionality. Solving this differential equation, we have:

$$y = y_0 e^{kt} \tag{5.2}$$

This is equivalent to the growth of money under compound interest. The proportionality constant k can be viewed as an "interest rate." The only major difference is that interest is compounded periodically, while exponential growth is compounded continuously.

Exponential trends are conventionally plotted on semilogarithmic scales, as in Fig. 5.2, where they then appear as straight lines. This can be seen by taking logarithms of both sides of Eq. (5.2):

$$\ln y = Y = \ln y_0 + kt \tag{5.3}$$

In fitting a trend to historical data, the forecaster regresses the logarithm of the actual trend values versus time, using the linear form in Eq. (5.3). The constant term of the regression is the logarithm of the initial value y_0, and the slope term of the regression is the growth rate k.

If an exponential trend is fitted to the data plotted in Fig. 5.2, the constant term is found to be 42.558 and the growth rate is 0.0582. That is, the regression gives a speed of 42.558 mph in year zero (actual year of record minus 1900) and an average growth of almost 6 percent per year after that.

5.3 An Example of a Forecast

We will illustrate the use of exponential trends through what is actually a "postdiction" rather than a "prediction." We will use data on U.S. electric power consumption from 1945 through 1965 to forecast consumption for 1970. The data come from Table A.1 and is included on the TEKFOR disk.

The TEKFOR output is shown in Table 5.1. A plot of the data is shown in Fig. 5.3. The data alone suggest a straight line when plotted on semilog paper, hence an exponential trend is an appropriate method for forecasting. Only the fitted curve is shown in the figure; the confidence bounds are so close to the fitted curve they would only clutter the graph.

The forecast power production for 1970 is 1,780,009 million kW · h. This is only 4 percent greater than the actual production of 1,639,771 million kW · h. The historical trend in growth of power production continued for 5 years after 1965. However, the forecast power production for 1975 is 2,593,658 million kW · h. This is about 29 percent higher than the actual production of 2,003,000 million kW · h. Moreover, the actual production is below the 50 percent confidence limit for 1975. This deviation is the result of the oil embargo levied by the Organization of Petroleum Exporting Countries in the early 1970s, which resulted in higher energy costs and reduced consumption in the United States. Hence, after 1970, the U.S. power production departed from its historical trend. This problem of external events affecting a trend is something we will take up again in Chap. 8.

As with any other forecasting technique, the use of exponential trends cannot be simply a mechanical process of fitting curves to data. Selecting the proper data and determining that an exponential trend is appropriate are part of the forecaster's task before the trend is fitted. Interpreting the trend is part of the forecaster's task after the trend is fitted. The use of confidence bounds makes the task of interpretation simpler and provides some guidance about the degree of deviation to be expected between trend and actual values.

TABLE 5.1 TEKFOR Output of U.S. Electrical Power Production (Millions of Kilowatt-Hours)

RESULTS OF EXPONENTIAL FIT TO ELECPWR.DTA
CONSTANT TERM = 6.8411378000388E-059
GROWTH RATE = .075290200584577
CORRELATION COEFFICIENT = .99497744267493
STANDARD ERROR OF GROWTH RATE = .0017377202604459

X	Y	BEST FIT	LOWER CONF	UPPER CONF
1945	271255	270999.89521	261425.79761	280924.62135
1946	269609	292191.27372	281975.32088	302777.35005
1947	307400	315039.75442	304129.07184	326341.85961
1948	336808	339674.91775	328011.05993	351753.53468
1949	345066	366236.47691	353754.78500	379158.56607
1950	388674	394875.07029	381503.96403	408714.81252
1951	433358	425753.11572	411413.31507	440592.72976
1952	463055	459045.73164	443649.40279	474976.37190
1953	514169	494941.73021	478391.55068	512064.47093
1954	544645	533644.68813	515832.82498	552071.60030
1955	629101	575374.10122	556181.09590	595229.42939
1956	684804	620366.62917	599660.18207	641788.07614
1957	716356	668877.43778	646511.08433	692017.56569
1958	724752	721181.64604	696993.31574	746209.40379
1959	797567	777575.88641	751386.33479	804678.27419
1960	844188	838379.98713	809991.08964	867763.87026
1961	881495	903938.78606	873131.68150	935832.87178
1962	946526	974624.08630	941157.15610	1009281.0786
1963	1011417	1050836.7649	1014443.4329	1088535.7140
1964	1083741	1133009.0462	1093395.3821	1174057.9116
1965	1157583	1221606.9534	1178449.0614	1266345.4004
1966		1317132.9509	1270074.1234	1365935.4037
1967		1420128.7948	1368776.4079	1473407.7693
1968		1531178.6046	1475100.7327	1589388.3497
1969		3505138.2911	3352857.0605	3664335.8838
1970		1780008.5539	1713007.9217	1849629.7721
1971		1919199.8808	1845903.5215	1995406.6610
1972		2069275.5517	1989053.8660	2152732.7048
1973		2231086.6897	2143248.6127	2322524.6888
1974		2405550.9731	2309338.2558	2505772.1491
1975		2593657.8399	2488238.8093	2703543.1508

5.4 A Model for Exponential Growth

Empirically, many technologies do grow exponentially. This idea is commonly accepted today. It is even widely deplored as meaning that technology is somehow "out of control" and growing too fast. However, the idea was not always recognized or accepted. Henry Adams in 1918 expressed the notion that the growth of technology is similar to the behavior of a mass introduced into a system of forces previously in equilibrium. The motion of the mass will accelerate until a new equilibrium is reached. Arthur Conan Doyle, in a story entitled "The Great Keinplatz Experiment," written late in the nineteenth century, made the statement, "Knowledge begets knowledge as money begets interest." M. Petrov and A. Potemkin, writing in the Russian journal *Novy Mir* (*New World*) (June 1968, pp. 238–252), attribute the first formulation of

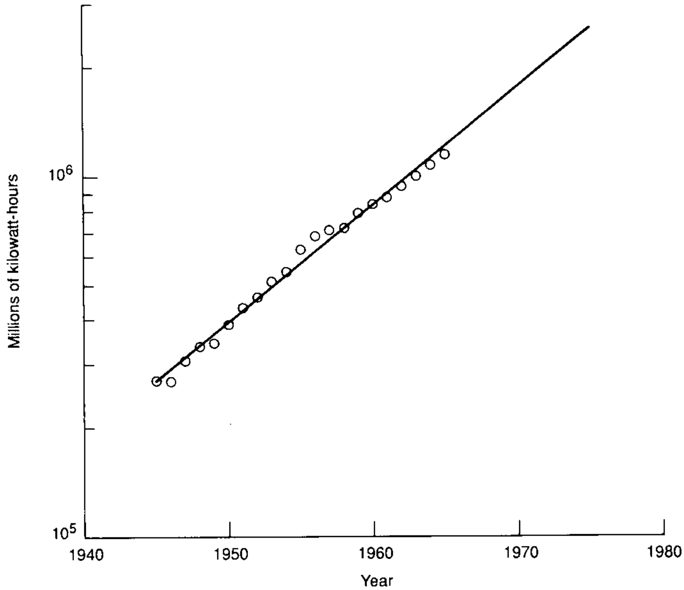

Figure 5.3 Electric power production in the United States.

the law of the exponential growth of science to Engels, who used it to refute Malthus's argument that population grew geometrically (i.e., like compound interest) while food production grew only arithmetically. Engels's argument was that an exponentially growing technology would allow resources to keep up with population.

Thus, the idea of exponential technological growth has a respectable intellectual history, as well as considerable empirical foundation. However, both the empirical evidence and intellectual acceptance present us with a brute fact. While exponential growth has been accepted, it has not been explained.

Attempts have been made to explain exponential growth through a competitive model. Holton (1962) was apparently the first to advance such a behavioral explanation. Observing the exponential growth of the operating energy of nuclear-particle accelerators, he wrote:

> One research team will be busy elaborating and implementing an idea— usually that of one member of the group, as was the case with each of the early accelerators—and then work to exploit it fully. This is likely to take from two to five years. In the meantime, another group can look so to speak over the heads of the first, who are bent to their task, and see beyond them an opportunity for its own activity. Building on what is already known from the yet incompletely exploited work of the first group, the second hurdles the first and establishes itself in new territory. Progress in physics is made not only by marching, but even better by leapfrogging.

Holton need not have confined his remarks to the field of physics. Many other fields exhibit competition of this sort, with each participant attempting to get an edge on the others by leapfrogging them. Seamans (1969) exploited this idea of competition in a model that appears to provide a satisfactory explanation of the phenomenon of exponential growth through competition. This model will now be described.

Assume there are two competitors, A and B. They are in a situation, be it military, commercial, or scientific, where each wants to be ahead of the other in the level of some functional capability. This may arise from the desire to meet a military threat, to obtain a commercial advantage, or to obtain priority for some piece of scientific research. We assume that A desires to be m percent ahead of B, and B desires to be n percent ahead of A (m and n need not be equal). We also assume that A has a response time S from the time a decision is made to initiate a new project until that project has reached its goal; B has a response time T. A has a response fraction f, where f is a number between zero and one and represents the fraction of B's response time that A is willing to wait before initiating its own leapfrogging project after B has started a project to surpass A's current level. We assume that B has a response fraction g. A and B may be involved in a completely symmetrical situation, of course, where both have equal response times and response fractions, and both desire the same percentage lead over the other. The allowance for asymmetry does not unduly complicate the model and provides more generality.

We assume that A starts at time t equal to zero with a level of functional capability equal to 1. At time t equal to zero, B initiates a project designed to overcome A's lead. This project will have the goal of achieving a level of $1 + n$, which will be reached at time T.

Once B's project is initiated, A will wait until time fT to initiate a counterproject. This project will have the goal of achieving a level of $(1 + m)(1 + n)$, which will be reached at time $fT + S$.

B will then initiate a project at time $fT + gS$. This will have a goal of $(1 + m)(1 + n)^2$, which will be reached at time $(f + 1)T + gS$.

A will then initiate a project at time $2fT + gS$. This will have a goal of $(1 + m)^2(1 + n)^2$, which will be reached at time $2fT + (g + 1)S$.

B will initiate a project at time $2fT + 2gS$. This will have a goal of $(1 + m)^2(1 + n)^3$, which will be reached at time $(2f + 1)T + 2gS$.

A will initiate a project at time $3fT + 2gS$. This will have a goal of $(1 + m)^3(1 + n)^3$, which will be reached at time $3fT + (2g + 1)S$. We now see how the model works, and can generalize the above expressions.

A will start one of its leapfrogging projects at a time given by $pfT + (p - 1)gS$, where p is any positive integer. The project will have a goal of $(1 + m)^p(1 + n)^p$, which will be reached at time $pfT + [(p - 1)g + 1]S$.

B will start one of its leapfrogging projects at a time given by $qfT + qgS$, where q is any positive integer. The project will have a goal of $(1 + m)^q(1 + n)^{(q+1)}$, which will be reached at time $(qf + 1)T + qgS$.

We can now see that the interval between the introduction of "new models" by A will be a time period $fT + gS$, and that the ratio between the level of functional capability of the new model and that of its predecessor will be $(1 + m)(1 + n)$. Similarly, B will introduce new models at intervals of

$fT + gS$, each with a level of functional capability $(1 + m)(1 + n)$ times that of its predecessor. Part of the time A will have the superior capability, and part of the time B will have the superior capability. The fraction of the time that either is in the lead will depend on their respective response times and response fractions.

The buildup of capability described above is actually geometric in nature because of the discrete steplike nature of the process. If we imagine technology as growing smoothly, however, between the introduction of successive models, this process is equivalent to exponential growth. The exponent can be evaluated as:

$$e^{a(fT+gS)} = (1 + m)(1 + n) \tag{5.4}$$

or

$$a = \frac{\ln[(1 + m)(1 + n)]}{fT + gS} \tag{5.5}$$

As would be expected, the shorter the response times and the smaller the response fractions of the two competitors, the larger the exponent, or equivalently, the greater the rate of growth. Similarly, the greater the advantage each desires over the other, the greater the rate of growth.

The situation with many competitors is qualitatively the same, although quantitatively more complex. We can imagine each planning to outdo the current leader by a margin sufficient to provide a lead for a satisfactory length of time. The competitors will also try to keep track of the ongoing projects of all the others to avoid being leapfrogged shortly before or after the new model is ready. (We will look at how this information can be gained, and some ways of using it, in Chaps. 10 and 14.) It could be argued that under these circumstances, all the competitors would tend to adopt the same percentage increase for their goal and the same response fraction. Those whose response time was much longer than that of their competitors would soon be forced out of the market; hence, there would also be a tendency for all to achieve the same response time. Under these circumstances progress would be exponential, and all competitors remaining in the market would tend to take the lead alternately.

While this model provides a theoretical explanation for exponential growth and requires only measurable variables (at least in principle), it still has some shortcomings. It does not link the response times of the competitors to the possible growth rate of the technology, i.e., the model assumes that if a competitor sets a goal, that goal can be achieved, although possibly at some considerable time and cost. Clearly, in practice neither the response times nor the percentage improvements can be set arbitrarily at the discretion of the competitors. Projects that are too ambitious, trying to achieve "too much too soon," will be failures, or at best will slip, either in the performance actually achieved or in the actual delivery date. Thus, this model provides an explanation of why in a competitive situation growth should be exponential so long as technology permits it. However, it does not explain why exponential growth should be possible.

As an illustration of the application of this model, we will consider the introduction of commercial passenger transport aircraft. Table 5.2, which is extracted from the data on transport aircraft found in Table A.4, gives the year of introduction and the productivity, in passenger-miles per hour, of passenger transport aircraft that were "leaders" at the time of their introduction. The data are from three major manufacturers who have been in the market consistently for over 60 years. Table 5.2 also gives the interval in years between successive pacesetting models for each manufacturer and the productivity ratio of the two successive models. For each manufacturer, we can calculate the mean interval between the introduction of successive models and the mean

TABLE 5.2 Pace-Setting Transport Aircraft Introduced by Three Major Manufacturers

Year	Passenger-miles per hour	Interval, years	Performance ratio
	Boeing Aircraft Co.		
1933	2000		
		5	3.8
1938	7524		
		11	4.3
1949	32,250		
		10	3.5
1959	112,833		
		10	2.8
1969	313,600		
	Douglas Aircraft Co.		
1934	2982		
		1	1.6
1935	4620		
		5	2.5
1940	11,550		
		7	1.9
1947	21,420		
		7	1.8
1954	38,855		
		4	2.8
1958	109,431		
	Lockheed Aircraft Co.		
1934	1920		
		6	2.0
1940	3808		
		6	5.5
1946	21,056		
		4	1.6
1950	34,040		
		8	1.3
1958	44,100		

value of the natural logarithms of the performance ratios. We can then take the ratios of the average natural logarithm to the average interval between successive models. For the three manufacturers, these turn out to be as follows: Boeing, 0.143; Douglas, 0.144; and Lockheed, 0.131. The overall industry average of these three values is 0.139. This turns out to be not too much different from the exponent describing the overall growth of passenger-mile productivity for all transport aircraft, including those that were not pacesetters at the time of their introduction (see Prob. 2).

Since we have three competitors, the model derived above is not strictly applicable, and the simple form of the exponent cannot be used. Furthermore, at least since World War II, the response time of the manufacturers has been governed more by the question of whether the airlines have paid off their last purchases and are now ready to buy more aircraft than by technological considerations. Nevertheless, the competitive aspects of the market are clear, with each of the competitors introducing successive aircraft models that exhibit about the same growth rate in performance and with different manufacturers holding the lead at different times. While the illustration of the model would be more satisfactory if we could derive the "right" exponent directly from a knowledge of the rate at which the airlines could absorb new aircraft, which determines the interval between successive aircraft models, the illustration nevertheless shows the fundamental correctness of the model in a competitive situation.

5.5 Qualitative Trends

In the preceding sections of this chapter, we have discussed trends that can be described in quantitative terms. Not all trends, of course, can be thus described, but this does not mean that they are any less real. It does mean, however, that it is harder to define them and that forecasts based on them will of necessity be less precise.

Table 5.3 illustrates a qualitative trend. Since the time of the Wright brothers, aircraft have been designed so that more and more of the total aircraft can be adjusted, varied, or otherwise moved under the control of the pilot while in flight. The wing-warping capability of the original Wright Flyer was only the first step in a long sequence of such capabilities. Attempts to quantify this trend (e.g., in terms of the percentage of empty weight that can be varied in flight) may not be completely successful since one movable portion may later be broken down into several movable portions, as illustrated by the step from simple flaps to complex flaps. Nevertheless, each item that is listed in the qualitative trend must be a discrete capability, which can be defined unambiguously so that it is possible to determine when it was actually achieved.

What can be done with a qualitative trend such as this? As with the trends we have been considering in previous sections, it is probably safe to extrapolate it. Aircraft will continue to be designed with more and more parts that can be moved or adjusted in flight. However, it does not appear to be possible to forecast just which parts will next be designed to be adjustable.

TABLE 5.3 Growth in Complexity of Variable-Geometry Aircraft

				Thrust deflectors →
				Thrust reversers →
				Ejectable capsules →
				Variable nozzles →
				Variable inlets →
				Flight fold wing tips →
			Swing tails	Variable sweep wing →
			Segmented elevons	Swing noses →
			Eject seats	Droop noses →
			Large cargo doors	Tilt wings →
			Refueling booms	Tilt propellers →
			Multistore combination	Tilt engines →
			External stores →	
		Slats	Speed brakes →	
		Flying tabs	Drag chutes →	
		Gear doors	Segmented spoilers →	
		Cargo doors	Lead edge flaps →	
		Multinspect doors →		
		Bomb bay doors →		
		Hatches →		
	Adjustable Stabilizer	Movable stabilizer →		
	Ground folding wings →			
	Arresting hooks →			
	Simple flaps →	Complex flaps →		
	Retractable gear →			
	Trim tabs →			
Rudder →				
Elevator →				
Ailerons →				
				Time →

SOURCE: Lamar (1969), Fig. 5. (Reprinted with permission.)

Another form of qualitative trend is displayed in Fig. 5.4. This shows a sequence of milestones in space flight. As with the items in Table 5.3, the events making up the sequence of milestones must be discrete events that can be defined unambiguously. Each successive event should represent an increasing level of difficulty. To the extent possible, there should be roughly equal increments of difficulty between the successive events (either on a linear or ratio scale).

In other areas where qualitative trends might be discernible, the same limitations as described above may apply. It may be possible to extrapolate a qualitative trend, in the sense that one forecasts that the phenomenon represented by the trend will continue. It may not be possible to forecast exactly what specific events or milestones will be achieved. Likewise it may not be possible to define the relative levels of accomplishment precisely.

Nevertheless, in the absence of well-behaved quantitative trends such as those described earlier, there may be no alternative to the use of qualitative trends for obtaining a forecast. Forecasts based on qualitative trends may well be better than no forecasts at all, and they can certainly provide useful inputs

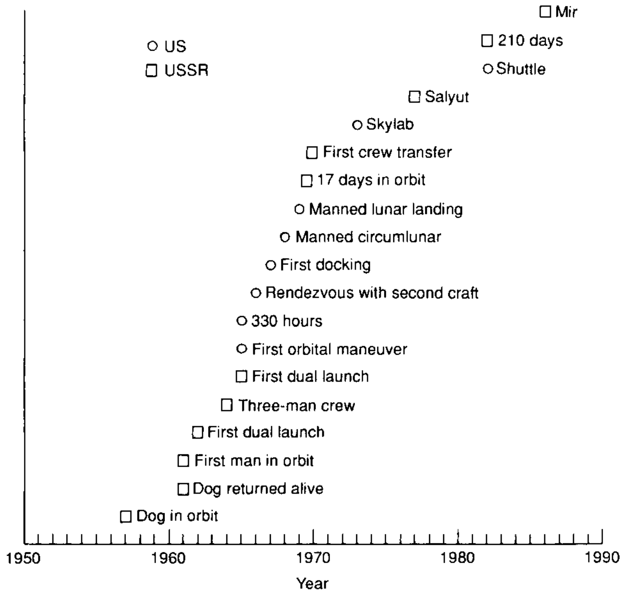

Figure 5.4 Milestones of space flight.

to a decision by indicating that a particular qualitatively described phenomenon can be expected not simply to continue at its present level but to continue to change in a manner similar to past changes.

5.6 A Behavioral Technology

In Chap. 1, it was claimed that behavioral technologies such as procedures and techniques are within the province of the technological forecaster. However, the examples shown so far have all been "hardware" technologies. It is worth showing an example of a behavioral technology that exhibits the same sort of behavior as do hardware technologies.

For the past three centuries or so, the scope of engineering projects has been increasing. One measure of the scope of an engineering project is simply its dollar magnitude. However, there is more to a project than simply spending money. A project must be managed so that the parts all come together at the right time and the tasks are accomplished in the proper order.

The techniques of engineering management have improved over the years, and it is now possible to manage projects that would have been beyond the capabilities of earlier managers and management techniques. Thus, dollar magnitude is not only a measure of project scope but also a rough measure of the capability of engineering management techniques.

Figure 5.5 shows the dollar cost of a variety of engineering projects completed over the past three centuries. At every time point, there is a project that is the largest thus far completed. However, many smaller projects are also carried out. Thus, we are interested in the outer envelope, or frontier, of

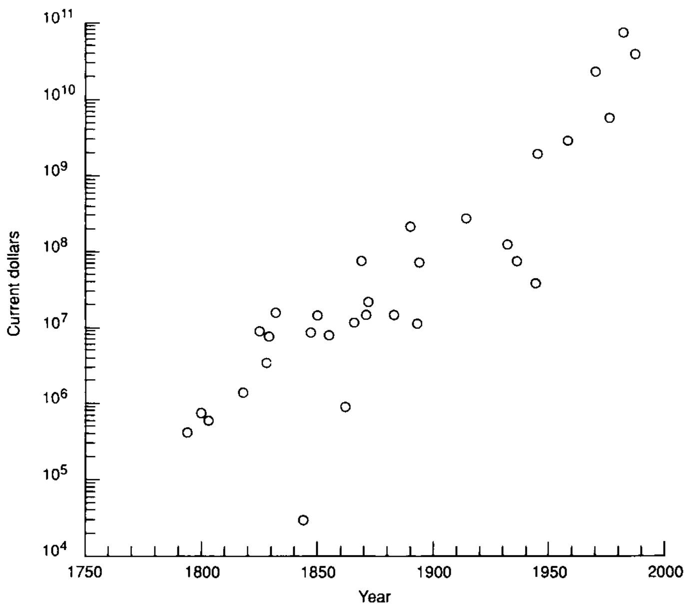

Figure 5.5 Magnitude of engineering projects.

project size versus time, rather than, say, the average size of those projects done at a particular time. The envelope of the plot of project magnitudes forms a long-term exponential trend, which indicates that engineering management capability exhibits the same kind of behavior as we have already seen for hardware technologies. Thus, the methods of technological forecasting can be applied to behavioral technologies as well as to the more conventional hardware technologies.

5.7 Summary

Any technical approach to achieving some functional capability will have an inherent upper limit beyond which that approach cannot go. However, forecasting beyond this upper limit may still be possible. The forecaster cannot and need not state what technical approach will be used to achieve the increased level of capability. Only the historical rate of improvement need be observed, especially where there has already been a succession of technical approaches. This rate of improvement can then be projected into the future. This can be done by fitting an appropriate trend curve (usually an exponential curve) to the historical data.

In some cases the overall technology will be approaching a fundamental limit, such as the speed of light, that affects all technical approaches alike. In such a case, continuation of a historical trend is impossible, and some alternative forecasting method is needed. However, in the absence of any such limit the historical trend should be projected and used as a forecast. Confidence limits can be placed on the forecast, where appropriate, to provide an estimate of the spread about the trend to be expected.

In any case, however, the forecaster cannot assume that an extrapolated trend will come about through passive waiting, especially when the past trend was established through aggressive action by the participants in the advancement of the specific technology. The forecaster must assume, however, that the aggressive actions that shaped the trend in the past will continue. If it is known that these actions will not be continued, then some other method of forecasting is appropriate, since the trend will not continue in the absence of the actions that produced it in the past.

References

Holton, G., "Scientific Research and Scholarship: Notes Toward the Design of Proper Scales," *Daedalus*, **91**:362–399 (1962).

Lamar, W. E., "Military Aircraft: Technologies for the Next Generation," *Aeronautics & Astronautics* (July 1969).

Martino, J. P., "Examples of Technological Trend Forecasting for Research and Development Planning," *Technological Forecasting & Social Change*, **2**(3,4):247–260 (1971).

Petroski, H. J., "Trends in the Applied Mechanics Literature," *Technological Forecasting & Social Change*, **10**(3):309–318 (1973).

Seamans, R. C., Jr., "Action and Reaction," *Technological Forecasting & Social Change*, **1**:17–32 (1969).

Problems

1 Take the data for the gross weight of U.S. single-place fighter aircraft from Table A.9. Plot this data on rectangular-coordinate paper and fit a free-hand curve to it. How easy would it be to extrapolate this curve? Plot the same data on semilog paper. Fit a free-hand straight line to it. Also, fit an exponential trend to the data using TEKFOR. How well does your free-hand fit agree with the regression fit? What are the advantages of plotting data in a coordinate system where the trend appears as a straight line? What are the advantages of using an objective procedure such as linear regression for fitting a trend to data?

2 Fit an exponential trend to the data on commercial aircraft productivity in passenger-miles per hour for Table A.4, using the data from 1926 to 1963. Project this trend to 1969. How well does the Boeing 747 fit the trend?

3 Fit an exponential trend to the data on the accuracy of measurements of mass from Table A.39. Are you satisfied that the accuracy of measurement of mass improved exponentially over the time period of the data?

4 Consider the following innovations in automobiles:

Self-starter	Power brakes
Fluid drive	Power steering
Automatic shift	Cruise control

What did each of these innovations mean in terms of the strength and skill required of the driver? What might be some possible steps in the continuation of this qualitative trend?

Chapter

6

Measures of Technology

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a very meagre and unsatisfactory kind."

WILLIAM THOMSON, LORD KELVIN

6.1 Introduction

Previous chapters presented methods for extrapolating some technological parameter. Nothing was said, however, about how to choose the parameter that represents the technology. This chapter will present methods for selecting or developing appropriate measures of technology for use in forecasting. It will also describe methods for dealing with multiple parameters.

The starting point for a measure of a technology must be the function the technology performs. Any measure of the technology must be related to that function and quantitatively describe how well the technology performs it. Individual technology measures are sometimes called "figures of merit" or "measures of effectiveness," implying that they measure how well the technology performs its function. Hence, the first step in selecting a measure of technology is to identify clearly the function the technology performs. This is particularly true when the forecaster is dealing with a succession of technical approaches. Each of these technical approaches can be put on a common basis only in terms of the function they all perform. Once this function has been identified, the forecaster can then attempt to select one or more measures of how well the technology performs the function.

It is also necessary to distinguish between "technical" and "functional" parameters of a technology. *Functional parameters* are those that measure utility to the user. *Technical parameters* are those that the designer manipulates to obtain the desired user utility.

For instance, for an electric light, functional parameters would include light output (lumens) and electrical power input (watts). These are of direct interest to the user, who wants adequate illumination at as low a cost as possible. Technical parameters might include filament temperature (for an incandescent lamp) and internal gas pressure (for a fluorescent lamp). These are the parameters that the designer varies to obtain the desired user characteristics.

For a jet engine, functional parameters would include total thrust, ratio of thrust to weight, and fuel economy. These are the parameters that lead the user to select one engine over another. Technical parameters might include such things as turbine inlet temperature and compressor pressure ratio. The engine designer adjusts these characteristics to obtain the combination of functional parameters best suited for a particular application (e.g., commercial transport, fighter aircraft, or vertical takeoff aircraft).

A measure of technology may be either a functional or a technical parameter. However, it is important to choose the appropriate parameter on the basis of how the forecast will be used. A technical parameter might be more appropriate for a forecast to be used for *research and development* (R&D) planning. A functional parameter might be more appropriate for a forecast to be used for marketing planning. In general, however, the forecaster may select either type of parameter as a measure of technology.

The measure chosen to describe a technology should meet the following criteria:

1. *The selected measure must actually be measurable, i.e., it must be possible to assign a numerical value to it on the basis of the performance of a device:* The most desirable measure will be objectively measurable in terms of the characteristics of the device or procedures that embody the technology being measured, i.e., it should be possible to apply a measuring instrument of some kind to the device and get a reading of the value of the measure. In some cases, an objective measure of this kind may not be possible. Then the forecaster will have to make sure that the measure can be scaled judgmentally by someone familiar with the technology. One example of this type of measure is "aircraft-handling quality." Test pilots are asked to rate an aircraft on a ten-point scale as to how well it "handles." Since "handling" is a subjective matter, this judgmental approach is the only way to get a value.

2. *The selected measure must be a true representation of the state of the art:* It must tell the full story of the way in which the technology performs its function. For a given state of the art, the designer often has the freedom to make trade-offs among various technical parameters. The user also has the freedom to make trade-offs among functional parameters. The selected measure of technology must capture these trade-offs by including all the parameters that the designer can vary. For instance, in electronic amplifiers, bandwidth can be increased if gain is reduced, and vice versa. Thus a widely used measure of performance is the gain · bandwidth product. Over a wide range of operating conditions, the gain · bandwidth product of a transistor, vacuum tube, or other amplifying device will be constant. This product represents the state of the art, and using it as a measure of technology captures the trade-offs available to the designer.

3. *The selected measure must be applicable to all the diverse technical approaches the forecaster will consider:* This leads back to the starting point above: The measure of a technology must be based on the function performed. Only then can diverse technical approaches be put on a common basis. As an example of this, see Figs. 4.2 and 5.1. Figure 4.2 displays two technologies: incandescent and fluorescent lamps. These can be compared only if a common measure exists. The common measure, in this case, is light emitted (lumens) divided by power consumed (watts). Other light sources could be compared with the two technical approaches to electric lighting if the light output is expressed in lumens and the rate of energy input is expressed in watts. This is done in Table A.6. For instance, a paraffin candle could be plotted on the same scale as electric lights if the light output is measured in lumens and the rate of energy consumption in the paraffin is converted to watts. Figure 5.1 likewise compares propeller-driven and jet-propelled aircraft. Here the common measure is aircraft speed, allowing the two different kinds of engines to be compared.

4. *The selected measure of technology should be one for which data are actually available:* Moreover, the available data should be related to a large number of devices, techniques, etc., that are historical embodiments of the technology. If a large number of devices is included in the data sample, there is less chance for the results to be distorted by the peculiarities of a single device or class of devices. In addition, the data should cover as long a time span as possible. A long time span reduces the standard error of any curves fitted to the data by regression and helps to eliminate distortions that might arise from circumstances peculiar to a particular time period (e.g., war or depression). However, it may not be possible to satisfy this criterion readily. The technology measure the forecaster prefers may be useless because data are simply not available. Some less desirable measure may have to be used simply because it is the only one for which adequate data can be obtained. An example of this situation was mentioned in Chap. 4 regarding the substitution of plastics for metal in automobiles. What the forecaster really wants to know is the amount of metal replaced by plastic. The information available, however, is the amount of plastic actually used.

5. *The measure selected should be one for which the data are consistent with respect to the stage of innovation represented:* If at all possible, all the data points should represent devices at the same stage of innovation. If the data come from devices at different stages of innovation, the forecast may be distorted. This is especially true if data from one stage of innovation are grouped at one end of the time span and data points from another stage are grouped at the other end of the time span. For instance, suppose the early data points all represent the stage of commercial introduction, while the late data points all represent the stage of laboratory prototypes. The result will be that the fitted curve will be too steep, giving an incorrectly high value for the forecast. If the grouping is the other way, the forecast will be too low. Even if data from early and late stages of innovation are mixed randomly, the standard error of re-

gression of the fitted curve will be larger than if consistent data had been used. The forecaster may not be able to obtain consistent data, in which case there is no choice but to use what data are available. However, the possibility that inconsistent data may distort the forecast should be recognized.

In many cases, the forecaster will use a measure of technology that is derived directly from the technical characteristics of the device. This may be a simple parameter such as speed or efficiency, or it may be a complex one such as gain · bandwidth product, which captures an engineering trade-off. The remainder of this chapter will present some examples of measures of technology that involve multiple parameters and the use of judgmental scaling.

6.2 Scoring Models

The scoring model has long been used in operations research as a means of ranking or rating several alternatives. It is used when the alternatives have several characteristics and the value or importance of an alternative depends upon a combination of the characteristics rather than on any single one. More recently, the scoring model has been adapted to technological forecasting. It provides a means for obtaining a measure of technology when several parameters or characteristics are important and there is no analytical procedure for combining them into a composite measure.

The scoring model can best be illustrated with an example. Delaney (1971) developed a measure of technology for aircraft-hazard (fire and explosion) detectors. These devices are used in portions of aircraft such as fuel tanks and engine compartments to detect the presence of fires or explosions. They may simply warn the pilot or, instead, automatically activate a fire extinguisher. Each hazard detector has several technical or operational parameters that describe specific aspects of its action. No single parameter gives a complete measure of hazard-detection technology. However, there is no theoretical basis on which these parameters could be combined to give a single measure of technology. Delaney solved the problem by using a scoring model, which is given by:

$$T = SRMGKH/t \tag{6.1}$$

where his seven variables are as follows:

1. S is the specificity of the detector (its ability to react only to hazards) and is equal to the product of A and B. $A = 50/(L + \delta L)$, where L is the wavelength (in angstroms) and δL is the range of wavelengths to which the detector is sensitive. B is determined by the electronics used. $B = 1$ if discriminating electronics are not used; $B = 2$ if discriminating electronics are used; and $B = 10$ if both optical redundancy and electronic logic are used.

2. R is the reset capability, i.e., the ability to determine when the hazard is no longer present and be ready to detect another occurrence. $R = 1.0$ for four

or more cycles or reset capability; $R = 0.9$ for three cycles capability; $R = 0.5$ for one cycle capability; and $R = 0.1$ for a half cycle capability.

3. M is the reliability, in *mean-time between failures* (MTBF) in hours divided by a standard MTBF of 10^5 hours.

4. G is the false alarm rate, in *mean-time between false warnings* (MTBFW) in hours divided by a standard MTBFW of 10^6 hours.

5. K is the coverage factor, which is based on the detector's ability to give warning of a hazard throughout the volume protected. $K = 10$ if volume sensors are used; $K = 8$ if lineal sensors are used; and $K = 5$ if point sensors are used. These factors were based on experimental data that showed that a point sensor would detect only 50 percent of the fires in an engine compartment; lineal sensors would detect 80 percent of the fires; and volume sensors would detect 100 percent of the fires.

6. H is the maximum temperature at which the sensor can still operate. There are two different upper limits to the maximum temperature, depending upon the use of the hazard detector. If it is used in an engine or structural compartment, the upper limit is 1255 K, set by the structure itself. If it is used in a fuel tank, the upper limit is 483 K, set by the fuel. To place detectors for both applications on the same basis, the actual upper limit for a fuel-tank detector was rescaled by 1255/493, or 2.76. Finally, to keep the number for this factor comparable to the numbers for the other factors, the maximum temperature is multiplied by 0.1.

7. t is the time in seconds required for the detector to respond to a hazard and provide an alarm. To keep this value comparable with the other values the response time is multiplied by 10.

A plot of the measure of technology for nine different hazard detectors introduced between 1948 and 1970 is shown in Fig. 6.1. The measure of technology shows exponential growth over a 22-year period. Despite the highly subjective nature of some of the elements in the scoring model, the essential features of the technology have evidently been captured.

Next, we take up a formal procedure for developing a scoring model for a measure of technology. The procedure has three steps:

1. Identify the factors to be included
2. Weight the factors
3. Construct the model

Identifying the factors involves four steps in itself. The first of these is to list all the important factors that relate to how well the technology performs its function. These might include things like speed, weight, power consumption, efficiency, delay or waiting times, precision, and accuracy. The factors must be chosen to cover all the different technical approaches that may be

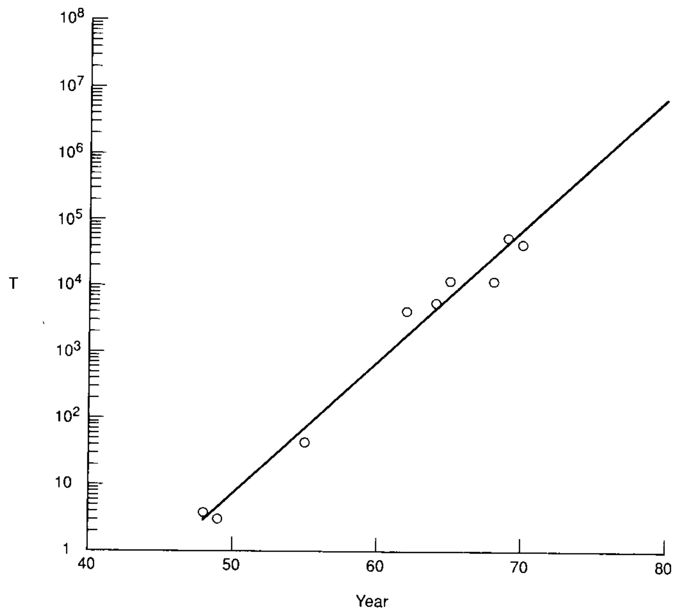

Figure 6.1 Measure of technology for aircraft-hazard detectors.

represented among the devices to be compared (e.g., the hazard detectors described above included thermocouples, infrared detectors, and ultraviolet detectors). In most cases, the forecaster will find it useful to work with a specialist in the technology to identify the important factors.

The second step in identifying the factors is to refine the initial list to eliminate overlaps and double-counting. It is also important not to mix technical and functional parameters. If at all possible, the list of factors should include only technical parameters or only functional parameters. In some cases, it will be impossible to avoid including some of each. In those instances, it must be remembered that technical parameters are "inputs" while functional parameters are "outputs." Technical parameters are used by the designer to achieve the desired functional parameters. It is particularly important, then, not to double-count functional parameters by including those technical parameters used to achieve them. At the end of this step, the forecaster will have a list of factors that are important to the technology and that measure independent aspects of that technology.

The third step is to verify that the factors can be measured or given judgmental ratings and that historical data or expert judges are available. If some factors cannot be measured or scaled, or data are not available, it may be necessary to repeat the first step of the identification process. At the end of this step, a list of factors will be available, and the forecaster will have either objective historical data or a means of obtaining judgmental data about each factor.

The fourth step in the identification process is to group the factors appropriately. One issue that always arises in the construction of scoring models is whether the factors should be multiplied or added together. This depends on the following considerations. If a factor is of such an overriding nature that it must be present, then it must multiply everything else in the scoring model. Thus, if it receives a score of 0 for some particular device (the factor is absent), the device automatically receives a technology measure of 0. On the other hand, some factors are not overriding but are subject to trade-offs. A low value on such a factor may be offset by a high value on another factor. Hence, a 0 value on one of these factors would not require a 0 for the overall measure of technology. Factors that can be traded against one another should be added together. Thus, a group of such factors will enter the model as a sum. This sum may in turn be multiplied by one or more other group sums if it is the case that at least one factor in each group must have a nonzero value. Finally, the combination of sums and products may be multiplied by one or more overriding factors. Note that in some special cases a trade-off may be multiplicative rather than additive. This is the case with the gain · bandwidth product mentioned earlier. However, in most cases these multiplicative trade-offs involve factors that must have nonzero values if the measure of technology is to have a nonzero value. That is, an electronic device with 0 gain or 0 bandwidth would properly receive an overall technology measure of 0. Hence, even though there is a trade-off involved, both factors should enter the model as multipliers. (In this case, we can look upon the logarithms of the parameters gain and bandwidth as being tradable.)

Once the factors have been selected and grouped, they must be weighted. Weighting involves three steps. The first of these is to assign numerical weights to each of the factors, reflecting their relative importance.

Consider a group of factors that can be traded against one another and are, therefore, added together. One of these can be identified as being least important. It can be temporarily assigned a weight of 1. Each other factor in the group can then be assigned a weight reflecting its importance relative to the least important factor. Once each factor in the group is given a weight, the weights are normalized so they sum to 1 within the group. This is done by adding together the weights assigned to the factors in the group, then dividing each weight by the sum. The reason for this normalization is to assure that groups are weighted properly relative to each other. A group with many factors might improperly dominate the overall score simply because it outweighs other groups with a few factors. Normalization avoids this problem.

This weighting is carried out for each group of additive factors. Then the groups, which are multiplied together, must themselves be weighted. Clearly, if we assign a coefficient to such a group, the coefficient can simply be factored out as a multiplier of the entire model. A weight that is to remain associated with a group must be an exponent, i.e., the parenthetical expression for the group must be raised to a power. Note that if we have an expression of the form $X^aY^bZ^c$, the logarithm of this is $a\log X + b\log Y + c\log Z$. Thus in a product of groups of terms, raising the groups to constant powers can be looked upon as linearly weighting the logarithms of the group, just as we linearly weigh the variables within the groups.

In the same way as with groups, overriding factors that multiply the entire model must be weighted. If one is more important than another, it can be raised to a power to reflect that relative importance.

The second step in weighting the factors is to select a judgmental scale for those that cannot be measured but must be rated. A typical scale would range from 0 through 9. However, other ranges are possible, depending upon the degree of discrimination possible and desired. The important point is that the scale must be the same for all judgmental factors. If the scale for one factor ranges from 0 through 4 while the scale from another ranges from 0 through 9, the second is implicitly being given twice the weight of the first. Relative importance of the factors should be taken into account in the numerical weights assigned to the factors, not in the range of their scaled values. The weighting and the scaling of a factor should be two separate activities.

The third step in weighting is to convert the measurable factors to a scale with the same range as that for the judgmental factors. A convenient way of doing this requires that the mean and standard deviation of the values be computed for each measurable factor. A scale is then devised for each factor in which the mean value lies in the middle of the range (e.g., if the scale runs from 0 through 9, the mean value for the factor may be scaled at the division between a scale value of 4 and a scale value of 5). The remaining scale ranges are then defined in terms of a suitable fraction of the standard deviation. For instance, if the distribution of values for a factor is roughly normal (gaussian), almost all the values will lie within three standard deviations of the mean. Thus, if a scale from 0 through 9 is desired, intervals of half a standard deviation will divide the range of values conveniently. A scale value of 4 should include all factor values from the mean to half a standard deviation below the mean; a scale value of 5 should include all factor values from the mean to half a standard deviation above the mean. A scale value of 0 should include all factor values in the lower tail more than two standard deviations below the mean, and a scale value of 9 should include all those factor values in the upper tail more than two standard deviations above the mean. If the distribution of measurements is not symmetrical about the mean, it may first be transformed (e.g., by taking the logarithm) to make it more symmetrical before the mean and standard deviation are computed. By this procedure, the measured and scaled factors can all be put on a common basis so that their relative importances are reflected in their numerical weights.

Finally, the scoring model must be constructed. This operation is almost routine after the preceding steps have been completed. The model must have the groups of additive and multiplicative factors in the proper algebraic form. In addition, it is customary to construct the model as a fraction, with the desirable factors in the numerator and the undesirable factors in the denominator. Thus, in the hazard-detection scoring model all the performance factors except operating time were desirable and placed in the numerator. The operating time appeared in the denominator.

A scoring model developed according to this procedure might end up looking something like this:

$$\text{Score} = \frac{A^a B^b (cC + dD + eE)^z (fF + gG)^y (1 + hH)^x}{(iI + jJ)^w (1 + kK)^v} \tag{6.2}$$

Here the upper case letters represent factors and the lower case letters represent weights applied to the factors. In this model, A and B are overriding factors, such that a zero value on either of them means a zero score for the technology. Factors C, D, and E can be traded for each other, as can factors F and G, and factors I and J. The weights must be normalized so that:

$$c + d + e = 1$$
$$f + g = 1$$
$$i + j = 1$$
$$a + b + z + y + x = 1$$
$$w + v = 1$$

Factors I, J, and K are costs or in some other way undesirable. The bigger these factors are, the lower the score of the technology. Hence, they are placed in the denominator. The other factors are desirable—the bigger they are, the better the technology. Hence, they are placed in the numerator.

Note the factors $(1 + hH)$ and $(1 + kK)$. These represent special cases not mentioned above. In this model, the factors H and K cannot be traded off with any other factors, hence neither of them can be part of a group. Moreover, they may not always be present. Some devices may lack factor H or factor K entirely. Factor H is not overriding, in the sense that its absence justifies a score of 0. It may be an "option" that increases the score if present but does not affect the score if absent. Hence, by including it in the form $(1 + hH)$, we achieve the desired result. A nonzero value for H raises the overall score appropriately, but a 0 score for it does not reduce the score to 0. In the same way, just because undesirable factor K has a value of 0 for some devices does not mean that their score should be infinite. Including K by weighting it then adding 1 means that when K is present, it reduces the score as it should, but when K is absent it does not affect the score at all. This method allows us to deal with factors that must stand alone, but are not overriding either as desirable or undesirable factors. The use of this method is an exception to the rule that weights must be normalized to sum to 1.0. Since there is only one factor in each "group," however, lack of normalization does not distort the overall score.

To illustrate this procedure, we will develop a scoring model for the overall technology of fighter aircraft. Table A.57 lists several performance characteristics of U.S. fighter jet aircraft from 1944 through 1982. Characteristics of fighter aircraft changed during this period, as the missions changed. The first jet fighters were small and light. They flew at speeds that were lower than the speeds achieved by later jet fighters (although faster than the propeller-driven aircraft they replaced). Their comparatively low speed meant that they could maneuver quickly. However, during the 1950s the requirements placed on fighter aircraft changed. The major threat then was high-speed bombers carrying nuclear weapons. Fighters of that era were designed for speed. Multishock air inlets and afterburners on the engines gave high thrust even during supersonic flight. Thin, swept wings reduced drag, allowing higher speeds. However, these aircraft lacked maneuverability, as seen by the reduced values for instantaneous and sustained turn rates, i.e., even though overall aircraft technology improved, performance in one area was actually sacrificed to achieve greater performance in another area. With the advent of long-range surface-to-air missiles and ground-to-air missiles, the high-altitude, high-speed bomber became obsolete. Fighters designed in the 1960s, therefore, emphasized maneuverability rather than speed. Improvements in aircraft technology were taken in the form of greater than proportional improvements in maneuverability. Experience during the fighting in Vietnam gave further emphasis to maneuverability. The F-16 and the F-18 employed leading-edge and trailing-edge flaps on the wings that allowed limited reshaping of the wing in flight to increase maneuverability.

In addition to the cycle from high maneuverability to low maneuverability and back, there was a shift from guns to missiles and back again. With the addition of radar on fighter aircraft, there were further trade-offs. Long radar range meant a large-diameter antenna, with accompanying high drag, but it also meant the ability to detect and engage targets *beyond visual range* (BVR). Availability of BVR missiles opened the possibility of engaging more than one target at a time, which required a more complex radar, which was heavier and consumed more electrical power. Another way in which aircraft technology changed was the increase in reliability and maintainability of the aircrafts. These improvements showed up as reduced maintenance hours per flight-hour and greater mean flight-hours between failures.

All these factors need to be taken into account in devising a measure of fighter aircraft technology. The several parameters that together measure fighter aircraft performance will be grouped as follows:

1. Takeoff roll is a cost. Ideally, an aircraft would require no runway at all. Thus, this term will appear in the denominator of the scoring model.

2. Maneuverability includes instantaneous turn rate, maximum sustained turn rate, and sea-level climb rate. These will be weighted respectively at 0.3, 0.3, and 0.4.

3. Availability, the extent to which the aircraft is available for flight instead of being down for repair, is also a measure of performance. Availability involves time between failures and speed of repair. The data are given in maintenance hours per flight-hour, which is the standard way of reporting the

data; in the model, this will be inverted to use flight hours per maintenance hour. Meantime between failures, given in the table, will be used directly. These two factors will each be weighted at 0.5.

4. Range and payload can be traded for each other at the design stage, since the designer can split the aircraft's lifting capability between fuel and cargo. These two will, therefore, be grouped, and each will be weighted at 0.5.

5. Both maximum speed and cruising speed are important. The former is important because the faster aircraft is the one that controls whether there will be a fight (it can escape; its slower opponent cannot). Cruise speed is the speed for maximum fuel economy; the greater the cruise speed, the sooner the aircraft can reach the place where it is needed. Each of these speeds will be weighted at 0.5.

6. The avionics grouping includes radar range and maximum number of simultaneous targets. Each will be weighted at 0.5.

7. The weapons grouping includes the number of dogfight missiles, number of BVR missiles, range of dogfight missiles, range of BVR missiles, and presence or absence of guns. Each of these will be weighted at 0.2.

Avionics and weapons are each considered to be twice as important as the other factors, all of which are of equal importance. Thus, the model will be:

$$\text{Score} = \frac{\text{maneuverability} \times \text{availability} \times \text{range} - \text{payload} \times \text{speed} \times \text{avionics}^2 \times \text{weapons}^2}{(1 + \text{takeoff roll})}$$

Note that 1 is added to the factor takeoff roll to keep the score finite even for vertical-takeoff aircraft.

Next we need to account for the differences in magnitudes of the numbers. This is done by centering and standardizing the factors. We compute the mean and standard deviation for each factor. For each aircraft, we then subtract the mean value for a factor from the aircraft's value on that factor. Those aircraft having values below the mean for a factor will of course have a negative value for this difference. We divide the difference by the standard deviation. For each factor, this gives us a set of numbers roughly centered about 0. We then rescale the results so each factor has a range of about 0 to 9. For all variables but cruise speed, this is done by multiplying the standardized score by 2.5 and adding 5. Because the values for cruise speed are unsymmetrical about the mean, the rescaling is done by multiplying by 2.0 and adding 6. Each factor is, thus, scaled to the range 0 to 9. Moreover, despite the rescaling, the relative differences of the aircraft remain the same on each factor.

Substituting the standardized values into the scoring model, we get the results shown in Table 6.1. These scores are also plotted in Fig. 6.2. A best-fit exponential trend is also plotted in Fig. 6.2. The scores appear to "make sense" in that they reflect the known relative standings of the aircraft. The F80, the first American jet-propelled fighter, was a mediocre aircraft, inferior to its contemporaries the British Meteor and to the Soviet Mig-15. Its score places it below the long-term trend. The F86, whose score is well above that of any contemporary aircraft

TABLE 6.1 Performance Scores for Fighter Aircraft

Aircraft	Year of first flight	Score
F80	1944	505.25
F84	1946	2722.84
F86	1947	8128.66
F89	1949	4232.61
F94	1950	1962.75
F100	1953	6714.56
F101	1954	11,190.06
F102	1953	19,527.68
F104	1954	25,098.36
F106	1955	20,832.61
F8	1956	31,464.34
F5A	1959	21,090.51
F4E	1967	148,451.28
F14	1971	5,470,053.52
F5E	1971	219,488.36
F15	1972	12,958,468.11
F16	1974	1,295,227.63
F18	1978	4,857,418.53
F20	1982	3,148,197.11

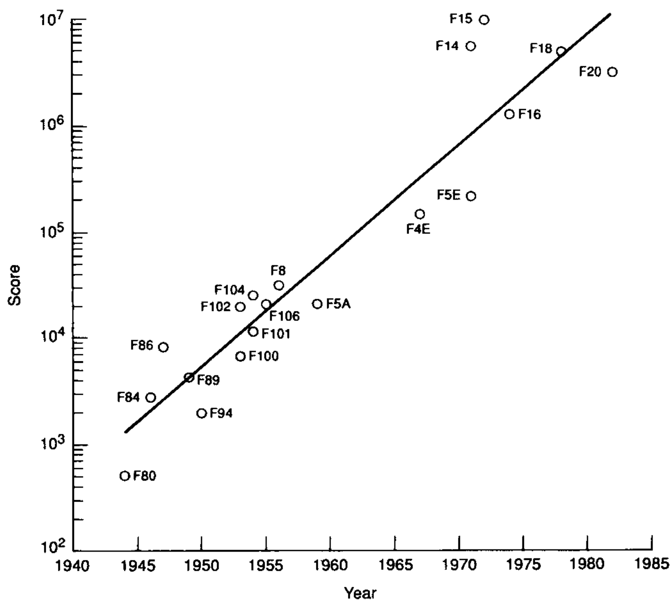

Figure 6.2 Performance scores of fighter aircraft.

and well above the long-term trend, turned in a magnificent performance during the fighting in Korea. The F4E was likewise an outstanding aircraft, giving a good account of itself not only in Vietnam but also in the Middle East. However, it was an upgraded version of a design that first flew in the 1950s. Even though its score is higher than the scores of its predecessors, it still falls below the long-term trend. The F15 was designed as an "air superiority" fighter. Its high maneuverability, heavy armament, and long range are reflected in its high score. It performed well in fighting in the Middle East. In contrast to the F4E and the F15, which were designed to be first-line fighters, the F5A and F5E were designed to be "lightweight" fighters, primarily for export to nations that could not afford first-line fighters. Their lower scores reflect the fact that they had short range and carried dogfight missiles but not BVR missiles.

By using a scoring model, we have succeeded in obtaining a single "score" for each aircraft in the set, despite the fact that they were designed for different missions and, therefore, represented different design trade-offs. By properly combining and weighting the relevant factors, we can put all the aircraft on a common basis and directly compare their overall technology scores. By extending the trend, we could forecast the overall score of some future aircraft. Note, however, that we do not have sufficient information to tell how that future aircraft's score would be "divided up" among maneuverability, speed, weapons, etc. By combining the factors into a single score, we have lost the disaggregated information.

Development of a scoring model using the procedure described above allows the forecaster to produce a measure of technology for a class of devices for which several parameters must be combined into a single number. With the scoring model, the forecaster can integrate both objectively measurable factors and factors that must be scaled judgmentally. In addition, the scoring model allows the forecaster to combine factors in an intuitively correct manner even though there is no theoretical basis for an exact combination.

6.3 Constrained Scoring Models

The scoring models described in the preceding section can take any form and, in principle, might produce a technology measure that either grows exponentially (i.e., the aircraft-hazard detector and the fighter aircraft) or exhibits a growth curve or some other behavior. Here we take up a form of scoring model that assumes that every technology, in the long run, exhibits a growth curve. The model is based on a theoretical upper limit for the technology and, thereby, allows a comparison of the state of the art from one technology to another. This method was first described by Gordon and Munson (1981).

The scoring model is of the form

$$M_i = 100\frac{C_i}{C*}\left[K_i\frac{X_{1i}}{X_i^*} + K_2\frac{X_{2i}}{X_i^*} + \ldots + K_N\frac{X_{Ni}}{X_N^*}\right] \tag{6.3}$$

where the X_i's are the different factors for the i-th device and the X^*'s are reference values to normalize the X's. A reference value may be a theoretical upper limit on the corresponding X, or it may be some other reference value that will not be exceeded. The X's are intended to increase with increasing performance. If some parameter decreases with increasing performance, such as the operating time of the hazard detector described in the preceding section, the corresponding X should be the reciprocal of the parameter, and the X^* the reciprocal of the lower limit or of the smallest value used as a reference. The K's are the weights reflecting the relative importance of the various factors. The term C represents one or more overriding factors that must have a nonzero value. A factor of 100 normalizes the maximum possible value.

In principle, the K's in this model could be obtained judgmentally, as were the weights in the scoring model for aircraft-hazard detectors. However, the originators of this model instead obtained the K's by regression on historical data for the set of devices for which the technology measure was desired. This was done as follows.

First, it was assumed that the technology measure would follow a Pearl curve with an upper limit equal to 100. The formula for the technology measure was thus incorporated in the Fisher-Pry form of the Pearl curve:

$$\tfrac{1}{2}\{1 + \tanh[A(t - t_0)]\} = 100C'(K_1X'_i + K_2X'_2 + \ldots + K_NX'_N) \qquad (6.4)$$

where the K's are identical to the above definition, but the primed variables represent the ratios in the form of the scoring model shown in Eq. (6.3). In this version of the Pearl curve formula, t_0 is the time at which the curve reaches 50 percent of its upper limit and A establishes the slope at time t_0. In Eq. (6.4), there are $N + 2$ unknown coefficients: A, t_0, K_1, K_2, ..., K_N. If there are at least $N + 2$ data points, we can solve for these coefficients. In the usual case, there will be more than $N + 2$ data points, and a nonlinear regression method must be used to solve for the K's.

The nonlinear regression method starts by choosing values for A and t_0; ordinary linear regression is then used to evaluate the K's. A measure of the goodness of fit of the data to the growth curve is obtained, such as the standard error of estimate. Another pair of values for A and t_0 is chosen, and the regression is repeated. This process is continued until a pair (A, t_0) is found that gives the smallest standard error of estimate for the fitted K's. A nonlinear regression computer program can be developed that will carry out this search automatically, obtaining a numerical solution for the "best" values of A, t_0, and the K's.

The advantages of this constrained scoring model are that it can be applied to a great variety of unrelated technologies and will allow a direct comparison of the degree of maturity of unrelated technologies. As with all scoring models, the model allows the inclusion of factors that are judged to be relevant, even in the absence of any theory to guide the selection. By virtue of the assumption that the technology measure must follow a Pearl curve, however, the need for subjective weights for the factors is eliminated. The weights can be obtained by an objective means—nonlinear regression. This objectivity may be an advantage when there is uncertainty about the relative importance of different factors in an overall measure of a technology.

6.4 Technology Frontiers

One of the reasons for using a composite measure of technology is the possibility of trade-offs among the several technical parameters of a complex device. The concept is that at any given point in the state of the art, the designer has the freedom to choose more of one parameter at the expense of some other parameter(s). The specific application of a particular device will determine which parameters will be emphasized by the designer and which will be sacrificed (the fighter aircraft example in Sec. 6.2 illustrates this point). Thus, several different devices representing the same state of the art may have different values for the various parameters, depending upon their particular applications.

This leads to the concept of a trade-off surface in the N-dimensional space defined by the N parameters that go to make up the state of the art and that can be traded off for one another. For instance, consider a scoring model such as those described in the preceding two sections. Consider two devices that have different values for the various factors in the model but have the same score. By definition they have the same value for the technology and represent the same state of the art. They could be plotted as two different points in an N-dimensional space of the N parameters in the scoring model. For instance, two different electronic amplifiers, built using the same type of transistors, could have different gains and different bandwidths, but the same gain · bandwidth product. If we were to make gain and bandwidth the two different axes of a plot, the two amplifiers would plot as different points but would represent the same state of the art.

More generally, *all* the devices that have the same score would fall on the same surface in the N-dimensional space of parameters. That surface represents all the possible combinations of factors at a given state of the art. It also defines all the trade-offs available to the designer at the given state of the art. If we plotted the logarithm of gain versus the logarithm of bandwidth, every amplifier with the same gain · bandwidth product would fall on a straight line in log-gain–log-bandwidth space.

An advance in the state of the art means that more of one parameter can be obtained without sacrificing any of the remaining parameters. Thus, an advance in the state of the art would mean moving to a "higher" trade-off surface. Conversely, a design represented by a point "inside" the frontier for the current state of the art would be "inefficient" or a "bad design," in the sense that the designer could have obtained more of at least one parameter without sacrificing any of the others.

Since the trade-off surface is the set of all points that are just barely reachable at the current state of the art, the surface represents a "technology frontier" in the N-dimensional parameter space. Points outside the frontier cannot be reached at the current state of the art. Points inside the frontier are below the current state of the art. The technology frontier, then, is an N-dimensional definition of the current state of the art.

The concept of trade-off surfaces or technology frontiers allows the forecaster to deal with technologies that must be described by multiple parameters. This means, however, that the forecaster is not attempting to forecast a future value for a single parameter but is trying to forecast the future location of the technology frontier in the N-dimensional parameter space of the technology, i.e., the forecaster is trying to forecast the set of possible combinations of the N parameters that will be possible at some future time.

6.5 Planar Technology Frontiers

One approach to defining technology frontiers, described by Alexander and Nelson (1973), treats these as hyperplanes in the N-dimensional space of the parameters that define the state of the art. The surfaces are given by the expression:

$$M = K_1P_1 + K_2P_2 + \dots + K_NP_N \tag{6.5}$$

where the P's are the values of specific parameters and the K's are coefficients that define a hyperplane. Note that this formula reduces to an ordinary plane in three-dimensional space and a line in two-dimensional space.

Figure 6.3 illustrates this for a space of three parameters. If plane M_3 in Fig. 6.3 represents the current state of the art (the technology frontier), the designer has freedom to move about in that plane, trading an increase in one parameter for a decrease in either or both of the other two parameters.

Each value of M corresponds to a plane representing a specific level of the state of the art. The planes are all parallel, and movement from one plane to another plane farther from the origin is equivalent to an increase in the state

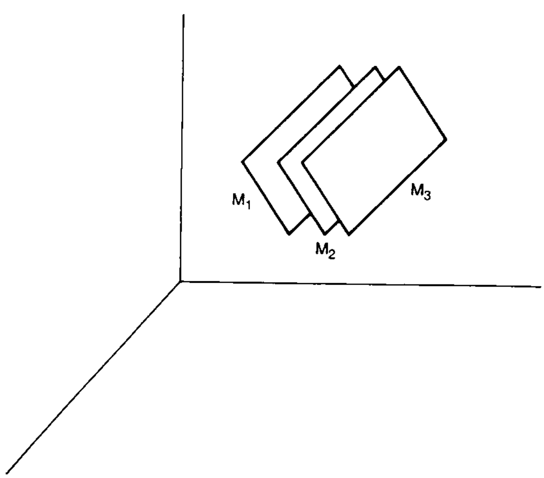

Figure 6.3 Planar trade-off surfaces.

of the art. The ratio K_i/K_j is the rate at which parameter P_j can be traded off for parameter P_i, i.e., if P_i increases by 1 unit, the score M increases by K_i units. In order to keep M constant, P_j must be reduced by K_i/K_j units.

If M_3 represents the current state of the art, to move toward the origin, via plane M_2, would represent an inefficient design. It would be regression to an earlier state of the art. A move away from the origin can happen only if the state of the art is increased.

Note that all the parameters should be selected such that an increase in the value of the parameter corresponds to an increase in performance. The idea of a trade-off would be nullified if a small value was better for one of the parameters while larger values were better for the rest.

This approach, as developed by Alexander and Nelson (1973), does not require that a numerical measure of the state of the art be developed. Instead, the forecaster obtains historical data on devices that represent differing levels of the state of the art. The mathematical expression for the technology frontier is obtained by regressing the year of introduction of each device on the parameter values for that device. The year of introduction is, thus, a "proxy" for the state-of-the-art measure M, and the K's are obtained by multiple linear regression. The "predicted" year of introduction, from the regression, then represents the actual technology level of a specific device, given its parameter values. Alexander and Nelson discussed the application of this approach to aircraft turbine engines. Hutzler et al. (1985) discussed its application to several other technologies.

To illustrate the approach, we will use the same set of data on fighter aircraft as was used for the scoring model of Sec. 6.2. We will regress year of first flight on the various parameters that describe the aircraft. It was stated above that the parameters should be chosen so that an increase in value corresponds to an increase in performance. This is not actually necessary in the Alexander and Nelson approach. A parameter that is "reversed" will simply enter the regression with a negative sign—more of it causes the device to have an earlier predicted year of introduction, corresponding to a lesser performance.

In regressing year of introduction on performance parameters, we encounter a problem we did not meet with the scoring model. In this example, there are almost as many parameters are there are aircraft, i.e., almost as many variables as cases. This means it would be "too easy" to get a good regression fit. In principle we would like to have more cases; in practice we will omit some of the variables.

Deciding which variables to omit actually turns out to be fairly simple. Those that do not significantly improve the goodness of fit can be eliminated through standard statistical tests.

The final equation, which includes all the variables found to be significant in the regression, is:

$$\begin{aligned}\text{Predicted year of first flight} = {} & 1938.740 + 5.431 \times \text{maximum mach number} \\ & + 4.036 \times \text{mean flying hours between failures} \\ & + 0.002 \times \text{payload} + 0.178 \\ & \times \text{range of BVR missiles}\end{aligned}$$

This equation explains 96 percent of the variance in the data, and all the coefficients are significant at the 0.1 level or better. The best variable, mean flying hours between failures, is significant at the 0.001 level. Substituting actual parameter values for a particular aircraft into this equation would give a predicted year of first flight, on the basis of its demonstrated performance.

The actual and predicted years of first flight are shown in Table 6.2. Note that the numbers to the right of the decimal in Table 6.2 are meaningless, since the original data was rounded down to the next lower whole year. In principle, the same should be done with the predicted values. The spurious precision has been preserved, however, to illustrate that the regression outcome is not necessarily an integer. The input data could be improved by utilizing fraction of the year as well as year of first flight. This would justify using more precision in the predicted value as well.

The actual and predicted years of first flight are plotted in Fig. 6.4. The line shown in the plot is not a regression line, but is the 45° equality line. If predicted and actual years were the same for a given aircraft, the data point would fall on the 45° line. An aircraft above the line can be interpreted as being "early" and is, therefore, an advance in the state of the art; an aircraft below the line is "late" and can be interpreted as being behind its contemporaries.

It is interesting to compare the results of this regression with the scoring model of Sec. 6.2. The F84, F86, F106, F8, F14, F15, and F18 all fall "above the line" in both Figs. 6.2 and 6.4. This implies they were more advanced than expected for the time. In both plots, the F94, F100, F4E, F16, and F20 fall below the line, implying they were less advanced than they "should have

TABLE 6.2 Actual and Predicted First-Flight Years for Fighter Aircraft

Aircraft	Actual year of first flight	Predicted year of first flight
F80	1944	1947.18
F84	1946	1947.76
F86	1947	1948.5
F89	1949	1947.65
F94	1950	1946.49
F100	1953	1950.3
F101	1954	1954.38
F102	1953	1952.21
F104	1954	1956.55
F106	1955	1955.21
F8	1956	1958.02
F5A	1959	1959.79
F4E	1967	1962.9
F14	1971	1971.87
F5E	1971	1972.49
F15	1972	1974.32
F16	1974	1969.8
F18	1978	1978.71
F20	1982	1980.9

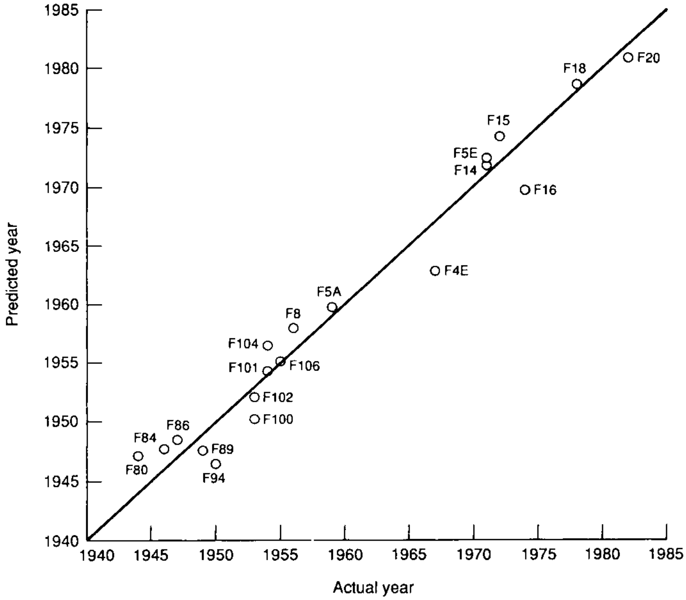

Figure 6.4 Predicted versus actual first-flight years for fighter aircraft.

been" for their actual date of first flight. Thus, for many of the aircraft in the sample, the two methods give comparable results.

However, the two plots differ for some aircraft. The scoring model puts the F80 below the trend line; the predicted year analysis shows it coming earlier than expected. Conversely, the F102 is later than expected, but has a score above the trend line. There are several other aircraft for which the two methods disagree.

What do these differences mean? First, it must be remembered that the scoring model includes parameters that did not enter the regression. For instance, the regression equation includes no parameters related to maneuverability (turn rate and climb rate) and only one parameter related to armament. Surely these are important for a fighter aircraft. Thus, the scoring model might be interpreted as including more information than the regression model. On the other hand, the weights incorporated in the scoring model were determined subjectively. They may not reflect the actual trade-offs faced by aircraft designers. Hence, the regression equation might be interpreted as more closely reflecting technological reality.

The issue of which of the two measures of fighter aircraft performance is "more correct" cannot be resolved here. In practice, it would be desirable to have data on more cases for either approach to measuring fighter aircraft technology. In addition, it would be desirable to apply the expert judgment of aircraft designers in selecting the parameters to be included in either method and in determining the proper weights for them in the scoring model.

In any case, this example illustrates the application of planar trade-off surfaces to the measurement of technology, using year of introduction as a proxy for the level of performance. This approach to technology frontiers determines empirically, rather than theoretically, what trade-off relationships were available to the designers of the devices included in the data set. The method can, therefore, be used to analyze cases where trade-offs are thought to exist but there is no theory available to guide the forecaster in combining the parameters of a device into an overall measure of technology. While expert judgment is required to determine what parameters should be tested for inclusion in the regression, no judgments are needed about their weights. The weights are computed as part of the regression.

6.6 Ellipsoidal Technology Frontiers

The planar technology frontier assumes that the trade-off rate between two parameters is always the same regardless of where a device is found on the trade-off surface and how much trading off has already been done. In many cases this assumption is unrealistic. For many technologies, the designer has to give up more and more of one parameter in order to achieve an additional unit of another parameter. That is to say, the technology frontier is convex outwards.

Dodson (1970) presented a technique for dealing with such convex trade-off surfaces by fitting an ellipsoid to them. Further details are presented in Dodson (1985). The essence of Dodson's procedure defines a surface in N-dimensional space as:

$$1 = \sum_{i=1}^{N} \left[\frac{X_{ij}}{a_i} \right]^n \tag{6.6}$$

where X_{ij} is the value of the i-th parameter for the j-th device, and a_i is the intercept of the ellipsoid on the i-th axis. Setting the left side equal to 1 is simply satisfying the equation for an ellipsoid.

Dodson limited n to the value 2, requiring the surface to be a true ellipsoid. He also fitted the surface to the data by a least-squares technique. This had the effect of giving great emphasis to outlying points.

Martino (1985) modified Dodson's work in two respects. He allowed n to take on any even value, permitting ellipsoids that were more "square shouldered" in N-dimensional space. Second, he fitted the surface to the data by minimizing the *mean absolute deviation* (MAD) rather than by minimizing the mean squared deviation. This resulted in a surface that represented the median of the data points and that was less influenced by extreme values.

In either case, however, the basic idea is the same. A set of devices is selected as representing a common level of the state of the art. For each device, the value of each of N parameters is obtained. A surface is then fitted to the data, according to Eq. (6.6). The resulting surface is taken to represent the state of the art in N-dimensional space. The mathematical expression for the surface can then be used to estimate the degree of advance over the current state of the art of a new device with a specific set of values for the N parameters, i.e., if its values are substituted in the right side of Eq. (6.6), the resulting value (presumably larger than 1) would represent the degree of advance over the current state of the art.

If data for several successive sets of devices is available, a sequence of surfaces can be fitted. The forecaster can then use the separation between successive surfaces to estimate where the technology frontier will be at some time in the future. The result is a forecast of the future technology frontier, i.e., a forecast of the entire set of trade-offs that will be possible at that time.

6.7 Summary

Because of the complexity of a technology, it is important that the parameters that measure the technology be selected carefully. They should be independent and should measure important aspects of the technology. To the extent possible, technical and functional parameters should be kept separate, rather than mixed. If the set of devices includes representatives of several technical approaches, the parameters chosen should apply to all the approaches. Finally, the form of the expression—theoretical figure of merit, scoring model, constrained scoring model, or technology frontier—should be chosen to reflect the way the several parameters actually interact in nature.

References

Alexander, A. J. and J. R. Nelson, "Measuring Technological Change: Aircraft Turbine Engines," *Technological Forecasting & Social Change,* **5**:189–203 (1973).

Delaney, C. L., "Technology Forecasting: Aircraft Hazard Detection," *Technological Forecasting & Social Change,* **5**:249–252 (1973).

Dodson, E. N., "A General Approach to Measurement of the State of the Art and Technological Advance," *Technological Forecasting & Social Change,* **1**:391–408 (1970).

Dodson, E. N., "Measurement of the State of the Art and Technological Advance," *Technological Forecasting & Social Change,* **27**:129–146 (1985).

Edwards, K. L., "Further Research into a Convention for Measuring the State-of-the-Art of Products or Processes," *Technological Forecasting & Social Change,* **24**:153–175 (1983).

Gordon, T. J. and T. R. Munson, "A Proposed Convention for Measuring the State of the Art of Products or Processes," *Technological Forecasting & Social Change,* **20**:1–26 (1981).

Hutzler, W. P., J. R. Nelson, R. Y. Pei, and C. M. Francisco, "Nonnuclear Air-to-Surface Ordnance for the Future," *Technological Forecasting & Social Change,* **27**:197–228 (1985).

Martino, J. P., "Measurement of Technology Using Trade-off Surfaces," *Technological Forecasting & Social Change,* **27**:147–160 (1985).

For Further Reference

Technological Forecasting & Social Change, **27(2/3)** (1985). Special issue on technology measurement.

Problems

1 Develop a scoring model for measuring automobile technology. Include the following factors: passenger capacity, luggage capacity, miles per gallon, range (miles per tankful of gasoline), time from 0 to 55 mi/h, and curb weight. Include any other factors

you think are important. Put all the factors in a mathematical expression, with weights that you consider to be appropriate. Obtain data on each factor for current-year automobile models from some source such as the *Consumer Reports* annual automobile issue, and score them.

2 Plot the data on computers in Table A.54 on log-log paper. For each year group, plot a best-fit line to the logarithms of operations per second and seconds per dollar. Do your results support the conclusion that computer designers were able to trade off speed for cost?

Chapter

7

Correlation Methods

7.1 Introduction

The preceding chapters have dealt with the direct forecasting of one or more measures of technological change. However, in some cases direct forecasting may not be the most appropriate means of gaining the information needed for a decision. The functional capability that interests us may be highly correlated with something else that can be measured directly or is easier to forecast. In such cases, a forecast based on the correlation may be better than one made directly, since it makes use of more readily obtained information.

In this chapter, we will examine several forecasting methods that are based on a correlation between the functional capability that interests us and other factors that may be more readily measured or forecast. In all such cases, the usefulness of the forecast depends upon the continued validity of the correlation on which it is based. This must be investigated for each case in which correlation methods are to be used.

The methods shown in the chapter are intended as examples only. In practice, many other correlations similar to the ones shown here can be found and used. The forecaster should try to discover such correlations in other cases and not assume that the ones shown here exhaust the possibilities.

7.2 Lead-Lag Correlation

There are cases where one technology appears to be a precursor to another. Thus, the trend, growth curve, or other time history of one will be similar to that of the other, except that one lags behind the other. Forecasting the lagging technology is then done by determining the lag and projecting the values of the functional capability *already achieved* by the leader.

An illustration of this situation is shown in Fig. 7.1, which displays a growth curve showing the increasing use of composite materials (plastics) in aircraft as a replacement for metal. The forecaster has estimated a 10-year lag between the first demonstration in an experimental aircraft and application in an operational aircraft, i.e., a 10-year lag between the operating prototype stage and the operational use stage. Projecting a continuation of this 10-year lag, we can forecast that nearly 50 percent of the

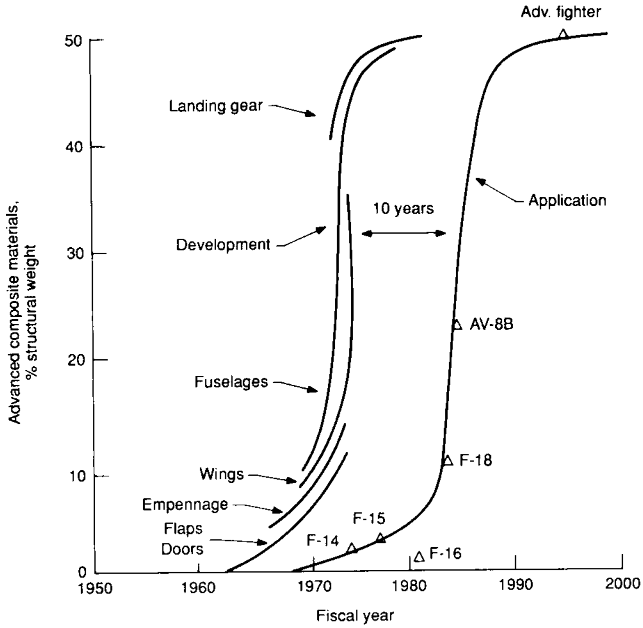

Figure 7.1 Time lag from development to application of advanced composites in aircraft. (Adapted from Hadcock, 1980).

structural weight of an advanced fighter will be composite materials by the early 1990s (Hadcock, 1980).[1]

The composites forecast was based on the transition of technology from experimental demonstration to use in practice. However, there are other forms of technology transfer that lend themselves to lead-lag forecasting. One common type of transfer is from an application where performance is the dominant requirement to later use in applications where cost is the dominant requirement.

There are numerous examples of this type of lead-lag arrangement. It is common to transfer technology from military aircraft to commercial aircraft once the technology has been proven and production costs are reduced. However, military applications are not the only ones from which high-technology products are transferred to more mundane applications. High-performance electronics technology is often first applied in laboratory instruments, then transferred to consumer products. Automotive technology is often applied first in racing cars, then is transferred to commercial automobiles. Even within the

[1]This 1980 forecast is presented primarily as an illustration of a point. Nevertheless, the reader may wonder how well this forecast holds up 10 years after it was made. In 1991, the U.S. Air Force selected Lockheed as the winner in the competition for its next-generation fighter aircraft. Lockheed's winning YF-22 prototype had a structural fraction of 23 percent composites, a figure that was expected to grow to between 35 and 40 percent by the final production design, which was to fly in the mid-1990s. In short, the forecast is holding up well. In addition, the Boeing 777 400-passenger transport, likewise scheduled to fly in the mid-1990s, will have about 10 percent of its weight in composites. Here composites are being used in a performance-sensitive application first, then are being adopted for a cost-sensitive application.

automotive field, it is common to incorporate new technology first in prestige vehicles, then transfer it to lower cost, mass-market automobiles.

Whenever a lead-lag situation such as these is encountered, we can forecast the use of a technology in the follower application once that technology has appeared in the leader application. In attempting to forecast the transfer of technology from a leading application to a lagging application, it is necessary to match the levels of functional capability in order to determine the lag. For instance, if the leading technology had a level of functional capability of 50 in 1980 and the lagging technology achieved a level of 50 in 1990, we could talk about a 10-year lag. If there were several matched pairs, where the functional capability in the leading technology was matched by the same level of functional capability in the lagging technology, we might average the lags to get an estimate of the time for the transfer of the technology or we might examine the lags to see if there was a trend, with the lag tending to increase or decrease over time.

The problem is that there are likely to be few if any matched pairs. The technology may actually be transferred from the leading application to the lagging application. However, it is very rare for devices embodying that technology to have exactly the same level of functional capability in the two applications. Thus, it is not possible to specify a lag time for devices of exactly the same level of functional capability.

Alternatively, we might look for differences in functional capability in devices introduced at the same time. Rather than estimating lag times for a given level of performance, we might try to estimate difference in performance at the same point in time. Unfortunately, the forecaster often finds that there are few if any cases where leading and lagging devices were introduced at essentially the same time.

Thus, although the lead-lag relationship might be quite clear, the problem becomes one of matching pairs to obtain either a time lag for the same level of performance or a performance lag at the same points in time. What must be done in order to obtain the equivalent of matched pairs is to interpolate between the performance of devices for either the leading or the lagging technology. This approach is illustrated in the following example.

It is already known that much of the technology introduced into commercial transport aircraft during the period 1945 through 1970 was first demonstrated in military aircraft. Thus, we would expect a lead–lag relationship between the performance of military aircraft and that of transport aircraft. One measure of functional capability is aircraft speed. Therefore, we will look for a relationship between the speeds of military aircraft and those of transport aircraft.

If we fit an exponential trend to the data on speeds of military aircraft (Table A.21), we obtain:

$$Y = 29.29e^{0.064047T} \tag{7.1}$$

where Y is speed, e is the base of the natural logarithms, and T is the year minus 1900. Using Eq. (7.1), we can interpolate between the speeds actually achieved by military aircraft to obtain the time when military aircraft "would have" had a speed equal to a specific commercial aircraft.

We can simplify the task by taking the natural logarithm of both sides of Eq. (7.1):

$$\ln(Y) = X = \ln(29.29) + 0.06404T \tag{7.2}$$

In general, if we designate X' as the performance of the lagging technology, we can write:

$$X' = A + B(T - D) \tag{7.3}$$

where A and B are identical to the constant term and the slope term, respectively, in the equation for the leading technology, and D is the lag in years. Thus, the expression for speed of commercial aircraft (in logarithmic form) becomes:

$$X' = 3.377 + 0.06404(T - D) \tag{7.4}$$

Now consider one of the lagging devices that was introduced in year T_j' with performance X_j'. Substituting this value of performance into the equation for the performance trend of the leading technology, we can solve for a time T_j that is the interpolated year at which the leader "would have" achieved performance level Y_j. Then $D_j = T_j' - T_j$. For instance, a commercial transport was introduced in 1925 that had a speed of 95 mi/h. The natural logarithm of 95 is 4.55388. Substituting this into the left side of Eq. (7.2), we obtain a T of 1918.374, the interpolated date (year and decimal fraction) at which a combat aircraft would have achieved a speed of 95 mi/h. This can be rounded to the nearest whole year, 1918, since the original data (for both combat aircraft and transports) was given only to whole years. Thus, a commercial transport achieved the performance of a military aircraft after a lag of 7 years. Without the use of the interpolation Eq. (7.1), we could not have achieved this result since there never was a military aircraft with a top speed of exactly 95 mi/h.

The resulting time lags are given in Table 7.1. This table gives the year a commercial transport was introduced, its speed, the interpolated year for a military aircraft of the same speed, and the lag in years.

We have thus far considered D to be a constant lag. However, D may itself change with time. We can determine this by regressing the lag times on the year of introduction for the follower technology. This would give an equation of the form:

$$D = a + bT \tag{7.5}$$

Substituting this into Eq. (7.3), we obtain:

$$X' = A - aB + B(1 - b)T \tag{7.6}$$

This is the lagged trend equation for the follower technology, which we can then use to forecast X' (the follower technology) on the basis of values of Y (the leader technology) already achieved.

The data in Table 7.1 can be used to obtain an expression for the lag time. However, it should be noted that subsonic transports introduced after 1959 were deliberately limited in speed for economic reasons, i.e., flying at high

TABLE 7.1 Lag of Transport Speed behind Combat Aircraft Speed

Year	Speed, mi/h	Year for same combat aircraft speed	Lag
1925	95	1918	7
1926	111	1921	5
1927	116	1921	6
1928	148	1925	3
1933	161	1927	6
1934	225	1932	2
1938	228	1932	6
1940	275	1935	5
1946	329	1938	8
1947	347	1939	8
1948	375	1940	8
1954	409	1941	13
1958	579	1947	11
1959	622	1948	11
1963	632	1948	15
1968	1550	1962	6

subsonic speeds increased fuel consumption completely out of proportion to the time saved. Hence, these transports should not be included in the computation of lag time because they did not take full advantage of the technology available from military aircraft. Thus, the lag is computed only on the basis of transports introduced between 1925 and 1959. Likewise, the supersonic Concorde, introduced in 1968, is omitted from the lag calculations.

Regressing the lag on year of introduction of the follower technology, we obtain:

$$D = -1.195 + 0.2056T \tag{7.7}$$

Substituting this into Eq. (7.6) we obtain:

$$X' = 3.454 + 0.0509T$$

or, alternatively, the transport aircraft speed X is given by:

$$X = 31.63e^{0.0509T} \tag{7.8}$$

The results are plotted in Fig. 7.2. The military aircraft speed trend, the lagged transport trend, and the actual transport data points are shown. The lag was fitted only to aircraft through 1959. The projection of the lagged trend beyond that date amounts to a forecast. As it turns out, the Concorde supersonic transport was "early," in that it appears above the lagged trend. This may in part account for the fact that it was never adopted on a large scale, it lost money for its manufacturers, and the handful ever built now serve only as charter aircraft for the "jet set."

This example illustrates the use of lead-lag correlation to forecast the transition of a technology from a leader application to a follower application. Whenever two technologies seem to show a leader-follower relationship, this method can be applied. It is necessary to fit a curve (e.g., a growth curve or an

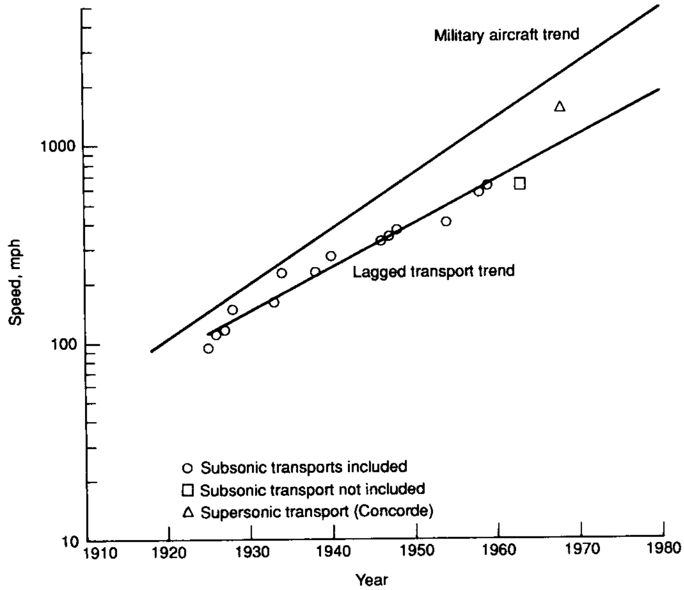

Figure 7.2 Correlation of combat aircraft and transport speed trends.

exponential trend) to the leader technology and use this curve to interpolate between the actual values achieved by the leader devices. The time intervals between these interpolated values and the actual values for the follower represent the lags. These lags should then be examined for any consistent patterns of change with time. The time lag (either a constant or a function of time) is then introduced into the interpolation equation for the leader, and the resulting equation is used to forecast the follower. The equation can be used to forecast the performance of the follower to a future date equal to the date of the most recent example of the leader plus one lag time of the follower. If the leader technology can be forecast reliably, the lag equation can be used to forecast the follower technology even farther into the future than the lag time.

Another kind of lead-lag correlation is the one that exists between indicators of some technology and its later deployment. For instance, patents and technical papers often serve as leading indicators of the later deployment of particular inventions or techniques. Martino (1982) demonstrated that the number of patents and technical papers can serve as leading indicators of new technology being brought to the market. The approach would be to identify the typical lead time of new technology in a specific industry, then use that lead time as the forecast for when a newly identified technology would reach the market.

7.3 Technological Progress Function

The technological progress function (Fusfeld, 1970) is based on an observed correlation between performance and cumulative production. This is illustrated in Fig. 7.3, which shows the improvement (decrease) in two perfor-

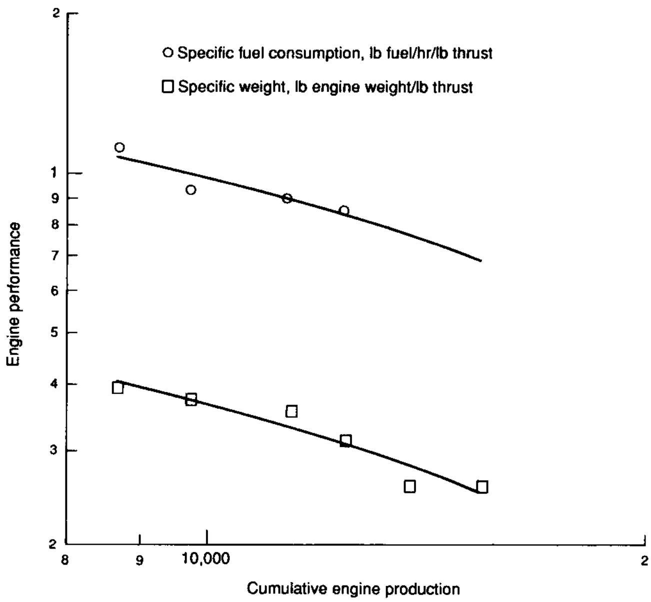

Figure 7.3 Improvement in jet engine performance versus cumulative production.

mance measures of jet engines (specific fuel consumption and specific weight) with cumulative engine production. This phenomenon has been observed for a wide range of technologies.

While the correlation is interesting, little success has been achieved in placing it on a sound theoretical basis. Nevertheless, if this behavior is observed for a technology, it might be used for forecasting improvements. The problem, of course, is that it is necessary to forecast production in order to use the progress function. However, where good forecasts of production can be made and a correlation is observed between cumulative production and performance, this method can be used.

The customary method for utilizing the progress function is to regress the logarithm of the level of functional capability on the function of the logarithm of the cumulative production.[2] The data should fall on or near a straight line when plotted on log-log paper. In effect, this means that equal percentage increases in cumulative production bring about equal percentage improvements in functional capability. The ratio of the two percentages is of course the slope of the curve. So as cumulative production grows, it becomes harder and harder to squeeze out the same increase in functional capability. This relationship seems to make sense, which gives reason to believe that there is some basis for the progress function. Perhaps additional research will show why the method works as well as it does and will identify those technologies for which it is appropriate.

[2]This can be done using the log-log regression capability of the TEKFOR program.

7.4 Maximum Installation Size

In many industries there is a relatively constant relationship between total industry capacity and the size of the largest single installation. This relationship, which was apparently first observed by Simmonds (1972), can be used for forecasting maximum installation size if the total production or total capacity can be forecast.

Figure 7.4, taken from Martino and Conver (1972), illustrates this behavior. It shows the size of the largest steam turbine electric generator in the United States by year and the total installed steam turbine capacity (data obtained from Table A.51). The growth in size of the largest single unit roughly parallels the growth in total capacity. This parallelism is disturbed by the appearance of two "freak" units, one in 1915 and one in 1929. These were significantly larger than units introduced in the immediately previous years; they were not exceeded by other units for many years. In the case of the generator introduced in 1915, there was a lag of 8 years before a slightly larger generator was introduced. In the case of the generator installed in 1929, there was a lag of 24 years before a slightly larger one was introduced. If the 15-year period of stagnation of the electric power industry (1930 to 1945) is subtracted, this lag becomes 9 years, which is similar to the lag following the 1915 generator. This is reasonable, since both units exceeded the prior state of the art by about the same percentage.

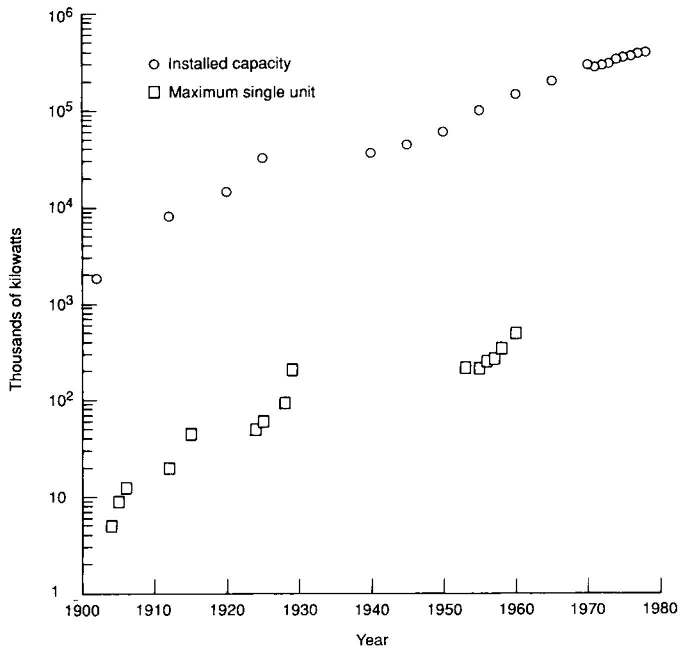

Figure 7.4 Total installed steam capacity and maximum steam turbine size.

The ratio of the largest single unit to the total capacity can be computed for the 11 units for which both items of data are available. The mean value of this ratio is 0.00415, and the standard deviation is 0.00192. If the 1929 unit is excluded from the calculation, the mean value becomes 0.00362, and the standard deviation drops to 0.000975. It can be concluded that the maximum size of steam turbines does represent a fairly constant proportion of the total installed capacity of steam plants in the United States. This proportion has remained nearly constant for over 40 years.

There seems to be no theoretical basis for the correlation between total capacity and the maximum size of a single unit. Nevertheless, when such a correlation is observed and total capacity can be forecast, the correlation can be used to forecast the size of the largest single unit. The advantage of this correlation, for the technological forecaster, is that variables such as total capacity and total production are often easier to forecast than are performance variables such as maximum single unit size. This is because maximum capacity is often related to such things as gross national product or population, which are more readily forecast. Therefore, the relation can be used advantageously by the technological forecaster when it is found to exist.

7.5 Correlation with Economic Factors

In some cases, the level of functional capability of some technology may be related to economic factors. For many technologies related to communications or transportation, the extent of development of the industry within a country is often closely correlated with the wealth of the country.

Consider a telephone system. One measure of its extent of development is the number of telephones per capita. This is an indicator of the degree to which the user can communicate with other people. The very first telephone subscriber in an area has a problem. There is no one to call. The more telephone subscribers there are, the more useful a telephone is to each subscriber.

The *gross national product* (GNP) is often used as a measure of national wealth. It is not a perfect indicator, since it may conceal sizable disparities in income. It can also be misleading in totalitarian countries, where the perquisites of the leaders do not show up as cash income. Despite its shortcomings, however, the GNP per capita is a useful measure of national wealth.

Figure 7.5 plots telephones per capita versus GNP per capita for 42 countries (data from Table A.32). A trend line is fitted to the data. The equation for the trend line is:

$$Y = (1.075 \times 10^{-5})X^{1.166}$$

where Y is telephones per capita and X is GNP per capita. The trend line appears to represent the data well, even though there is a great deal of scatter.

Once the relationship between GNP per capita and the number of telephones per capita is known, this equation can be used to forecast the development of the

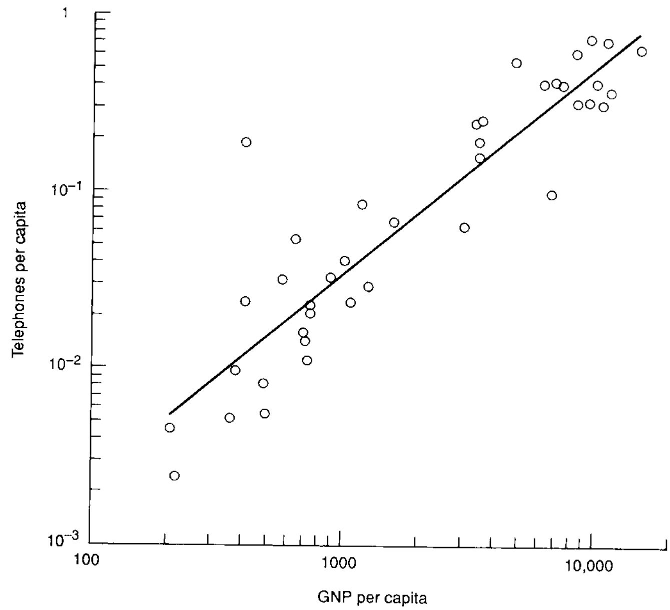

Figure 7.5 Telephones per capita versus GNP per capita for various nations.

telephone system in a nation. First, we forecast the GNP and the population, obtain their ratio, and then substitute this into the best-fit equation above. The advantage of this method is that there may not be any direct way to forecast something like telephones per capita. In such a case, this method can be useful.

Two caveats must be entered. First, this method is cross-sectional, whereas the forecaster really wants longitudinal data. Second, the forecaster must be wary of correlating the economy with itself.

What the forecaster really wants is a longitudinal relationship, i.e., one that can say something about the future of a particular nation on the basis of data about the past of that nation. However, this method does not offer any relationship between the past and the future of a single nation. It offers information about the present of many nations. In effect it assumes that the future of some countries will be like the present of other countries. In using this method, the forecaster risks subscribing to the "railroad track theory of history." This theory assumes there is only one path for development. It assumes that all countries follow the same track, passing through the same stations in the same order, with some having gotten an earlier start. Therefore, the future of the late starters will be like the past and present of the early starters. This theory may not be valid, and the forecaster should be careful not to adopt it inadvertently.

Nevertheless, in the absence of any other method of forecasting, the use of cross-sectional data in the place of longitudinal data can be defended as being better than throwing up one's hands in despair. Hence, while the forecaster should be aware of the traps involved, this method will nevertheless be useful.

There are many known relationships like the one between telephones per capita and GNP per capita. For instance, it is well known that both electric power and steel production per capita are closely correlated with the GNP per capita. However, items such as electric power and steel production are really significant elements of an economy, and correlating these with the economy is to some degree correlating the economy with itself. The results are not particularly meaningful. Use of this method, of correlating technology with the economy, should be restricted to those cases that are not equivalent to correlating the economy with itself, i.e., this method should be used only for technologies that do not represent a significant fraction of the nation's productive activity. Obviously, the value of the method depends upon there being a correlation between the technology to be forecast and the economy. But this correlation should result from some third factor relating the technology to the economy. The correlation should not arise simply because the technology is a significant fraction of the economy.

Despite these caveats, this method is extremely useful in particular circumstances, and the forecaster should take advantage of correlations with economic factors that can be used when direct forecasting is not possible.

7.6 Stages of Development

In some cases, information about a technology can be obtained early in the development of that technology. This early information might represent, for instance, a demonstration of the technology, which is not yet ready for market. This demonstration can sometimes be used as a leading indicator of subsequent market availability.

Appendix B, taken from Martino (1987), presents histories of several automotive innovations, showing progress from early demonstrations through availability in mass-market autos. There appears to be a definite pattern in these innovation histories. An innovation is first demonstrated on an experimental vehicle. It later appears in some manufacturer's prestige car. After that, it becomes available in a mass-market car. Table 7.2 lists several of these innovations and the time intervals from demonstration to prestige car, then to mass market. The average interval from demonstration to prestige car is 4 years and that from prestige car to mass-market car is 7 years (rounded to the nearest integer).

This information could be used to anticipate when a new automotive innovation will reach the consumer market. At the time the demonstration takes place, the forecaster could project a lag of 4 years before the innovation would appear in a prestige car. Once the innovation makes its appearance in such a car, the forecaster could project another 7 years before the innovation appears in a mass-market car.

For this method to be useful, the forecaster must identify the stages of development through which a typical innovation passes in a particular industry and must collect data on the time intervals between stages for past innovations in that industry. The average time lags can then be used as the basis for a forecast. The standard deviations of the lags can be used to estimate the degree of uncertainty in the forecast.

TABLE 7.2 Stages of Development of Automotive Innovations

Innovation	Demonstration	Prestige car	Mass-market car
Fuel injection	1971	1975	1987
Electronic engine control	1972	1978	1987
Electronic ignition	1970	1974	1979
Plastic body shell	1978	1982	1984
Plastic structural parts	1980	1983	1984
Turbosupercharger		1978	1984
Diesel passenger car	1979		1982
Mean interval		4.2	6.8
Standard deviation		1.1	3.8

7.7 Patents as Leading Indicators

Patents often serve as leading indicators of technological change. This technological change may be represented by a new type of product reaching the market or a new source of supply for some product from a competitor who has achieved a technological edge.

Table A.58 presents data on Japanese applications for U.S. patents on cameras and on the Japanese market share of imported 35-mm cameras. The data are plotted in Fig. 7.6. Japan had dominated the imported camera market in the United States prior to 1964, steadily displacing Germany as the major importer. However, the total import market nearly tripled between 1964 and

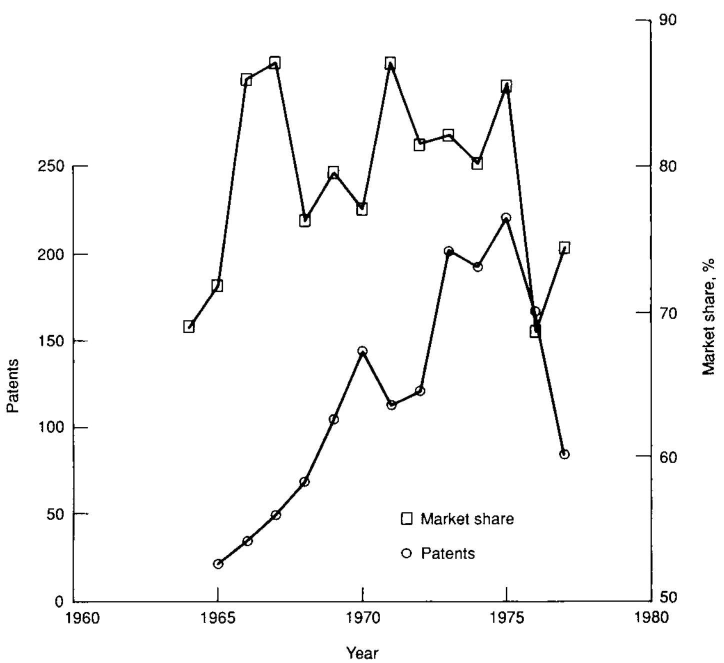

Figure 7.6 Japanese camera patents and market share.

1977, largely because the price of 35-mm cameras dropped significantly and their performance improved (i.e., more features and simpler operation). Japanese patenting activity grew dramatically between 1965 and the mid-1970s. This patenting activity reflected development of new technology that was responsible for the growth of the camera market. It also gained a larger share of this larger market for the Japanese camera firms, starting about 1970. A detailed analysis of the correlation between the two time series, patents and market share, found that Japanese patenting activity gave advanced warning of from 1 to 2 years of increased imports of Japanese cameras. (Note that in computing the cross-correlation between the two time series, a correction was needed for two recessions that reduced the total U.S. camera purchases. Failure to correct for these recessions would have given misleading results regarding the lead time available from patents as an indicator of new technology.)

Table A.59 presents data on German applications for U.S. patents on typewriters and on the German market share of imported typewriters. The data are plotted in Fig. 7.7. The number of German patent applications nearly doubled from 1966 to 1967, and for nearly a decade its level was more than 50 percent higher than the 1965 level. German market share started increasing in 1968, and through the 1970s, its level was nearly double its 1965 value. This represented an increasing market share of a growing market. A detailed analysis of the correlation between the two time series, patents and market share, found that German patenting activity gave advanced warning of nearly 3 years of increased imports of German typewriters.

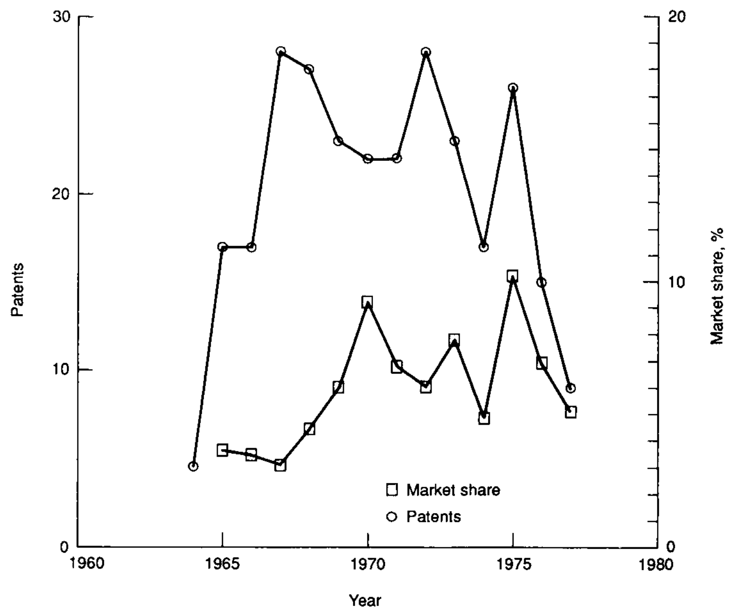

Figure 7.7 German typewriter patents and market share.

Table A.60 presents data on Japanese applications for U.S. patents on watches and on the Japanese market share of imported electrical and electronic watches. The data are plotted in Fig. 7.8. The number of Japanese patents started increasing significantly in 1967. Japanese share of watch imports increased dramatically about 1970 and accounted for about 10 percent of the market thereafter. A detailed analysis of the correlation between the two time series, patents and market share, found that Japanese patenting activity gave advanced warning of 2 years of increased imports of Japanese watches.

These examples indicate that patenting activity can be a useful indicator of technological change and of a market shift that derives from improved technology. Unfortunately, the lead time appears to be quite short—2 to 3 years at most. Patents may, thus, be a supplement to other leading indicators that can provide longer range warning, but with less detail.

7.8 Summary

This chapter has presented several methods for forecasting a technology on the basis of information about something else: a technological precursor, cumulative production, total capacity, and economic factors. These correlations are intended as illustrative examples. They do not exhaust the possibilities of forecasting by correlation. The forecaster should seek other types of correlations in those cases when it appears that a direct forecast may be difficult or impossible to obtain.

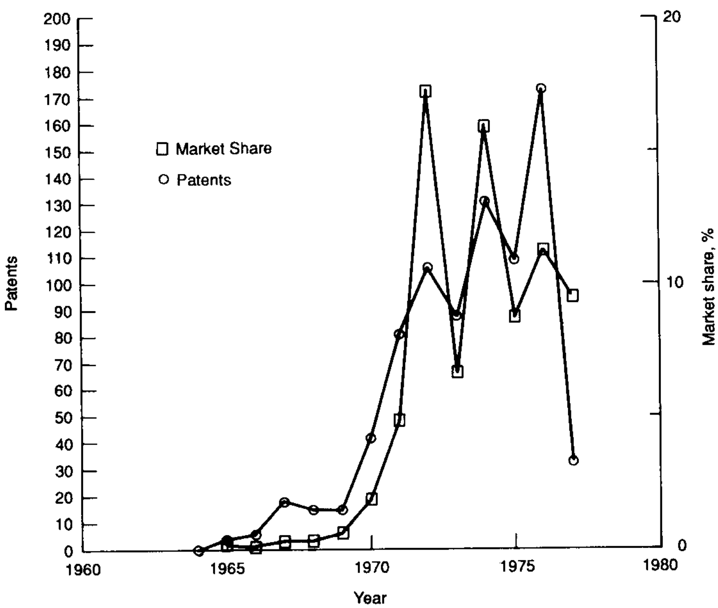

Figure 7.8 Japanese watch patents and market share.

References

Fusfeld, A. R., "The Technological Progress Function," *Technological Forecasting & Social Change,* 1:301–342 (1970).
Hadcock, R. N., "The Cautious Course to Introducing New SDM Technology into Production Systems," *Astronautics and Aeronautics,* 31–33 (March 1980).
Martino, J. P., "Technological Indicators of Market Shift," *Technological Forecasting & Social Change,* 21:77–83 (1982).
Martino, J. P., "Using Precursors as Leading Indicators of Technological Change", *Technological Forecasting & Social Change,* 32:341–360 (1987).
Martino, J. P. and S. K. Conver, "The Step-Wise Growth of Electric Generator Size," *Technological Forecasting & Social Change,* 3:465–471 (1972).
Simmonds, W. H. C., "The Analysis of Industrial Behavior and Its Use in Forecasting," *Technological Forecasting & Social Change,* 3:205–224 (1972).

Problems

1 Use the data on speeds of rocket and experimental aircraft (Table A.11) and the data on official speed records of aircraft (Table A.37). Assume that the experimental aircraft technology is a precursor to the military aircraft technology represented by the aviation speed records. Using the lead-lag correlation method, prepare a forecast of the speeds of future military aircraft. Does your forecast appear reasonable?

2 From Table A.4, use the data on the pace-setting commercial transport aircraft (i.e., those which were faster than any predecessor at the time of their introduction). From Table A.44, use the data on U.S. aircraft production. Fit a technological progress function to this data. If you needed to base a forecast of transport aircraft speed on this function, how might you estimate future aircraft production?

3 Use the data from Table A.46 on installed hydroelectric generating capacity and the data from Table A.52 on maximum size of hydroelectric turbines. Estimate the relationship between the total installed capacity and the maximum single unit. The U.S. Geological Survey has estimated that the ultimate installed hydroelectric capacity in the United States is 160,000 MW. Assuming this figure is correct, forecast the largest single turbine size that will be in use when the United States has reached the maximum installed hydroelectric capacity. Do you consider your forecast to be reasonable? Why or why not?

4 Using the data from Table A.32, correlate the number of radios per capita with the GNP per capita. Assuming a growth rate of 4 percent for the Spanish economy, when will Spain have 0.75 radios per capita?

Chapter

8

Causal Models

8.1 Introduction

In chapters 4, 5, and 6, we presented forecasting techniques that were extrapolative in nature. In chapter 7, we presented forecasting techniques that were based on leading indicators. Forecasts produced by these methods are rational and explicit. They can be analyzed in detail and critiqued by persons other than the forecaster, which cannot be done for intuitive forecasting methods such as Delphi. However, these methods still have shortcomings, which means that they cannot be applied in all circumstances.

The basic shortcoming of these methods is that they take no account of the causal factors involved in the growth of technology. They simply assume that whatever was causing the growth of technology will continue to cause that growth. These methods require no knowledge of the causes of growth.

The ability of these methods to make forecasts despite a lack of knowledge about the causal factors makes them very useful. However, the fact that they ignore the causal factors producing technological change means that they have the following serious shortcomings:

1. They are unable to provide a warning that there has already been a significant change in the conditions that produced the past behavior and that, therefore, this behavior will not continue.

2. They are unable to predict the future when the forecaster knows that an important factor affecting technological change will be altered. (This can be especially important when the change is the result of a policy decision, e.g., Man on the moon by the end of the decade.)

3. They are unable to give policy guidance about which factors should be changed and by how much to produce a desired technological change.

In Chap. 3, we discussed analogies, which implicitly take causal factors into account. In this chapter, we will examine methods that explicitly incorporate causal factors. These methods relate technological change to the specific factors that produce it, and they can warn the forecaster of the effects of change as well as provide policy guidance about the changes needed to produce the desired effects on the rate of technological change.

The development of causal models requires an understanding of what causes technological change. Here we face a major shortcoming. Existing theories of technological change are not adequate to allow us to develop elaborate and precise models. Nevertheless, enough theoretical work has been done to allow some useful models to be developed.

In the remainder of this chapter we discuss three types of models. The first type attempts to forecast technological change on the basis of factors internal to the system that produces technological change; the second type assumes that technology is driven by economic factors; and the third type includes some elements of the social and economic systems in which the technology is being developed.

Causal models can also be divided into two categories on the basis of their formal expression. The first category is the closed-form analytical model, in which the cause-effect relations are expressed as an equation or a set of equations. The second is the simulation model, which may be expressed as a set of differential equations but whose outcome cannot be expressed in a set of equations. As it turns out, the technology-only and the techno-economic models described below are both of the first category, while the techno-economic-social models are of the second. However, there is nothing inevitable in this grouping. In principle, technology-only models could require computer simulation, while techno-economic-social models could be expressed in equation form.

8.2 Technology-Only Models

This class of models assumes that technological change can be explained fully by factors internal to the technology-producing system. This does not mean that technology-only models assume technology to be "autonomous" or "out of control"; it means only that the external factors affecting technology are reflected in factors internal to the technology-producing system. Hence, a knowledge of the latter is sufficient for forecasting purposes. We will cover two growth curve models.

The growth of scientific knowledge. This model explains the growth of scientific knowledge or information on the basis of factors within science and technology. The model asserts that the rate of increase of scientific knowledge in some field is a function of the information already known, the maximum that can be known in that field, the number of people working in the field, and the amount of information transferred among workers. This relationship can be expressed as:

$$\frac{dI}{dt} = K\left[1 - \frac{I}{L}\right]\left[N + \frac{mfN(N - 1)}{2}\right] \tag{8.1}$$

where I is the amount of information available at time t, L is the maximum amount of information knowable, and N is the number of people working in the field. There is a maximum of

$$\frac{N(N - 1)}{2}$$

possible information-transferring transactions per unit time among the workers in the field; the factor f represents the fraction of these that actually take place. The factor m represents the average relative productivity resulting from one of these transfers, as compared with a researcher working alone for the same length of time. The factor m may be smaller or larger than unity. K is a proportionality constant.

If N is much greater than one, and therefore N^2 much greater than N, which is the case in most fields of technology, Eq. (8.1) can be simplified to:

$$\frac{dI}{dt} = K\left[1 - \frac{I}{L}\right]\left[\frac{mfN^2}{2}\right] \tag{8.2}$$

If we take N to be constant, this can be solved to give:

$$I = L - (L - I_0)e^{-KmfN^2t/2L} \tag{8.3}$$

This is a growth curve that has a value I_0 at time $t = 0$. It reaches an upper limit L as t approaches ∞. However, unlike the Pearl curve, it has a value of $I = I_0$ at a finite value of t.

If this model is applied to a new field of science, I is such a small fraction of L that their ratio can be ignored. However, once the field has more than 10 people working in it, N becomes negligible with respect to N^2. Therefore, we can write:

$$\frac{dI}{dt} = \frac{KmfN^2}{2} \tag{8.4}$$

If we now assume that the growth of the number of people in the field is exponential, that is,

$$N = N_0e^{ct}$$

we have:

$$\frac{dI}{dt} = \frac{kmfN_0^2e^{2ct}}{2} \tag{8.5}$$

which can be rewritten as:

$$\ln\left[\frac{dI}{dt}\right] = \ln\left[\frac{KmfN_0^2}{2}\right] + 2ct \tag{8.6}$$

The logarithm of dI/dt grows linearly at a rate that is twice the growth rate of the number of people working in the field. Figure 8.1 illustrates this behavior. It shows the production of papers in the field of masers and lasers from 1957 to 1968. The growth rate from 1958 through 1964 is 0.5349; that from 1964 through 1968 is 0.1901. In terms of the model, we can interpret this as saying that while the field was small enough that rapid communication was possible among most of the active researchers, the field grew faster than did the number of researchers. When the field became too large for such easy communi-

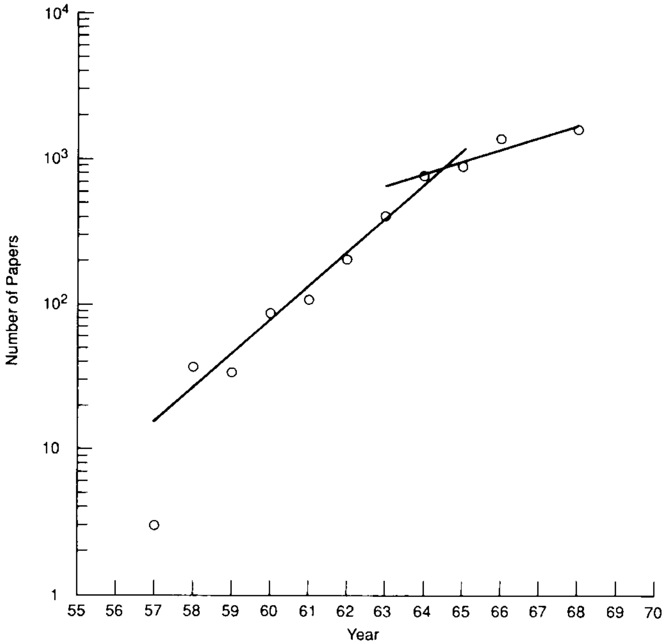

Figure 8.1 Number of papers published on masers and lasers.

cation, the rate of growth slowed down to match the growth of the number of researchers. Note that in this example we have taken the number of papers published as the measure of information available. This is not an exact measure, but the two quantities are strongly correlated, and it is the best measure available.

There are many possible variations of these information growth models. However, they are most useful at the two ends of the size spectrum: new, very small fields and very large fields. In the new fields, there is active communication among almost all workers. In very large fields, only a small fraction of the possible number of interactions among researchers actually take place. At these two extremes, these models and their variations can be useful in forecasting the growth of knowledge.

A universal growth curve. This model was developed by Floyd (1968), and it represents an attempt to explain growth toward an upper limit on the basis of effort expended by active researchers. Floyd's model is given by the following equation:

$$P(f,t) = 1 - \exp\left[-(F - f)\int_{-\infty}^{t} KNWdt' \right] \tag{8.7}$$

where f is the functional capability (Floyd called it "figure of merit"), F is the upper limit on f, W is the number of researchers working in the field to advance f, N is the number of attempts per unit time per worker to advance f, K is a constant, and $P(f,t)$ is the probability of achieving level f by time t.

By making some assumptions about the rate of growth of the number of workers in the field, Floyd converted Eq. (8.7) to the following form:

$$P(f,t) = 1 - \exp - \left[-(F - f) \int_{-\infty}^{t} (f - f_c) T(t')dt' \right] \tag{8.8}$$

where f_c is the functional capability of the competitive technology and $T(t)$ is a slowly varying function of time representing the total activity to advance f.

After some manipulation, the integration in the exponent is carried out, with the following results:

$$P(f,t) = 1 - \exp - \left[\frac{-\ln(2)\,(C_1 t + C_2)}{\ln(Y - 1) + Y + C_2} \right] \tag{8.9}$$

where the variable Y is given by:

$$Y = \frac{F - f_c}{F - f} \tag{8.10}$$

If we set $P(f,t) = 0.5$, Eq. (8.9) can be manipulated to yield

$$\ln(2) = \frac{\ln(2)(C_1 t + C_2)}{\ln(Y - 1) + Y + C_2} \tag{8.11}$$

or

$$\ln(Y - 1) + Y = Ct \tag{8.12}$$

where the subscript on C has been dropped.

For each value of f and its corresponding Y, Eq. (8.12) is used to compute the product Ct. If there are two data points available, the Ct values can be plotted against t and a straight line can be drawn through them. The projection of this line becomes the forecast. If more than two data points are available, a best-fit value for C can be computed as the slope of the regression of Ct on t. Note that for a set of data, Eq. (8.12) lends itself readily to computation of Ct on a spreadsheet.

To prepare a forecast, then, we select the future value of t of interest, multiply that by C to obtain Ct, and solve Eq. (8.12) for Y. Given Y, since F and f_c are known, Eq. (8.10) can then be solved for the future value of f.

We will do an example using Floyd's model. The first two columns of Table 8.1 are an abbreviated list of speed records set by propeller-driven aircraft. All but the last two entries are obtained from Table A.37. The last two entries are unofficial records set by the XP-47H and the XP-47J and are presented in Table A.36.

We assume a value of 600 mi/h for F as the maximum speed for propeller-driven aircraft. The performance of the competitive technology, f_c, is assumed to be 100 mi/h, about the maximum speed for a train or an automobile. Using

TABLE 8.1 Aircraft Speed Records

Year	Speed	Y	Ct
1912	108.2	1.017	- 3.077
1913	126.7	1.056	- 1.819
1920	194.5	1.233	- 0.233
1921	205.2	1.266	- 0.056
1922	222.9	1.325	0.205
1923	266.6	1.499	0.806
1924	278.5	1.555	0.967
1932	294.3	1.636	1.182
1933	304.9	1.694	1.330
1935	352.4	2.019	2.039
1937	379.6	2.269	2.507
1939	469.2	3.823	4.523
1943	490.0	4.545	5.811
1944	504.0	5.208	6.645

Eq. (8.10), we obtain the values of Y in the third column of Table 8.1. Using Eq. (8.12), we obtain the values of Ct in the last column.

We can now plot Ct versus t, the year in which the record was achieved. This is shown in Fig. 8.2, where the first seven data points from Table 8.1 are plotted. The line is fitted to these seven data points by linear regression and extended to 1950 as a forecast. From the regression, $Ct = 0$ in 1920 and $C = 0.299$.

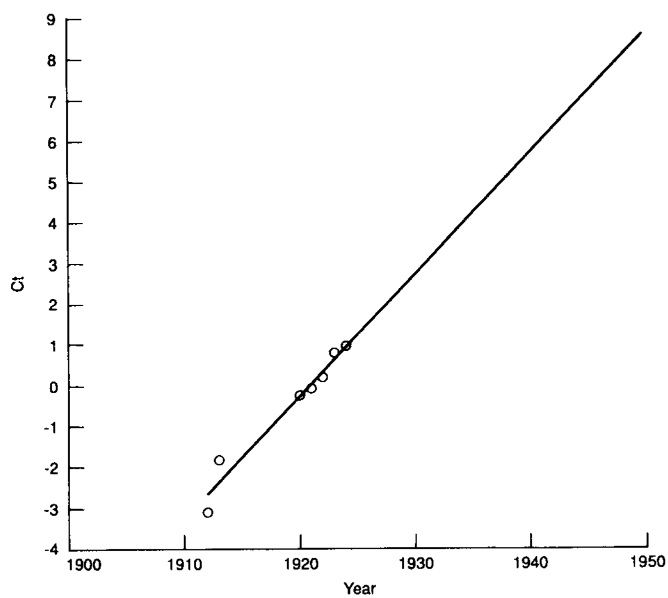

Figure 8.2 Plot of Ct versus time for aircraft speed.

Using the regression equation, we obtain the values for Ct shown in Table 8.2. These must be converted to aircraft speeds. To do this we must first obtain the Y values corresponding to each value of Ct. While Eq. (8.12) can readily be computed going from the right side to the left, solving for Y given a value for Ct is more complicated. However, we can take advantage of the fact that both terms on the left side increase with increasing Y. We can estimate values for Y and compute the corresponding Ct. If the Ct value is too large, we need to reduce the estimated value of Y; conversely if Ct is too small, we need to reduce the estimate of Y. Iteratively, we can home in on a sufficiently accurate estimate of Y. Let us determine the speed corresponding to the fitted Ct value of 1.253, corresponding to 1925 in Table 8.2. Take 1.7 as the first estimate of Y. Solving Eq. (8.12), Ct is 1.343, which is too large. Thus, the next estimate of Y should be smaller. The following table shows one sequence of estimates that converges on a good estimate for Y.

Y	Ct
1.7	1.343
1.6	1.089
1.65	1.219
1.66	1.244
1.67	1.269
1.665	1.257

In the same manner we can fill in the remainder of the second column of Table 8.2. Solving Eq. (8.10) for f, we can use the estimated Y tables to obtain the final column of Table 8.2.

These results are plotted in Fig. 8.3. The plotted curve is taken from the last column of Table 8.2. The data points shown as circles are those used to fit the curve, and those shown as triangles are the remainder of the data points from Table 8.1. It is striking that a model based on aviation progress from 1912 to 1924 gives a remarkably good forecast for performance in the late 1930s and the 1940s. However, the model does not do well at all for the four speed records set in the early 1930s. It is possible that the Great Depression slowed down aviation progress, contributing at least in part to the overoptimistic forecast for 1932 to 1935.

TABLE 8.2 Floyd Curve Fitted to Aircraft Speed Records

Year	Ct	Y	Speed (mi/h)
1915	− 1.738	1.061	128.25
1920	− 0.243	1.229	193.17
1925	1.253	1.665	299.70
1930	2.748	2.405	392.09
1935	4.244	3.375	464.86
1940	5.739	4.490	488.64
1945	7.234	5.688	512.09
1950	8.730	6.950	528.06

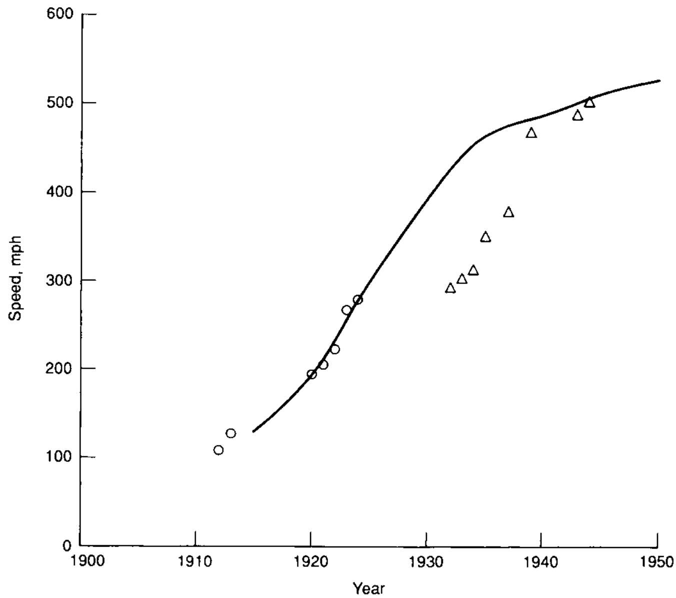

Figure 8.3 Floyd curve of aircraft speed versus time.

Floyd originally developed a nomogram to solve Eq. (8.12) since computation by numerical means is tedious, as illustrated above. However, the availability of computers has eliminated the need for Floyd's nomogram. The program TEKFOR uses an iterative method to solve Eq. (8.12), lifting that burden from the user. TEKFOR treats Floyd's model as simply another growth curve, asking for the competitive technology f_c in addition to the upper limit and the confidence bounds.

The Floyd model seems suitable for forecasting the progress of a single technical approach. The upper limit in performance and the performance of the next-best competing technology must be known. At least two historical data points must be available to permit the evaluation of the slope and intercept of Ct versus t. If more data points are available, a regression program can be used to fit a curve to the data.

8.3 Techno-Economic Models

In Chap. 4, we used a Pearl curve to forecast the substitution of one technology for another. However, we needed data from the early years of the substitution to forecast the remainder. Mansfield (1968a and 1968b) introduced a techno-economic model for forecasting substitution. He obtained data on the adoption history of three innovations in each of four industries. His measure of substitution was the number of firms that bought or installed at least one of the new devices.

The innovations are shown in Table 8.3. The number of firms that ultimately adopted the innovation is given as n_{ij}. p_{ij} is a measure of the profitability of the j-th innovation in the i-th industries. It is the ratio of the average payback time required by the industry to the payback time for the innovation (shorter payback time of the innovation means greater profitability). S_{ij} is a measure of the cost of adopting the innovation. It is the initial investment required by the j-th innovation in the i-th industry divided by the average assets of the n_{ij} firms that adopted the innovation.

The parameters a_{ij} and b_{ij} are the location and shape parameters, respectively, obtained by fitting a Pearl curve to the substitution history. As noted in Chap. 4, the location and shape of a Pearl curve are controlled independently by a and b, respectively. By fitting a Pearl curve to his data and obtaining the parameter b_{ij}, Mansfield in effect collapsed an entire substitution history into a single number. The parameter r_{ij} is the correlation coefficient for the regression fit of the Pearl curve, and the last column is *root-mean-square* (RMS) error of the fit.

The rapidity of adoption is given by the parameter b_{ij}. Mansfield hypothesized that the more profitable the innovation, the more rapidly it would be adopted (i.e., larger b_{ij}), and the more costly it was to adopt the innovation, the more slowly it would be adopted (i.e., smaller b_{ij}). He, therefore, regressed the computed values of the b_{ij} on the values of S_{ij} and p_{ij} for the innovations. The actual regression equation he used was:

$$b_{ij} = b_1 d_1 + b_2 d_2 + b_3 d_3 + b_4 d_4 + a_1 p_{ij} + a_2 S_{ij} \tag{8.13}$$

This equation makes use of a trick frequently employed in situations such as this. The d_i terms are dummy variables that are assigned the value of 1 when the data come from the i-th industry; otherwise they are assigned the value of 0. This trick allows us to use all the data from all the industries to estimate

TABLE 8.3 Innovations in Three Major Industries

	Sample data			Parameter estimates			RMS
Innovation	n_{ij}	p_{ij}	S_{ij}	a_{ij}	b_{ij}	r_{ij}	error*
Diesel locomotive	25	1.59	0.015	- 6.64	0.20	0.89	2.13
Centralized traffic control	24	1.48	0.024	- 7.13	0.10	0.94	1.52
Car retarders	25	1.25	0.785	- 3.95	0.11	0.90	5.02
Continuous wide-strip mill	12	1.87	4.908	- 10.47	0.34	0.95	0.90
By-product coke oven	12	1.47	2.083	- 1.47	0.17	0.98	0.84
Continuous annealing	9	1.25	0.554	- 8.51	0.17	0.93	1.42
Shuttle car	15	1.74	0.013	- 13.48	0.32	0.95	2.03
Trackless mobile loader	15	1.65	0.019	- 13.03	0.32	0.97	1.66
Continuous mining machine	17	2.00	0.301	- 14.96	0.49	0.98	2.22
Tin container	22	5.07	0.267	- 84.35	2.40	0.96	3.00
High-speed bottle filler	16	1.20	0.575	- 20.58	0.36	0.97	0.95
Pallet-loading machine	19	1.67	0.115	- 29.07	0.55	0.97	1.58

*RMS, root-mean-square.
SOURCE: E. Mansfield, 1968*b*., p. 149. (Reprinted by courtesy of W. W. Norton.)

a_1 and a_2, while still obtaining separate values of b_i for each industry. Mansfield obtained the following values from this regression:

$$b_{ij} = \begin{bmatrix} -0.29 \\ -0.57 \\ -0.52 \\ -0.59 \end{bmatrix} + \underset{(0.015)}{0.530p_{ij}} - \underset{(0.014)}{0.027S_{ij}} \tag{8.14}$$

The constants in the large brackets apply, from top to bottom, to the brewing industry, the bituminous coal industry, the steel industry, and the railroad industry, respectively. An algebraically large number means greater speed of adoption. The coefficients of p_{ij} and S_{ij} have the expected signs, i.e., an increase in profitability increases b_{ij} whereas an increase in the required initial investment decreases b_{ij}. The numbers in parentheses below the coefficients are the standard errors of the coefficients. Both coefficients are significantly different from zero.

We can interpret the regression results as follows: The effects of the profitability and size of the required initial investments are the same for all the industries, but for innovations with comparable profitabilities and sizes of investment, some industries will adopt more rapidly than others. In particular, the brewing industry is far more prone to adopt innovation rapidly than are the other three industries. We can now compare the values of b_{ij} given by Eq. (8.14) with those obtained by fitting a Pearl curve to the data for adoptions. The result is shown in Fig. 8.4, where the b_{ij} values computed from the Pearl curve are plotted against the b_{ij} values computed from Eq. (8.14). The 45° line indicates equality, i.e., if the values from the regression equation agreed perfectly with the Pearl curve fits to the data, all the points would lie on the 45° line. The points lie close to the line if the profitability and size of an investment explain a great deal of the variation in the rate of substitution for innovations in different industries. In addition, some industries appear to be more innovative than others.

To extend Mansfield's approach to some other industry, it would be necessary to compute the industry innovativeness coefficient [the term in large brackets in Eq. (8.14)]. To obtain this coefficient, it would be necessary to repeat Mansfield's work for that industry: First, Pearl curves are fitted to the adoption history of several innovations, and the b values regressed on the profitability and investment data. Once the industry innovativeness coefficient was known, the model could be used to forecast adoption of other innovations in the same industry.

Several extensions have been made to Mansfield's work. Most of these extensions involve attempts to replace the industry innovativeness coefficient with other variables that can be measured directly.

Blackman et al. (1973) developed an extension that allows the forecaster to estimate the industry innovativeness coefficient from other data on *research and development* (R&D) expenditures and revenue from new products as a percent of

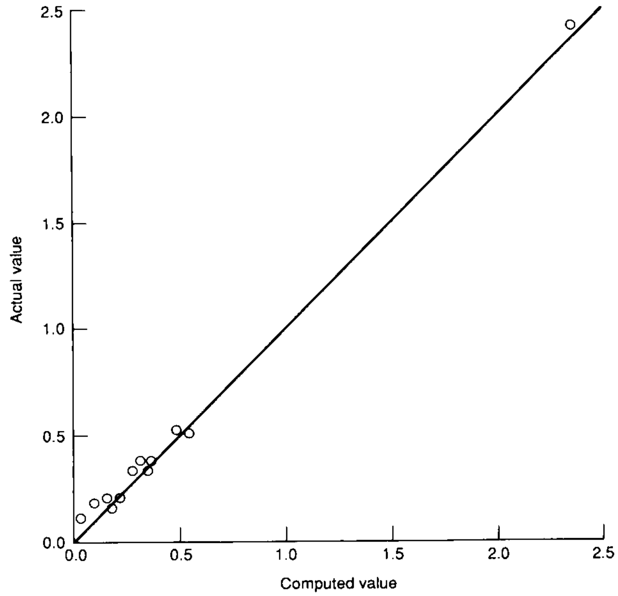

Figure 8.4 Actual versus computed rate coefficients for 12 innovations. (Adopted from Mansfield.)

total sales. If the data to estimate this coefficient are available, then it is not necessary to compute the coefficient from prior adoption histories.

Martino et al. (1978) developed a more elaborate model, using an approach related to Mansfield's. Their measure of substitution or adoption was the fraction of total industry output obtained using the innovation. (By contrast, Mansfield counted a firm as "adopting" when it had installed one of the new devices, even though the new device might account for only a small fraction of the firm's output.) They identified several factors from the literature on adoption of innovations (Rogers and Shoemaker, 1971) that had been found to affect speed of adoption. These were the independent variables in their model. They then fitted Pearl curves to adoption histories for 41 innovations in 14 industries. This resulted in 41 values for b. These b values were regressed on the variables already selected as causal factors. The result was a model that estimated the Pearl curve shape factor b entirely from industry-related economic variables that can be known at the time the innovation is introduced. The model was found to be:

$$\begin{aligned} Y = {} & 0.08312 + 0.00210X_1 + 0.22933X_3 + 0.25044X_4 + 0.00314X_7 \\ & - 0.00060X_8 + 0.00127X_9 + 0.00626X_{10} + 0.00311X_{11} \\ & + 0.00097X_{12} - 0.00069X_{13} \end{aligned} \tag{8.15}$$

where Y = diffusion rate (fitted Pearl curve b value)
X_1 = market share of four largest firms (percent)

X_3 = compatibility of innovation with existing work force [rated judgmentally on scale of 1 (incompatible) to 4 (no change in work force skills needed)]
X_4 = regulatory encouragement of innovation in the industry [rated +1 for encouragement of innovation by regulators; −1 for discouragement of innovation by regulators; and 0 for regulatory neutrality or unregulated industry]
X_7 = median age of work force in years
X_8 = fraction of facilities less than 10 years old
X_9 = growth rate of productivity in the specific industry (percent per year)
X_{10} = new capital investment per production worker (normalized to all-industry average)
X_{11} = total capital per production worker (normalized to all-industry average)
X_{12} = dollar sales per employee
X_{13} = growth rate of industry (percent per year)

Missing subscripts indicate variables originally included in the analysis that turned out not to be significant causal factors.

This model explains about 49 percent of the variance in adoption rates and uses variables that can be measured at the time the innovation is introduced. It does not require any "calibration" on previous innovations in a particular industry, as does the Mansfield model.

Other forecasters have also developed approaches to causal models that involve variables believed to be causally linked to technological change or technological substitution. Some of these are discussed here.

Day (1983) developed a model for technological substitution in which the rate coefficient of the growth curve for a new technology is estimated on the basis of the relative cost of the old and new technologies. Lakhani (1979) showed that the growth rate of a Gompertz curve for a particular substitution (filter-tip cigarettes and menthol cigarettes for regular cigarettes) could be explained by a model incorporating price and advertising. Stier (1983) used the sulfate pulp process as an example to show that the upper limit of the Gompertz curve describing diffusion of this technology could be explained on the basis of the gross national product and the price of wood pulp.

Ayres (1985) developed an interesting model for market share captured by an innovation. This model assumes that the firm introducing the innovation can act as a monopolist, at least for the period of time when no other firm is offering an equivalent technology. This firm will then attempt to set a price path over time that maximizes its discounted profit. The profit, of course, depends not only on price but the cost of production. Ayres assumes that the production cost obeys the widely used "learning curve," declining as cumulative production increases. As the price declines, the market grows, i.e., the new technology attracts users from the old technology. However, there is a limit to how rapidly users can be attracted. If users delay adoption of the new technology, they lose some benefit during the delay period, but they gain the benefit of price reduction. Hence, price cannot be reduced so rapidly that it pays users to delay adoption. The end result of this model is two equations. The growth rate for market penetration is given by:

$$kF = \sigma(r - \delta) \tag{8.16}$$

and the price trajectory for the product is given by:

$$P = P_0 e^{-(r-\delta)f} \tag{8.17}$$

where k = rate coefficient of a Pearl curve
F = upper limit of the Pearl curve
f = fractional market penetration of the innovation
σ = price elasticity of demand for the innovation [elasticity is (− 1) × (percent change in sales)/(percent change in price)]
r = minimum expected rate of return demanded by potential adopters of the innovation (in the absence of other information, this might be approximated by the adopting industry's rate of return on investment before taxes)
δ = adopter's time discount rate (in the absence of other information, this might be approximated by the interest rate on bonds in the adopting industry)

In principle, at least, all these variables can be measured or estimated for the innovation and for the industry in which the innovation is to be adopted. Thus, the optimal price profile for the seller of the innovation can be determined, and the market penetration rate can then be derived.

The important point in all these efforts is that they permit the forecaster to go beyond the "brute fact" of certain types of behavior (e.g., Pearl curves or exponential trends) and explain the coefficients of the models on the basis of factors that are plausibly linked to the process of technological change. By identifying causal factors, they permit the forecaster to predict rates of technological change or rates of substitution even when it is known that conditions have changed or even for technologies for which there is no prior history on which to "calibrate" a curve for extrapolation.

8.4 Economic and Social Models

As the preceding section showed, it is possible to develop causal models that include economic as well as technological factors. However, as indicated in Chap. 3 on analogies, there are often other factors, including social, cultural, and political factors, that may hasten or delay the development or adoption of a technology. It may be necessary to include these factors in a model used to forecast a particular technology's change in performance or rate of substitution.

All the models so far developed to take into account these other factors have been computer simulation models rather than the closed-form analytical models presented in the preceding sections. Below, we will examine two different approaches to simulation modeling of technological change. The first, KSIM, has the advantage of simplicity. The second, the differential equations type of model, is more difficult to use than is KSIM but has greater generality and flexibility.

KSIM. Kane (1972) has presented a simulation model, KSIM, that is simple to use and yet powerful enough to analyze many moderately complex real-world problems. KSIM has the following properties:

1. Variables are restricted to the range 0–1. Actually, this is not a severe restriction. In the real world, all variables are bounded above and below. By a change of scale, any real-world variable can be converted to a variable with a range of 0–1.

2. A variable increases or decreases according to whether the net impact of the other variables in the system on it is positive or negative, respectively. A variable may also affect itself, and this is accounted for in computing the net impact.

3. The response of a variable to impacts from other variables decreases as the variable reaches its extremes of 0 or 1, and is a maximum near values of 0.5. Put another way, a variable is most sensitive to impacts from other variables in the midrange of its values and is least sensitive near the extremes.

4. Other things being equal, the impact produced by a variable will be larger, the larger the value of the impacting variable.

5. Complex interactions are decomposed into pairwise interactions, and only the pairwise interactions are specifically incorporated into the model.

It is this last property that produces the most serious limitation on KSIM's ability to model real-world problems. This property means that the strength of impact of one variable on a second variable depends only on the magnitudes of those two variables and is not affected by the magnitude of any other variable. Wakeland (1976) has pointed out that this is not true for many systems that the forecaster might want to model. For instance, because of this property, KSIM is incapable of producing a genuine Pearl curve. Kane has developed more advanced versions of KSIM that are capable of handling such interactions. However, they are beyond the scope of this discussion.

KSIM operates by starting each variable with an initial value, stepping forward a time increment δt, and computing a new value for each variable on the basis of the impacts from all the variables during the time increment. The change in a variable x_i during a time increment δt is given by:

$$x_i(t + \delta t) = x_i(t)^{p_i} \tag{8.18}$$

where the exponent p_i is obtained from:

$$p_i = \frac{1 + \frac{\delta t}{2}\sum_{j=1}^{N}(|\alpha_{ij}| - \alpha_{ij})x_j}{1 + \frac{\delta t}{2}\sum_{j=1}^{N}(|\alpha_{ij}| + \alpha_{ij})x_j} \tag{8.19}$$

Here the α_{ij} are the impacts of x_j on x_i, and δt is the time increment. The form of Eq. (8.19) assures that p_i will always be positive since raising a variable whose value lies between 0 and 1 to any positive power will keep the values within that range. Equation (8.19) can be expressed as follows, to make its purpose more clear:

$$p_i(t) = \frac{1 + \delta t \mid \text{sum of negative impacts on } x_i \mid}{1 + \delta t \mid \text{sum of positive impacts on } x_i \mid} \tag{8.20}$$

When the negative impacts are larger than the positive impacts, the numerator of p_i will be larger than the denominator, p_i will be larger than 1, and $x_i(t + \delta t)$ will be *smaller* than $x_i(t)$. Note that in the numerator, subtracting the cross-impact factors from their absolute values doubles the negative impacts and cancels the positive impacts. In the denominator, where the impact factors are added to their absolute values, the effects are just the opposite. This is the reason for the factor of ½ in both the numerator and the denominator: to restore the impacts to their original size.

From Eq. (8.18), we can see that a given magnitude p_i will have less effect when x_i is near either 0 or 1 than when x_i is near 0.5. From Eq. (8.19), we can see that the larger the value of x_j, the stronger its impact on x_i will be. Finally, condition (4) above simply means that all the effects of the variables on each other are expressed by α_{ij}.

To use KSIM, we need to specify the initial values for each of the variables, their impacts on the others (i.e., the values for each α_{ij} for each pair of variables x_i and x_j), and the impact of the "outside world" on each variable. The outside world is a variable that affects but is not affected by the variables in the system. Formally, this means that the matrix of impact coefficients has one more column than it has rows. The last column represents impacts from the outside world but does not correspond to a variable that can receive impacts from the variables in the system.

The TEKFOR program has KSIM as one of its options. The user is asked for variable names, a one-character plotting symbol for each variable, initial values, and the matrix of α_{ij}s. The program then carries out the simulation.

Before using KSIM, we need to identify the variables that will go into the model, and once these are identified, the interactions among them have to be specified. This procedure is really the heart of the modeling process. KSIM then simply traces out the implications of whatever has been built into the model by specifying the impacts among the variables.

To illustrate this process, let us use KSIM to prepare a forecast of the use of solar energy in homes for space heating and hot water. The system will be described by the following variables:

1. The proportion of homes using solar energy for space heating and hot water. The current value of this variable is 0.1 of its possible maximum.

2. The price of fossil fuel. We take this as currently being 0.3 of its possible maximum.

3. The performance of solar technology for a given cost of installation and operation. This is taken as 0.3 of its possible maximum value (i.e., when fully developed, the performance of a unit of given cost will be over three times today's performance).

4. The amount spent on R&D for solar home-heating technology. This is taken as being 0.2 of its possible maximum value.

Next we consider the impacts on each of these variables.

The proportion of homes using solar energy is affected by itself, in that greater use implies a more widespread availability of technicians, sales outlets, and advertising. It is also affected by the price of fossil fuel, since the higher the price, the more incentive there is to install a solar energy unit. The proportion of homes using solar energy is also affected by higher performance, as this increases its cost effectiveness and makes it more attractive.

The price of fossil fuel is affected negatively by the use of solar energy, since greater use of solar energy decreases the demand for fossil fuel. The price of fossil fuel is affected by itself, as increased price decreases demand and reduces the pressure for further price increases. The price of fossil fuel is affected by demand in the "outside world," since home heating is not the only use for fossil fuel. Its price would rise with increasing scarcity, even without the demand for home heating.

The performance-price ratio of solar energy installations is affected by increased use, as cumulative production brings design improvement (the technological progress function discussed in Chap. 7). It is also affected by R&D, as cumulative R&D brings design improvements.

The amount spent on R&D for solar energy equipment is affected by the degree of use, since firms often tend to spend a fixed percentage of sales on R&D. Hence, the greater the cumulative use, the greater the cumulative expenditures for R&D.

The impacts among the variables are shown in Table 8.4, where they have been assigned numerical weights on a scale from -10 (maximum unfavorable impact) to 10 (maximum favorable impact). To plot the results of the simulation, the variables are assigned the plotting symbols "S" for solar energy use, "F" for fossil fuel price, "P" for the performance of solar installations, and "R" for R&D expenditures. The initial values given above, the impact factors in Table 8.4, and these plotting symbols, must be loaded into the KSIM program.

TABLE 8.4 Matrix of Impacts for KSIM Model

	Variable *j*				
Variable *i*	1	2	3	4	5*
1	3	5	4	0	0
2	- 2	- 2	0	0	0
3	3	0	0	7	0
4	8	0	0	0	0

*Outside world.

With the program set to print out every twenty fifth time increment, the results are shown in Fig. 8.5. As can be seen in the figure, all four variables increase toward their maximum values, but within the time horizon of the simulation, the fossil fuel price did not quite reach the maximum possible value, being only about 94 percent of its maximum.

Several points need to be mentioned about this simulation run. First, the model chosen is quite simple. Other variables might be included. However, they can be readily added to the model if desired.

Second, there may be some disagreement about the initial values of the variables. Is current solar energy use really at 10 percent of the maximum possible? It is relatively simple to test the sensitivity of the outcome to this value by changing the value and running the model again.

Third, there may be disagreement about the structure of the interactions as reflected in the matrix of impacts. Again, the numbers can be changed, and the model can be rerun to determine whether the changes produce significantly different results.

Fourth, KSIM gives results that are only qualitative. The time scale on the printout plot does not correspond to a specific number of years. The purpose of KSIM is to explicate the consequences of the structure of the system. It is not capable of providing a precise numerical forecast.

Fifth, while the initial values can reasonably be scaled to objective data, it appears that the numerical values of the impacts are completely subjective. This is true, but it is not a significant shortcoming of KSIM. As Jay Forrester of MIT has noted, most of the knowledge in the world is not written down; it exists in people's heads, in their intuitive understanding of how the pieces of the world fit together. The only way this intuitive knowledge can be made "public" is in assigning numerical values according to some scale. This quantified knowledge is superior to the original mental model from which it is derived, since it can be compared with the models produced by other people, and the reasons for the differences can be explored. Hence, the subjectivity of the impact matrix in KSIM does not justify rejecting the entire concept.

Despite its simplicity, KSIM allows the user to "get inside" a problem quickly and gain some understanding of the consequences of the perceived interactions. Thus, KSIM is a useful tool for the technological forecaster. It can

```
SOLAR_USE          = S
FOSSIL_PRICE       = F
SOLAR_PERFORMANCE  = P
SOLAR_R&D          = R

    0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1.0
     +....+....+....+....+....+....+....+....+....+....+
   0 |    S    R    *                                   |
  25 |     S   R     PF                                 |
  50 |      S   R     P  F                              |
  75 |        S R      P   F                            |
 100 |         S R      P    F                          |
 125 |           SR      P     F                        |
 150 |             *      P      F                      |
 175 |              R S     P     F                     |
 200 |                R  S    P     F                   |
 225 |                  R   S  P     F                  |
 250 |                    R    S P     F                |
 275 |                      R     SP    F               |
 300 |                        R      *   F              |
 325 |                           R      * F             |
 350 |                             R      *F            |
 375 |                                R    F*           |
 400 |                                  R   FPS         |
 425 |                                    R  F PS       |
 450 |                                      RF   PS     |
 475 |                                        *   PS    |
 500 |                                        F R  PS   |
 525 |                                         F R  *   |
 550 |                                         F  R PS  |
 575 |                                          F  R *  |
 600 |                                          F   RPS |
 625 |                                          F   R * |
 650 |                                           F   R* |
 675 |                                           F   R* |
 700 |                                           F    *S|
 725 |                                            F   R*|
 750 |                                            F   R*|
 775 |                                            F   R*|
 800 |                                            F   R*|
 825 |                                             F   *|
 850 |                                             F   *|
 875 |                                             F   *|
 900 |                                             F   *|
 925 |                                             F   *|
 950 |                                              F  *|
 975 |                                              F  *|
1000 |                                              F  *|
     +....+....+....+....+....+....+....+....+....+....+
    0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1.0
```

Figure 8.5 TEKFOR output of KSIM model run of the solar energy problem.

often be used as a means of gaining the understanding necessary to develop a more complex model.

However, KSIM does have limitations. White (1981) has shown that KSIM can produce an equilibrium solution that is purely an artifact of the specific differential equation implemented in KSIM and has nothing to do with the original model. Therefore, KSIM should not be used to model systems whose dynamics do not correspond to the differential equations underlying KSIM. Mohapatra and Vizayakumar (1989) have shown that KSIM does not adequately distinguish between state variables and rates of change of state variables. Thus, it is possible in KSIM to inadvertently produce closed loops that contain no state variables. These loops will have dynamics that incorrectly represent the true behavior of the system. (In a real system, every closed loop must contain at least one "level" or "accumulative" variable, representing the integral of a rate of change, because of the physical impossibility of measuring instantaneous rates of change.) Thus, despite the utility of KSIM, the forecaster may find that he or she needs to move on to a more complex model. This is the kind that we take up next.

Differential equations models. A differential equations model attempts to write the actual equations describing the rates of change of the variables in the model. In principle, some of these models may be solved in closed form, producing an analytical model. In practice, however, the models are integrated numerically, producing a computer simulation of the behavior of the system being modeled.

As an example of a differential equations model, we examine a model for the diffusion of some innovation among a population of potential adopters. This model is based on the one presented by Sharif and Ramanathan (1983).

We assume that some innovation is being introduced into a population of potential adopters. Depending on the nature of the innovation, these potential adopters may be individuals, households, firms, etc. For the purposes of this model, we need not specify the exact nature of the potential adopters.

At the time of introduction of the innovation, the population is entirely in the "undecided" category, i.e., they have not yet made a decision about whether or not to adopt. Some fraction of the undecided population will become "tryers," who try out the innovation. Of these tryers, some will become "adopters" and some will become "rejecters." As the innovation is improved with the passage of time, some of the rejecters may reconsider and reenter the undecided population. From the outset, other members of the undecided population become "disapprovers," i.e., they not only will not try the innovation, they think it should not be adopted even if it works the way it is supposed to work. The flow of population members among these groups is depicted in Fig. 8.6.

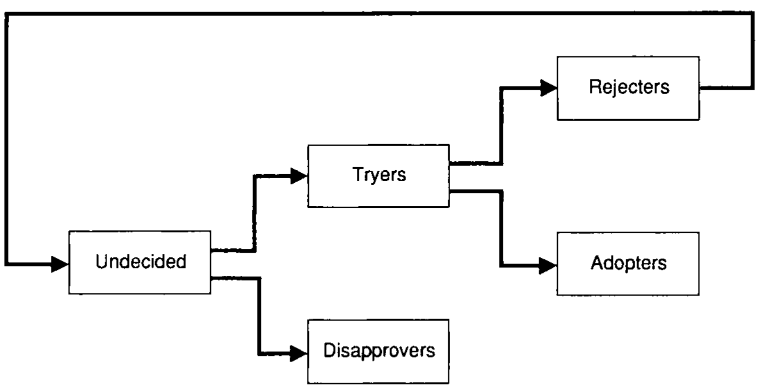

Figure 8.6 Flow of potential adopters of an innovation.

We will now define the variables specifically. Note that four of the variables change by accumulating flows or rates of change. In the model, then, these variables must be obtained by integrating their rates of change. For these variables we must specify an initial value and provide equations for the rates of change. Rather than specify actual numbers for the initial values, we will specify them as a fraction of the relevant population. They are:

Variable	Symbol	Initial value
Undecided	U	0.99
Adopters	A	0.01
Rejecters	R	0
Disapprovers	D	0

Those who are initially undecided can flow into tryers or into disapprovers. Assuming that the logic we used in Chap. 4 to justify a Pearl curve applies here, the number of tryers should be proportional to both the number of undecided and the number of adopters. We will let the proportionality constant be 0.5. Thus, the variable T is defined as in the following table. Note that if either U or A is zero, the number of tryers will remain zero. This is why the number of adopters was started at 0.01, and undecided was started at 0.99 (the sum of the two being the entire population). A Pearl curve cannot "get started" unless there are already some adopters.

The flow from undecided to disapprovers should decrease with time; those who are going to become disapprovers will probably do so early on. The variable Z in the following table expresses the nature of this decline in flow rate.

Some fraction of tryers will become adopters. With the passage of time, this fraction should increase, because the innovation itself will become more attractive with improvements in technology. We model the rate of change of tryers into adopters as the variable X in the following table.

Some fraction of tryers will become rejectors. With the passage of time, this fraction should decrease, for the same reason that the complementary fraction becoming adopters should increase. This flow rate will be the variable Y in the following table. As time passes and the innovation is improved, some rejecters may revert to undecided. The flow rate from rejecter to undecided is given by the variable W in the table.

Variable	Description
T	$0.5 \times U \times A$
W	Starts at zero, tapers linearly to 0.5 at 10 years, then remains at 0.5.
X	Starts at 0.75, tapers linearly to 1.0 at 10 years, then remains at 1.0.
Y	Starts at 0.25, tapers linearly to 0 at 10 years, then remains at 0.
Z	Starts at 0.05, tapers linearly to 0 at 10 years, then remains at 0.

The variables W, X, Y, and Z are plotted in Fig. 8.7.

We can now express the rates of change of the first four variables in terms of these additional variables. These are given by Eqs. (8.21) through (8.24).

$$\frac{dU}{dt} = -T - U \times Z - R \times W \tag{8.21}$$

$$\frac{dA}{dt} = T \times X \tag{8.22}$$

$$\frac{dR}{dt} = T \times Y - R \times W \tag{8.23}$$

$$\frac{dD}{dt} = U \times Z \tag{8.24}$$

These rates of change are integrated to obtain the values of the variables. The results of this computer simulation are shown in Fig. 8.8. The number of adopters follows a growth curve to a value of about 75 percent of the population, the number of disapprovers grows to just under 25 percent of the population, the number of undecided shrinks steadily, and the number of rejecters peaks at just below 1 percent of the population, then shrinks steadily.

The time-varying rate coefficients and the coefficient in the equation for tryers were selected arbitrarily for this example. However, these values could be determined in an actual population either by analogy with "similar" previous innovations or through market research on potential adopters' attitudes and practices.

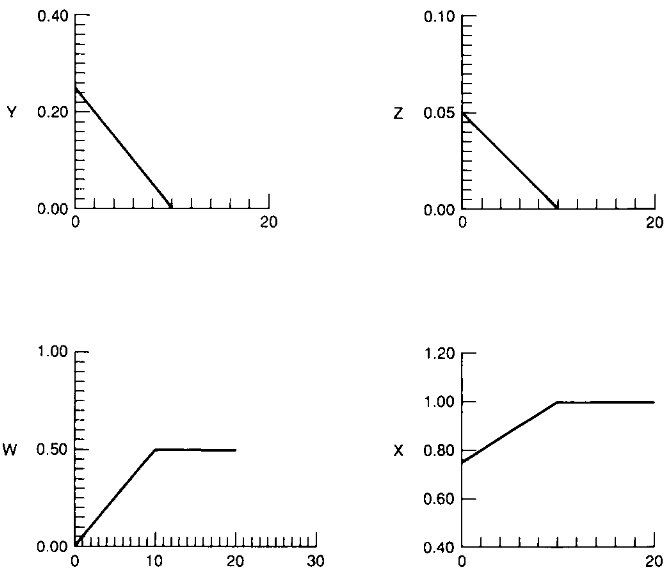

Figure 8.7 Rate coefficients as a function of years after introduction.

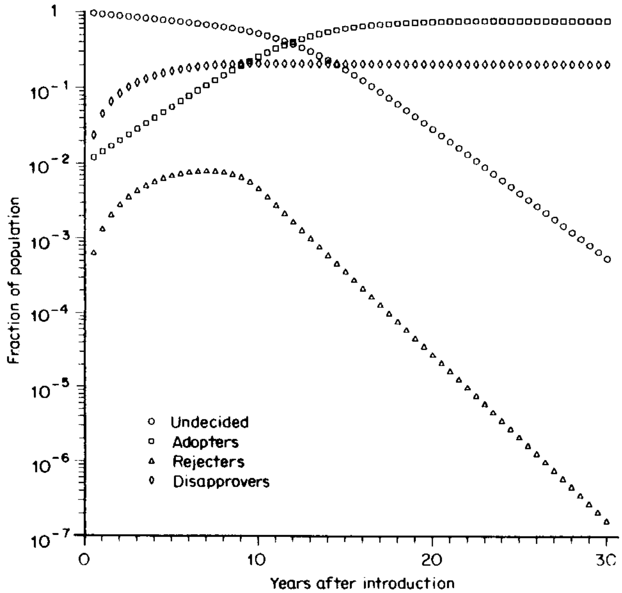

Figure 8.8 Simulation results of diffusion model.

While a model such as this diffusion model is more complex than a KSIM model, it is capable of dealing with a wider range of situations. Any behavior that can be expressed in a set of differential equations can be modeled, and future behavior can be forecast.

8.5 Model Construction

Both the KSIM and differential equations models involve multiple "output" variables. This is generally true of computer simulation models. In this respect, they differ from models such as the Floyd model and from the extrapolation methods of previous chapters. In developing such a model, the forecaster must, in effect, proceed backward. The starting point for model design is the nature of the desired answer. This determines the remainder of the model. In designing a model, the forecaster should ask the following questions, in the order given.

1. What decision is to be made? What change in resource allocation or organizational structure is being considered?

2. What information is needed to make that decision? What does the decision maker need to know in order to make the best choice? How is "best" determined?

3. What variables must be forecast to provide (part of) the needed decision information? The determination of variables to be forecast depends upon the nature of the decision, not vice versa. In most cases, only part of the decision information comes from a technological forecast; hence, the

variables being forecast need provide only that portion of the total decision information.

4. What variables affect the variables to be forecast? What variables in turn affect them? This step must be repeated until the chain of causation has been traced backward, from the variables whose future values are needed to all the variables that affect the "output" variables directly or indirectly.

This sequence of questions produces a list of variables that must be included in the model. Building the model then is a matter of determining the linkages among the variables and expressing these in the appropriate form (e.g., a KSIM matrix or a set of differential equations). The resulting model will usually turn out to be a network, in that the chains of variables, from input to output, are often cross-linked. This complexity is, of course, the reason a computer model was selected as the forecasting tool. In many cases, the network of linked variables will contain feedback (i.e., closed) loops, where a variable is found to affect the variables that in turn affect it. Any model purporting to describe the real world must contain feedback loops since the real world is full of them.

The critical issue is that the starting point for designing a simulation model is the decision to be made and the information needed to make that decision. The model must be designed to provide that information if it is to be useful.

8.6 Summary

Causal models are patterned after the models of the "exact" sciences. The intent is to go beyond simple correlation or extrapolation and to identify causal linkages among events. However, a major problem with causal models of socioeconomic activities is that there is no essential reason for people to behave in a fixed way. In particular, there is no reason why technologists must behave in the ways presumed by causal models of technological change. Thus, although an explanatory model fits past behavior well, there is no guarantee that it will fit future behavior.

In addition to the lack of "necessity," models of technological change suffer from another shortcoming. They are based on a hypothetical structure for the workings of the technology-producing system, which is described in a mathematical model. Coefficients of the model are then determined from past data, and the completed model is used for forecasting. At the very least, this past data contains errors of rounding and aggregation. It may also contain errors of mistranscription or miscopying.

Finally, some data may be missing through distortion or a deliberate refusal to disclose them. Hence, the coefficients of the model are in error to the degree that the data are in error. Even if the theory behind the model is correct, the model will produce erroneous forecasts to the extent that it was fitted to erroneous data.

Schoeffler (1955) described these and other errors of models in detail. Although he wrote of economic models, what he said applies equally well to causal models of technological change.

Despite these shortcomings, causal models remain an important tool of the technological forecaster. They permit the forecaster to take into account known changes in the factors that have produced technological progress in the past. They also allow the forecaster to explore the consequences of alternative policy choices and to examine the sensitivity of forecasts to particular factors whose values are uncertain.

References

Ayers, R. U., "A Schumpeterian Model of Substitution," *Technological Forecasting & Social Change*, **27**:375–383 (1985).

Blackman, A. W. Jr., E. J. Seligman, and G. C. Sogliero, "An Innovation Index Based on Factor Analysis," *Technological Forecasting & Social Change*, **4**:301–316 (1973).

Day, E. D., "Expected Utilities and Substitution Rates," *Technological Forecasting & Social Change*, **24**:299–312 (1983).

Floyd, A. L., "A Methodology for Trend Forecasting of Figures of Merit," in J. R. Bright (ed.), *Technological Forecasting for Industry and Government*, Prentice-Hall, 95–109, Englewood Cliffs, NJ, 1968.

Kane, J., "A Primer for a New Cross-Impact Language—KSIM," *Technological Forecasting & Social Change*, **4**:129–142 (1972).

Lakhani, H., "Empirical Implications of Mathematical Functions Used to Analyze Market Penetration of New Products," *Technological Forecasting & Social Change*, **15**:147–156 (1979).

Mansfield, E., *The Economics of Technological Change*, Norton, New York, NY, 1968*a*.

Mansfield, E., *Industrial Research and Technological Innovation*, Norton, New York, NY, 1968*b*.

Martino, J. P., K.-L. Chen, and R. C. Lenz, Jr., "Predicting the Diffusion Rate of Industrial Innovations," University of Dayton Research Institute Technical Report UDRI-TR-78-42 (1978). (Available from National Technical Information Service as PB-286693).

Mohapatra, P. K. J., and K. Vizayakumar, "Revisiting Causality in System Dynamics and KSIM Models," *Technological Forecasting & Social Change*, **36**:363–387 (1989).

Rogers, E. M., and F. F. Shoemaker, *Communication of Innovations*, The Free Press, New York, NY, 1971.

Schoeffler, S., *The Failures of Economics: A Diagnostic Study*, Harvard University Press, Cambridge, MA, 1955.

Sharif, M. N., and K. Ramanathan, "Polynomial Innovation Diffusion Models," *Technological Forecasting & Social Change*, **21**:301–323 (1983).

Stier, J. C., "Technological Substitution in the United States Pulp and Paper Industry," *Technological Forecasting & Social Change*, **23**:237–245 (1983).

Wakeland, W., "QSIM2: A Low-Budget Heuristic Approach to Modeling and Forecasting," *Technological Forecasting & Social Change*, **9**:213–229 (1976).

White, P. K., "The Equilibria of KSIM Modes: Existence, Stability, and Implications," *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-11(8)**:564–570 (August, 1981).

For Further Reference

Technological Forecasting & Social Change, **9(1, 2)** (1976). The entire issue is devoted to the use of models in technological forecasting, and many advanced applications are discussed.

Technological Forecasting & Social Change, **14(4)** (1979). This issue is devoted to structural modeling in technological forecasting, including several papers that treat applications of KSIM.

Problems

1 Using Floyd's model and the data on the efficiency of incandescent lamps from Table A.40, fit the model to the data from 1888 through 1940 and forecast the efficiency for 1959. Use the value of 19 lm/W, calculated in Chap. 4 as the upper limit on efficiency of incandescent lamps, for the maximum possible efficiency. Use the efficiency of a candle from Table A.6 as the competing technology.

2 Use KSIM to forecast the rate of adoption of the electric car (fraction of total number of cars on the road). Include the following variables in your model:

Price of gasoline
Price of electricity
Cumulative R&D on batteries
Price of electric cars
Price of gasoline-powered cars
Real per-capita income (i.e., with inflation removed)
Fraction of population living in urban areas

Add other variables if you believe that they are important. Estimate the current (initial) values for each of these variables. Estimate the entries in the matrix of impacts, and run this baseline model to see the consequences of your estimates. Next test the sensitivity of the results of your estimates by making the following changes, one at a time, and rerunning the model.

Change the initial value of one variable by a factor of two.
Incorporate a policy intervention (impact from the outside world) on battery R&D.
Change your initial estimate by a factor of two if you initially incorporated that impact; incorporate it with a value of five if you did not initially incorporate it.

Chapter

9

Probabilistic Methods

9.1 Introduction

The forecasting methods presented in the preceding chapters have all produced point estimates of future conditions. Moreover, the forecast methods assumed that the factors operating to produce the time series (level of functional capability or adoption of an innovation) were deterministic in their effect, even if their magnitude varied with time. Given a set of historical data, there was only one possible future value for the parameter of interest. The only concession to uncertainty was confidence bounds placed on extrapolations generated by least-squares fitting.

In contrast, a probabilistic forecast utilizes different assumptions. One type of probabilistic forecast might give a range of possible future values and the probability distribution over that range. Another possible type of probabilistic forecast might be based on a probability distribution for the magnitude or effectiveness of the factors operating to produce change. The essential point is that a probabilistic forecast involves a range of values and a probability distribution over that range.

Probabilistic methods are quite new in technological forecasting and have not been refined to the same extent as the deterministic methods presented in earlier chapters. However, probabilistic methods do have some advantages in particular situations, and these methods are presented in this chapter.

9.2 The Distribution of Discoveries

Before looking at probabilistic forecasts of performance or adoption of individual technologies, we will examine some results on the probabilistic nature of technological discoveries.

Sahal (1983) has examined the frequency of occurrence of discoveries, by which he means both invention (the development of something new) and innovation (the marketing of a previous invention). His analysis was based on a set of 62 inventions made between 1874 and 1961 for which he was able to obtain both the year of invention and the year of innovation. His data are presented in Table A.55.

Sahal argued that discoveries involve both chance and accumulated learning. He proposed that the negative binomial distribution was appropriate for describing situations involving both a cumulative mechanism and random chance. Thus, he postulated that the probability of having x discoveries in a single year is given by:

$$P(x) = \left[1 + \frac{m}{k}\right]^{-k} \frac{(x + k + 1)!}{x!(k - 1)!} \left[\frac{m}{(m + 1)!}\right]^{x} \qquad (9.1)$$

where m is the mean of the distribution and k is a parameter. The best estimate of the mean is the sample mean. The parameter k is estimated by iteratively solving the following equation:

$$\frac{N_0}{N} = \left[\frac{m}{m + k}\right]^{k} \qquad (9.2)$$

where N_0 is the number of years in which there were no discoveries out of a total number of N years. The probabilities for different values of x can then be obtained recursively from the equations:

$$P(0) = \left[\frac{m}{m + k}\right]^{k} \qquad (9.3)$$

and

$$P(x) = \left[\frac{m}{m + k}\right]\left[1 - \frac{1 - k}{x}\right]P(x - 1) \qquad (9.4)$$

For Sahal's data, the computed parameters are given in Tables 9.1 and 9.2. The negative binomial appears to give an excellent fit to the data on inventions and innovations.

Sahal observed that one major implication of this finding is that discoveries (both inventions and innovations) do not occur at a constant rate. Instead, they tend to cluster, with several occurring in the same year. Most years in the time interval covered had no discoveries. However, there were 7 years in which there were 2 or more inventions, and 6 years in which there were 3 or more innovations. Utilizing these results, it would be possible to forecast the number of discoveries to be expected in a particular year. However, the forecast would not be a single number, but a statement of the probability of 0, 1, 2, or 3 + discoveries in a single year.

TABLE 9.1 The Distribution of the Number of Major Inventions per Year, 1850 through 1970

	Number of years	
Number per year	Observed	Theoretical*
0	81	81.02
1	27	25.9
2	7	9
3 +	6	5.01
Total	121	120.9

*Mean (m) = 0.51, exponent k = 0.85, chi-squared = 0.78, P = 0.40.

TABLE 9.2 The Distribution of the Number of Innovations per Year, 1850–1970

	Number of years	
Number per year	Observed	Theoretical*
0	83	82.92
1	21	23.76
2	11	8.64
3 +	6	5.59
Total	121	120.9

*Mean (m) = 0.51, exponent k = 0.65, chi-squared = 0.96, P = 0.36.

Since Sahal had information on both the invention and the marketing of the same discovery, he also examined the "gestation period" between invention and innovation. He argued that the bringing of an invention to market is likely to involve many different factors, including both those related to getting the invention ready for market (debugging, refining, etc.) and those related to obtaining the necessary resources (production machinery, advertising, and distribution). This multiplicity of factors suggests that the gestation period should vary probabilistically. Sahal proposed that the number of inventions x requiring a gestation period equal to or greater than y is given by:

$$y = cx^{-p} \tag{9.5}$$

where c and p are parameters of the distribution. This is the Pareto distribution (Johnson and Kotz, 1970, Chap. 19). Equation (9.5) can be linearized by taking the logarithms of both sides:

$$\log y = \log c - p \log x \tag{9.6}$$

If the gestation periods are then ranked in descending order of length (longest is rank 1, second-longest is rank 2, etc.), and the number of years for gestation is plotted against rank on log-log paper, we would expect that gestation periods that have a Pareto distribution would fall on a straight line with slope $-p$. This was the case for Sahal's set of discoveries except for the higher ranks (30 and over out of 62 discoveries) whose slope tended to be steeper than $-p$, i.e., the higher-ranked innovations (those with shorter gestation periods) actually had a slightly shorter gestation period than would be expected from Eq. (9.5).

A consequence of Eq. (9.5) is that the rank-time product for innovations will tend to be a constant, i.e., the second-longest gestation period will tend to be half the longest, etc. This rank-time product may vary from one industry to another, but would be expected to be constant for a particular industry. For instance, Sahal found that for the telecommunications industry, the rank-time product turned out to be approximately 39. The longest gestation period was 39 years (magnetic tape recording); the tenth-longest was 3 years (long-playing record).

Both invention and innovation are the outcome of the operation of many independent factors. The consequence of this is that both processes are random or probabilistic in nature. However, they do turn out to be described by particular probabilistic behavior patterns. The numbers of inventions and innovations per year seem to follow a negative binomial distribution. The gestation periods from invention to innovation seem to follow a Pareto distribution. Hence, even though particular events of invention or innovation may not be forecastable, it is still possible to say something useful about the rate of occurrence of inventions and innovations, and about the lag from one to the other. What can be said, however, is only probabilistic.

9.3 A Probabilistic Biological Analogy

Kwasnicka et al. (1983) have proposed a probabilistic model for multiple substitution based on a biological model. They assumed that several technologies (or products) compete in a market with limited demand (which may vary with time). Each product has a quality index ("fitness") that reflects factors such as cost of production, market price, performance, reliability, and any other factors relevant in that particular market. If a theoretical basis for combining these factors into a single index does not exist, a scoring model may be used. In any case, it is assumed that such an index is available for each technology. The idea is that those technologies with the greatest "fitness" at one period will "reproduce" more rapidly than those with less "fitness" and will, therefore, be a larger fraction of the population at the next time period. However, this increase is not deterministic, rather it is probabilistic.

The model as presented by the authors is quite complex. To illustrate the basic ideas, a simplified version will be presented here.

We assume that we wish to forecast the respective market shares of three competing technical approaches. T_1, the "old" approach, starts with a dominant market share. The other two, T_2 and T_3, have small but different initial market shares. We assume that in each time period t_i, the sales of technical approach T_j is a random variable whose expected value is a function of both its sales at t_{i-1} and its fitness.

The authors assumed that sales are governed either by a gamma distribution or by a Poisson distribution. In either case, the parameters of the distributions are dependent on previous sales and on fitness.

For simplicity, we will assume that the fractional change in a technical approach's sales from one period to the next is described by a uniform distribution. The mean of that distribution will be the product of the approach's market share and the deviation of its fitness from the average fitness of all sales in the preceding period. The range of the uniform distribution will be 0.2 (i.e., the mean $\pm$ 0.1).

We assume the following values for the three technical approaches:

	T_1	T_2	T_3
Initial sales, x_0	85	5	10
Fitness, f	0.2	0.6	0.5
Initial market share, %	0.85	0.05	0.10

At t_0, the market-share-weighted average fitness is 0.25. Thus, the deviation of T_1 from the average fitness 0.2 − 0.25 or −0.05; that of T_2 is 0.35; and that of T_3 is 0.25. Thus, the mean fractional changes in t_1 for T_1 is -0.05×0.85 or -0.0425; for T_2, 0.0175; and for T_3, 0.025. The actual fractional changes will be random numbers drawn from a range of ± 0.1 about these means. For T_1, the sales in t_1 will be a random number drawn from the range $(85 - 85 \times (-0.0425 - .1) = 72.9,\ 85 + 85 \times (-.0425 + .1) = 89.9)$; for T_2, the sales will be in the range (4.6, 5.6); and for T_3, the sales will be in the range (9.25, 11.25). Note that the ranges for the second period depend upon the random outcome of the changes for the first period, and cannot be computed until the first period outcome is known.

What kind of outcome would we expect, given the information in the table above? Clearly T_1 is inferior to the other two in fitness; it will lose market share to one or both of them. In the long run we would expect T_2 to dominate the other two because of its superior fitness. However, it starts with a lower market share than does T_3. This might allow T_3 to capture some of the market held by T_1.

In order to determine the outcomes, we must conduct a computer simulation of this system. This simulation is known as a Monte Carlo simulation, since some of the transitions from one state to the next depend upon random outcomes. Thus, a computer program is needed that draws random numbers as needed and makes the appropriate computations.

The results of 30 such simulation runs are shown in Figs. 9.1 through 9.4. Figure 9.1 shows the market share of technical approach T_1. The solid line is the average market share over 30 simulation runs. The dashed lines show the average plus and minus one standard deviation, computed over the 30 runs. The market share starts at 85 percent, then very quickly drops to about 10 percent, followed by a slow taper toward 0. Since sales cannot be negative, the lower confidence bound (mean minus one standard deviation) reaches 0 and stays there after about $t = 4$.

Figure 9.2 shows the mean market share for T_2 over 30 simulation runs and one standard deviation about the mean. Because market share cannot exceed 100 percent, the confidence bound is flattened at a value of 1.0. Since T_2 was the fittest technical approach, it is reasonable to expect that it would dominate the market.

Figure 9.3 shows the mean market share for T_3 over 30 simulation runs and one standard deviation about the mean. Note that its mean market share rises quickly to about 50 percent, then drops off to less than 10 percent. Because of

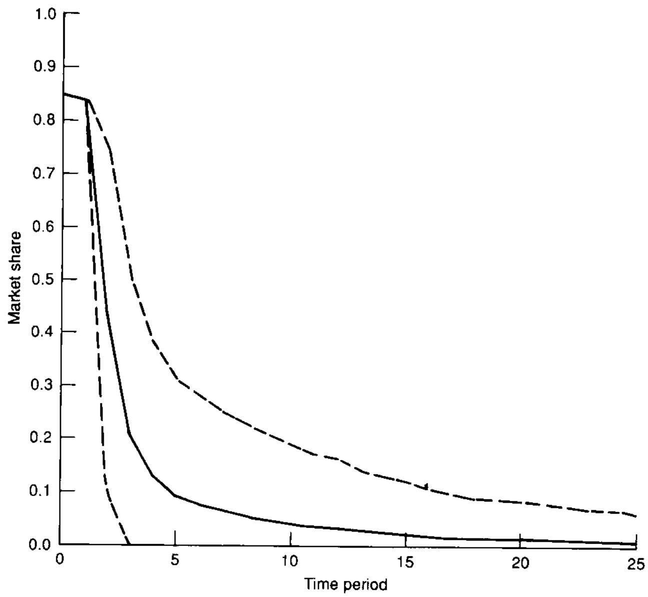

Figure 9.1 Market share of technical approach T_1.

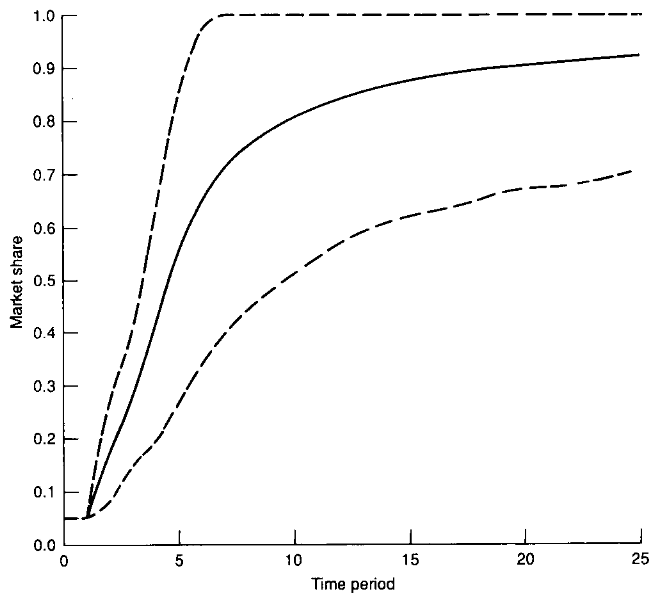

Figure 9.2 Market share of technical approach T_2.

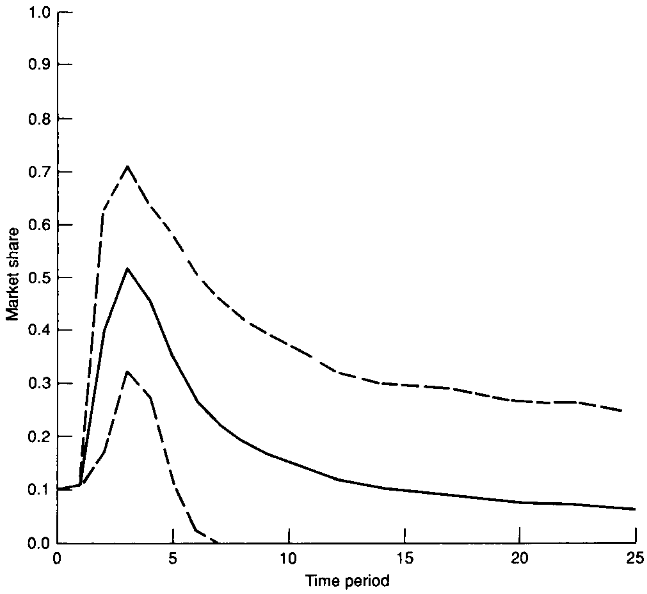

Figure 9.3 Market share of technical approach T_3.

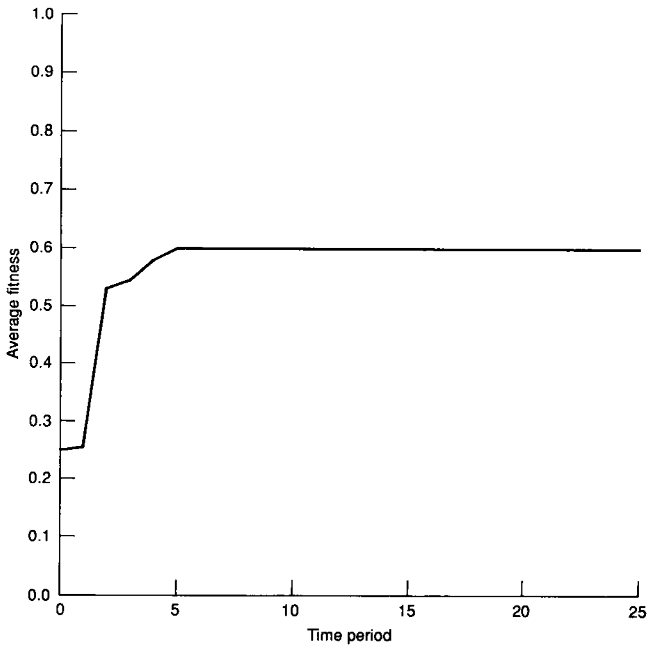

Figure 9.4 Average fitness of all the technical approaches.

the superiority of T_3 over T_1, it quickly captures market share from the latter. However, in the long run it cannot compete with T_2. Thus, T_3 is driven out of the market almost as quickly as is T_1.

Figure 9.4 shows the average fitness of all the items sold in the market. As T_2 comes to dominate the market, the average fitness rises to equal the fitness of T_2, or 0.6. Perhaps the only surprising thing about this figure is the rapidity with which the average fitness rises to its maximum value.

This illustrates the use of a probabilistic approach to modeling the competition of several technical approaches in the same market. The principle can be applied to other similar situations. The critical issue is to identify the factors that are important in the competition (i.e., the "fitness") and determine how they establish the parameters of the probability distribution that describes the competition. Once this is done, simulating the behavior of the system is fairly straightforward.

The program that was used to carry out the Monte Carlo simulation of this system is presented in App. 9A of this chapter. The primary use of this program listing is to serve as a model for application programs in specific circumstances. In particular, the function RAN_NUM would have to be replaced with a function that utilized the appropriate probability distribution (see App. 9C of this chapter on how to approximate probability distributions). The calls to that function would likewise have to be replaced with statements that passed the necessary variables to the function. The important point is that this program is an example of what needs to be done. It is not directly usable in actual forecasting situations.

9.4 Stochastic Projection

Extrapolation methods of forecasting make a deterministic projection of the past performance of a technology. Confidence bounds may then be placed on this projection to account for the uncertainty inherent in fitting a curve to the data. However, an alternative approach is to incorporate the randomness directly by making a stochastic projection instead of a deterministic one. (The material in this section is based on the papers by Murthy, 1979*a* and *b*, and by Murthy and Staib, 1984, who pioneered this approach.)

In the deterministic approach, the historical data is characterized in the form of parameters for some curve (e.g., the location and rate parameters for a growth curve or the initial value and growth rate for an exponential trend). In the stochastic approach, the historical data is instead characterized in the form of parameters of a probability distribution, such as the mean and standard deviation. Once these parameters are available, the forecast is obtained by drawing future changes in performance value from the appropriate distribution(s).

We will illustrate this by forecasting aircraft speed records. Figure 9.5 displays official speed records for aircraft from the period 1906 through 1939 (circles). A change in regime seems to have taken place about 1913; hence, only those records set from 1913 through 1939 will be used to forecast post-1939

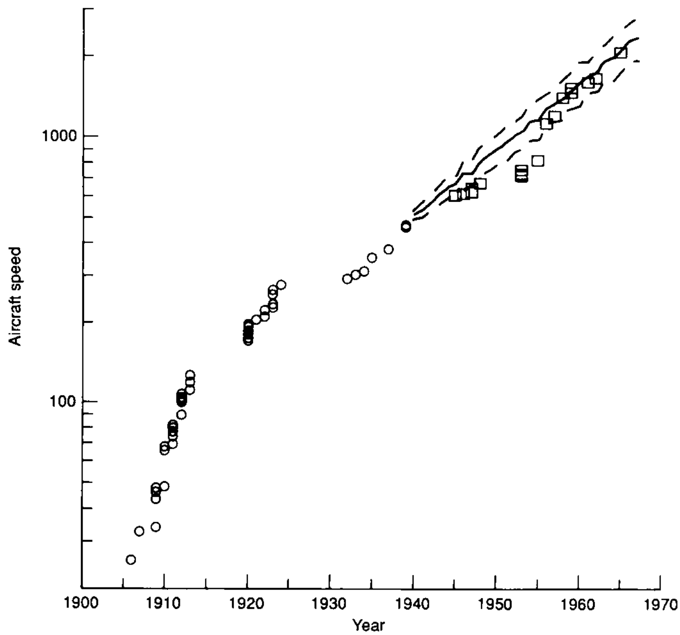

Figure 9.5 Actual and forecasted aircraft speed records.

speed records. The choice of 1939 as the end point is deliberate. The speed record set in 1939 was the last one set by a propeller-driven aircraft. All speed records subsequent to that were set by jet-propelled aircraft. The question to be considered is, can data for propeller-driven aircraft be used to generate a successful forecast for jet-propelled aircraft?

First, we obtain the intervals between successive speed records for the period 1913 to 1939. This is done by converting the day, month, and year of a record into a year and decimal fraction of a year. The intervals then can be computed as years and decimal fractions of a year.

Next, we obtain the speed changes from one record to the next. Instead of the *difference* between successive record speeds we take the *ratio* of successive record speeds. (Note that this implies an exponential growth of aircraft speed.) We next compute the mean and standard deviation for time intervals and for speed ratios.

An oversimplified approach to stochastic projection of aircraft speed records might be done as follows: We choose probability distributions that fit the observed distributions of interval and speed, and adjust them to have the computed means and standard deviations. Then we simply draw intervals and speed ratios from those distributions. Each step ahead is made by adding the randomly drawn interval to the previous time and multiplying the previous speed by the randomly drawn speed ratio.

The thing that is wrong with this approach is that time intervals and speed ratios are correlated. The longer the time interval between attempts at record setting, the greater the increase in speed. Hence, we need to account for this correlation in constructing our stochastic projection. In fact, the correlation coefficient between the two is 0.565. The length of the interval, thus, explains $0.565^2 = 0.319$ or about 32 percent of the variance in speed ratios.

We take the correlation into account by regressing speed ratios on the time intervals. This turns out to be given by:

$$\text{Speed ratio} = 1.039 + 0.022 \times \text{interval} \tag{9.7}$$

This gives us the correlated portion of the speed increase. The time interval will be drawn randomly from a probability distribution. Once it is known, the correlated portion of the speed ratio is fixed. The *residuals* of the regression then provide us with the random portion of the speed increase.

We compute the standard deviation of the residuals (their mean is by definition 0) and fit a probability distribution to the residuals. We likewise fit a probability distribution to the time intervals. In this example, the observed distributions of the time intervals and the residuals were approximated by straight-line segments (see App. 9C). These are easy to integrate in closed form and, hence, can readily be adjusted to have the proper means and standard deviations. Moreover, the analytical forms, which give cumulative probability as a function of the independent variable, can readily be inverted to give the independent variable as a function of cumulative probability.

We then begin with the last 1939 speed record, i.e., both the speed and date. We draw a time interval from the appropriate distribution, and compute the correlated portion of the speed ratio. We then draw a residual from the appropriate distribution to get the uncorrelated portion of the speed ratio. We add this algebraically to the correlated portion of the speed ratio and obtain the complete speed ratio. We, thus, have a new "simulated record" at a time given by adding the interval to the 1939 date and multiplying the 1939 speed by the ratio. This process can be continued, adding each interval to the last simulated record date and multiplying the last simulated record speed by the ratio, until we have projected speed records as far ahead as we wish.

Carrying out this process once gives a simulated future trajectory of the speed records, but this is only one possible outcome of a stochastic process. To gain some knowledge of the variability of the process, we can repeat it some suitably large number of times and compute the mean and standard deviation of simulated speed records for each "future" year.

This has been done fifty times, and the results are displayed in Fig. 9.5. The actual speed records for the years 1940 to 1965 are shown as squares. The mean simulated speed record for each year, 1940 to 1965, is shown as a solid line. The line has some wiggles in it. These arise from the fact that if a randomly drawn time interval is longer than a year, some years may be "skipped" in a particular simulation run. Hence, the results will not be smooth. The dashed lines show one-standard-deviation bounds above and below the mean. The simulation was produced by a program similar in principle to the one in App. 9A.

We note that for the period 1955 through 1965, the mean forecast agrees very well with actual results. The actual speed records are not only within the one-standard-deviation confidence bounds, they are quite close to the mean projection. However, the speed records for 1945 through 1955 fall well below the lower confidence bound. One possible explanation for this behavior is that during that time, all speed records were set by military aircraft. Military leaders, particularly those in the U.S. Air Force, were reluctant to allow official records to be established by operational aircraft, as this would reveal the performance of those aircraft. Hence, the official records tended to be below the actual capabilities of then-current first-line aircraft. The speed record set in 1962 by the Soviet Union, for instance, was announced at the time as having been set by an experimental aircraft. Only much later did it become known that the record had actually been set by a combat aircraft that had been stripped of its weapons to allow it to achieve its maximum possible speed. In the 1960s, as work began on commercial supersonic aircraft, little was to be gained by concealing the speed of combat aircraft; hence, official speed records more closely matched actual aircraft capabilities.

Why did record-setting speeds essentially come to a halt after the 1960s? Heppenheimer (1991) presented one explanation for this. There was no longer any commercial or military utility to be obtained from higher aircraft speeds; hence, there was no incentive to build transports or fighter aircraft that could fly faster than about 2000 mi/h. Moreover, the cost of building a "hot rod" aircraft, simply for the sake of setting a record, was too large for any nonmilitary or noncommercial organization to pay.

The stochastic projection approach provides an alternative to deterministic projections. To make a stochastic projection, the forecaster must characterize the historical data in statistical terms, e.g., the mean, standard deviation, and probability distributions. The forecast is obtained by repeatedly drawing values from the appropriate distributions then stepping the forecast ahead on the basis of each "draw." To obtain the mean forecast and an estimate of the uncertainty in the mean forecast, the process must be repeated many times, and the statistics of the projections must be computed.

9.5 Distribution of Time Lags

In Chap. 7, we examined lead-lag correlations as a method of forecasting. One time series was assumed to lead another time series by a known amount. The occurrence of an event in the first time series was taken to be advanced warning of an event in the second time series.

In this section, we will relax the requirement that the leading time series have a fixed time relationship with the lagging series. Instead we will allow a probabilistic relationship. Given an event from the leading time series, there may be several possible lags, each with its own probability, i.e., instead of assuming a fixed lag, we obtain a probability distribution over the possible range of lags.

This can be illustrated by an example. Table B.12 lists time of commercialization of several aluminum alloys and the first flight of the first aircraft utilizing each alloy. The lag time between commercialization and application can thus be determined for each alloy. If a new aluminum alloy has just been commercialized, we would like to be able to say something about the lag time we might expect before the alloy is used in an aircraft.

To put this question in context, consider a manufacturer of commercial transport aircraft whose current models, introduced just a few years ago, are selling well. The chief competitor has just announced it will develop a passenger transport, based on current technology, to be available in 5 years. The fuel consumption of the new aircraft will be 5 to 10 percent lower than that of our aircraft manufacturer's current models. Since fuel costs make up about half the direct operating cost of a transport aircraft, even a percentage point or two improvement in fuel economy can add up to significant amounts over the life of the aircraft.

Our aircraft manufacturer is then faced with the decision of what to do to remain competitive. The manufacturer's technological forecasters have been tracking several new technologies that, in combination, promise to reduce fuel consumption by 60 percent over current aircraft models. Should the manufacturer plan to introduce a "me-too" model in 5 years, based on the same technology the competitor will use, or instead should the manufacturer plan for a dramatically improved model a few years later? (We assume the manufacturer cannot do both; the capital required is beyond its means.)

To answer this question, it is necessary to know when the new technologies will be available. However, a forecast of their availability cannot be made as a single date for each technology. Such a forecast would inevitably be off to some degree. The forecast must define the risk involved in waiting. That is, the forecast must present the probability distribution for the availability of each of the new technologies.

One of the new technologies is an aluminum alloy that has just been commercialized. Its strength is equivalent to those of current alloys but is 10 percent lighter. Can we make use of the historical data on aluminum alloys to generate a probability distribution for the new alloy? On the assumption that the process that generated the lag times in the past is still operative, the answer is yes.

The issue in developing a probability distribution is to account properly for both knowledge and ignorance. By assuming a particular probability distribution (e.g., normal, gamma, or beta) and fitting it to the available data, the forecaster may be assuming knowledge where there is really ignorance. A method is required that produces a probability distribution consistent with the available data but does not go beyond the data, and that generates a distribution incorporating both the forecaster's knowledge and ignorance.

Such a method is the *method of maximum entropy* (Tribus, 1969). The "entropy," E, of a discrete probability distribution is given by:

$$E = -\sum_i \mathrm{p}_i \ln(\mathrm{p}_i) \tag{9.8}$$

In this equation, the p_i are the probabilities associated with lags of i years, with i ranging from 0 to some maximum value (for computational purposes, the time lags will be taken as $i + 0.5$, to put them at the midpoint of each year).

The entropy is a measure of the *lack* of information in a probability distribution. The broader and more diffuse a distribution, the greater the entropy. If all the probability is concentrated at a single point in time (i.e., all the p_i are 0 except for one equal to 1.0), the entropy is zero.

To properly reflect our ignorance, we want the broadest, or highest entropy, distribution that is consistent with what we know about the behavior of the lags. For instance, we can use the historical data to compute that the mean lag from commercialization to first flight is 4.08 years, and the variance in lags is 6.1009 years. These values place constraints on the probability distribution of the lags. We want a distribution that has no other constraints imposed on it besides these (e.g., we do not assume a particular form, nor do we make assumptions about other parameters that we do not know). The distribution, then, should be as broad as possible while having these known parameters. That would be the maximum entropy distribution, subject to the constraints that it have a specified mean and variance.

Tribus presented a more detailed discussion of the method of maximum entropy. He presented a method for computing the maximum entropy distribution, utilizing Lagrange multipliers. A simpler, although somewhat slower method, is based on the Nelder-Mead optimization procedure described by Nash (1990). This method is suitable for programming on desktop computers and usually converges to the desired distribution within a few hundred iterations, taking only a few minutes. The algorithm is described in App. 9B.

The results of using the Nelder-Mead algorithm on the aluminum alloy data are presented in Table 9.3. The individual and cumulative probabilities are also plotted in Fig. 9.6. The distribution is fairly uniform over most of the range of years. The peak at 4.5 years comes close to the actual peak from 3 to 6 years lag in the original data. The cumulative probability reaches 50 per-

TABLE 9.3 Probabilities for Lag of First Flight after Commercialization of Aluminum Alloys*

Lag time, years	Individual probability	Cumulative probability
0.5	0.084	0.084
1.5	0.040	0.124
2.5	0.049	0.173
3.5	0.039	0.212
4.5	0.536	0.748
5.5	0.037	0.785
6.5	0.040	0.625
7.5	0.042	0.867
8.5	0.073	0.941
9.5	0.008	0.948
10.5	0.028	0.976
11.5	0.014	0.990
12.5	0.010	1.000

*The original and computed means are 4.080 and 4.827, respectively, and the original and computed variances are 6.101 and 6.100, respectively.

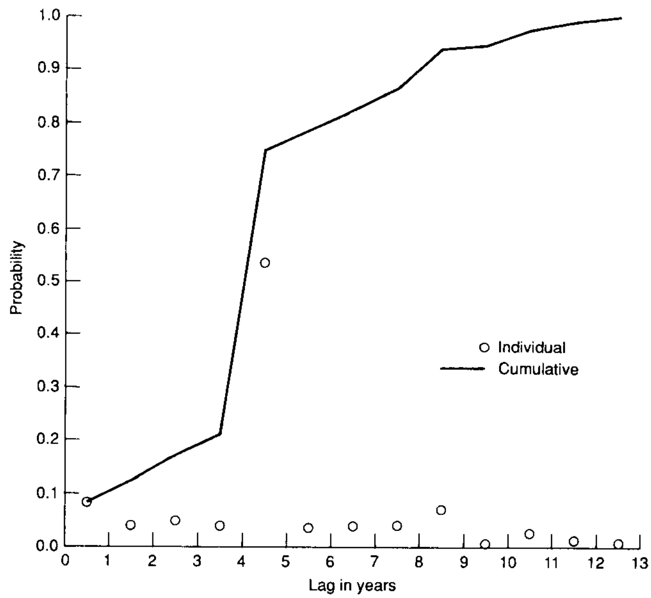

Figure 9.6 Probability for aluminum alloy lags.

cent in the fourth year after commercialization and exceeds 90 percent in the eighth year.

In this case, our hypothetical aircraft manufacturer might reason that it has roughly a 50:50 chance that the new alloy will be sufficiently tested and ready for a first flight about the same time the competitor is bringing a lesser-technology aircraft on the market. The manufacturer could then tell its potential customers that they can have a much-improved aircraft with only a short additional wait. However, if a 50:50 chance is too much of a risk, then the manufacturer is best advised to start work on a "me-too" aircraft in the immediate future. The point is that by using the historical data to generate a probability distribution, the risks of sticking with the current technology or switching to the new can be quantified. This quantification of risk allows a better decision to be made.

In Chap. 7, we considered data on automotive innovations. We observed that after feasibility of an innovation is demonstrated, the innovation is first applied in a prestige car and later appears in a mass-market car. Thus, demonstration of an automotive innovation can be taken as a precursor event for mass-market use. Likewise, appearance in a prestige car can be taken as a shorter-term precursor for mass-market use. Figure 9.7 shows the maximum entropy distribution of lags from demonstration to mass-market introduction. The median lag is about 10 years. Note that the distribution has a peak at 18.5 years, which corresponds to a peak in the actual data from about 16 to 19 years.

Figure 9.8 shows the maximum entropy distribution of lags from use in a prestige vehicle to mass market use. The median lag is about 4 years. The distribution has a peak at 11.5 years, corresponding roughly to the peak in the

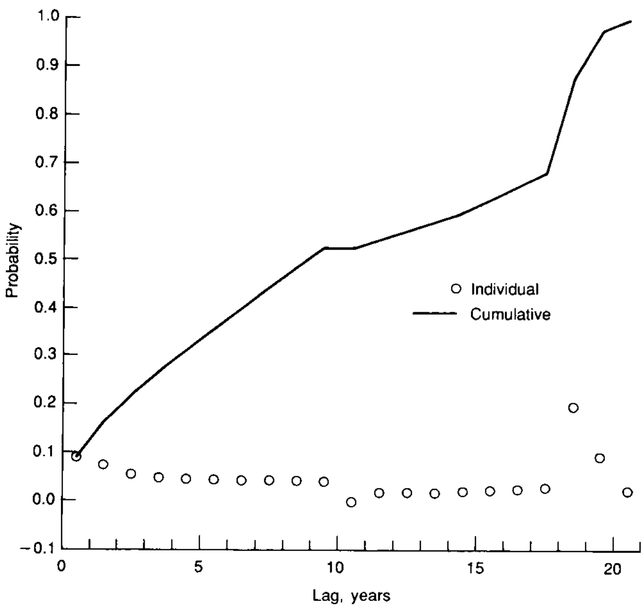

Figure 9.7 Automotive innovation lag, demonstration to mass market.

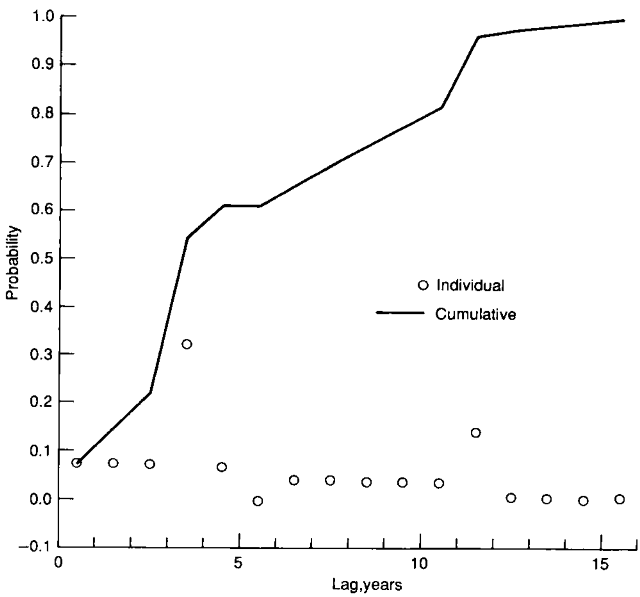

Figure 9.8 Automotive innovation lag, prestige car to mass market.

actual data at 12 years lag for fuel injection. (Time lags for the aluminum alloys and the automotive innovations were computed using the Nelder-Mead algorithm.)

This approach, of computing the distribution of time lags, is of general use. The forecaster must obtain the historical time lags for earlier instances of "similar" innovations, compute the mean and variance, and use these to compute the maximum entropy distribution subject to these constraints. The resulting probability distribution then describes the likelihood of lags of different amounts. The cumulative distribution can be used to estimate the risk associated with waiting for a technology for which a precursor event has already been observed. (Material in this section was drawn from Martino (1987) and Martino (1993).)

9.6 Stochastic Cellular Automata

Cellular automata were introduced by John von Neumann to model self-reproductive systems. They have since then been extensively studied as models of living systems. A cellular automaton can be visualized as a rectangular array of cells. Each cell can exist in more than one state. The system changes state at discrete time points. It is usually assumed that the fate of each cell is linked with, or determined by, the states of its immediate neighbors, i.e., influences on cells are purely local.

An example of a widely studied cellular automaton is the "Game of Life," invented by John Conway. In this automaton, the cells in the rectangular array are either alive or dead. The fate of each cell is determined by its own status and that of its neighbors. A nonliving cell comes to life if it has three living neighbors. A living cell continues to live so long as it has two or three living neighbors. A living cell becomes nonliving if it has fewer than two or more than three living neighbors.

Cellular automata of this type are purely deterministic. Their state at time $t + 1$ is completely determined by their state at time t. Bhargava, Kumar, and Mukherjee (in preparation) have introduced the idea of stochastic cellular automata. In such an automaton, the fate of each cell is determined randomly, although the randomness may be related to the states of other cells in the automaton.

They have applied the concept of stochastic cellular automata to diffusion of innovations. The following material is adapted from their work.

We will develop a cellular automaton to simulate the diffusion of an innovation through a population of potential adopters. We assume that once a cell adopts, it remains an adopter (i.e., in terms of the Game of Life, it never dies). We assume further that the probability that a cell will adopt the innovation is a function of the number of immediately neighboring cells that have already adopted it (i.e., adoption is by imitation only). We assume that there is some maximum number of potential adopters in the population, which is less than the total population. Finally, we assume that "newness" affects adoption. As time passes, potential adopters who have not yet adopted become less and less likely to adopt, because the element of "newness" wears off.

We will model this situation using a 100 by 100 array of potential adopters as our population. We will set a maximum of no more than 200 possible adoptions. Since in our model adoption occurs by imitation only, we must "seed" the population with some adopters at time $t = 0$. We will restrict the seeding to a 15 by 15 section at the center of the larger array. This is done to reduce the possibility of "edge effects" at the boundaries of the larger array. To model the loss of newness, we define a variable X that undergoes the following change from one time step to the next:

$$X_{t+1} = X_t + (1 - X_t) \times \frac{\text{number of adopters}}{\text{maximum number of adopters}} \tag{9.9}$$

We will specify X_0 as 0.9. From $t = 1$ on, X will increase, approaching 1.0 as an upper limit.

We define the probability that a cell will adopt the innovation as:

$$P_t = (1 - X_t) \times (\text{number of adopting neighbors})$$

Thus, the more neighbors a cell has that have already adopted, the more likely it is to adopt. As X increases with the passage of time, however, the probability of adoption decreases (newness wears off).

Figure 9.9 shows the results of 50 runs of this model, each taking 30 time steps, with one adopter "seeded" at $t = 0$. The mean number of adoptions levels out between 50 and 60 adopters at about $t = 15$. The one-standard-deviation confidence bounds range from just above 50 adopters to almost 70 adopters.

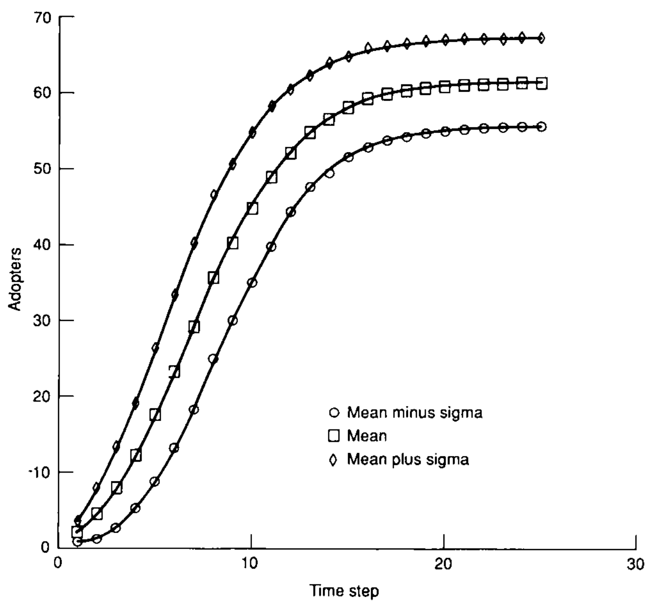

Figure 9.9 Number of adopters, with one initial seed.

Figure 9.10 shows the effect of increasing the number of seeded adopters at $t = 0$. The eventual number of adopters increases, as does the rate of adoption (steepness of the plot of cumulative adoptions).

One of the interesting conclusions reached by Bhargava et al., a conclusion that shows up fairly clearly in Fig. 9.10, is that the "takeover time" for the level of adoption to go from 10 to 90 percent of final number of adopters is relatively insensitive to the number of initial seeds and becomes less sensitive as the number of seeds increases, i.e., the steepness of the curve does not increase much going from 9 seeds to 17 seeds. As the number of initial seeds increases, however, the final number of adopters also increases. What seems to be happening is that a larger number of initial seeds induces cells to adopt before "newness" wears off. If the number of initial adopters is too low, newness wears off before the potential adopters have neighbors to imitate.

Some modifications to this model are readily apparent. For instance, the newness effect could be removed, allowing probability of adoption for a given number of neighbors to remain constant over time. This simply means allowing X to have the same value throughout the simulation. Another possible modification is to include in the probability of adoption equation a term for adopting that does not depend upon the number of neighbors. Thus, a "pioneer" might adopt even in the absence of neighbors to imitate. Other modifications might also be made to adapt the model to specific situations.

The stochastic cellular automaton is an example of a probabilistic model that can be used to forecast the rate of adoption of a technology. As with all probabilistic models, it does not give a single point forecast but instead gives a range of results and a probability distribution over that range.

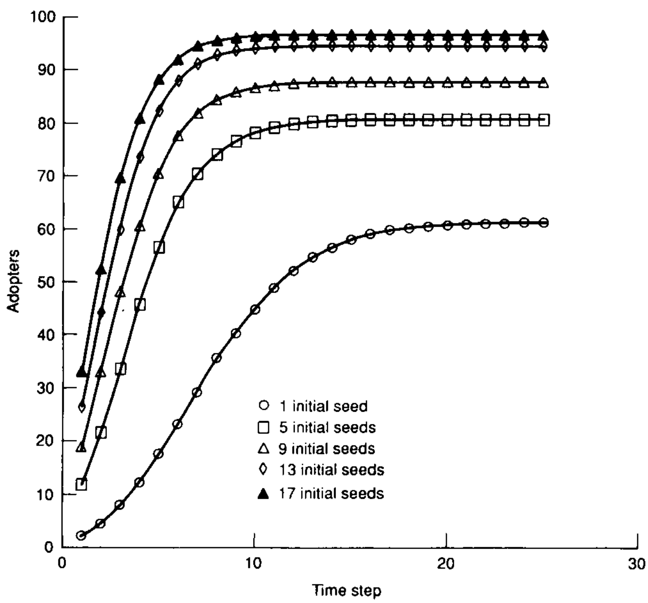

Figure 9.10 Adoption with varying initial seeds.

9.7 Summary

Probabilistic methods do not provide single point estimates of future performance or market share. Instead, they produce probability distributions that describe the frequency with which different values may be achieved. These probabilistic methods can be used in two ways. The first is to generate the probability distribution, as we did above for the case of the lags in application. The second is to use the probability distributions as inputs to a Monte Carlo simulation. The output of repeated simulations is then itself a probability distribution for the value of interest to the forecaster.

References

Bhargava, S. C., A. Kumar, and A. Mukherjee, "Stochastic Cellular Automata Model of Innovation Diffusion," *Technological Forecasting & Social Change* (in preparation).

Heppenheimer, T. A., "The Closing of the Aviation Frontier," *American Heritage of Invention & Technology,* 25–31 (Winter 1991).

Johnson, N. L. and S. Kotz, *Continuous Univariate Distributions – 1,* John Wiley, New York, NY, 1970.

Knuth, D. E., *Seminumerical Algorithms,* Addison-Wesley, Reading, MA, 1981.

Kwasnicka, H., R. Galar, and W. Kwasnicki, "Technological Substitution Forecasting With a Model Based on Biological Analogy," *Technological Forecasting & Social Change,* 23:41–58 (1983).

Martino, J. P., "Using Precursors as Leading Indicators of Technological Change", *Technological Forecasting & Social Change,* 32:341–360 (1987).

Martino, J. P., "Probabilistic Technological Forecasts," *Technological Forecasting & Social Change* 42:(4) (1993).

Meade, N., "Technological Substitution: A Framework of Stochastic Models," *Technological Forecasting & Social Change,* 36:389–400 (1989).

Murthy, D. N. P., "A Stochastic Model for Technology Forecasting," *Technological Forecasting & Social Change,* 14:27–37 (1979*a*).

Murthy, D. N. P., "Some Extensions to a Stochastic Model for Technological Forecasting," *Technological Forecasting & Social Change,* 15:273–280 (1979*b*).

Murthy, D. N. P. and K. C. Staib, "Forecasting Maximum Speed of Airplanes," *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-14(2) (March/April 1984).

Nash, J. C., *Compact Numerical Methods for Computers,* 2nd ed., New York, Adam Hilger, 1990.

Sahal, D., "Invention, Innovation, and Economic Evolution," *Technological Forecasting & Social Change,* 23:213–235 (1983).

Tribus, M., *Rational Descriptions, Decisions, and Designs,* Pergamon Press, New York, NY, 1969.

Appendix 9A. Program Listing for Competitive Substitution

```
10 NUM_TIMES% = 25
20 NUM_RUNS% = 30
60 Dim HIST1(NUM_TIMES%,2),HIST2(NUM_TIMES%,2),
       HIST3(NUM_TIMES%,2)
62 Dim AVG_FITNESS(NUM_TIMES%)

100 For J% = 1 to NUM_RUNS%
110 HIST1(0,1) = .85:
    HIST2(0,1) = .05:
    HIST3(0,1) = .10
```

```
120 FIT1 = .2:
      FIT2 = .6:
      FIT3 = .5
130 OLD _ SALES1 = 85:
      OLD _ SALES2 = 5:
      OLD _ SALES3 = 10
140 TOTAL = OLD _ SALES1 + OLD _ SALES2 + OLD _ SALES3
150 AVG _ FIT = (FIT1*OLD _ SALES1 + FIT2*OLD _ SALES2
  + FIT3*OLD _ SALES3)/TOTAL
160 AVG _ FITNESS(0) = AVG _ FIT
170 For I% = 1 to NUM _ TIMES%

200 Rem THIS SECTION GETS THE NEW VALUE FOR TECHNOLOGY 1
210 MEAN1 = HIST1(I%-1,1)*(FIT1-AVG _ FIT)
220 LOW1 = MEAN1-.1:
      UP1 = MEAN1 + .1
222 FRAC _ CHANGE1 = RAN _ NUM(LOW1,UP1)
230 NEW _ SALES1 = max(0,OLD _ SALES1*(1 + FRAC _ CHANGE1))
240 OLD _ SALES1 = NEW _ SALES1
242 If NEW _ SALES1 _ .00001 THEN NEW _ SALES1 = 0

300 Rem THIS SECTION GETS THE NEW VALUE FOR TECHNOLOGY 2
310 MEAN2 = HIST2(I%-1,1)*(FIT2-AVG _ FIT)
320 LOW2 = MEAN2-.1:
      UP2 = MEAN2 + .1
322 FRAC _ CHANGE2 = RAN _ NUM(LOW2,UP2)
330 NEW _ SALES2 = max(0,OLD _ SALES2*(1 + FRAC _ CHANGE2))
340 OLD _ SALES2 = NEW _ SALES2
342 If NEW _ SALES2 _ .00001 THEN NEW _ SALES2 = 0

400 Rem THIS SECTION GETS THE NEW VALUE FOR TECHNOLOGY 3
410 MEAN3 = HIST3(I%-1,1)*(FIT3-AVG _ FIT)
420 LOW3 = MEAN3-.1:
      UP3 = MEAN3 + .1
422 FRAC _ CHANGE3 = RAN _ NUM(LOW3,UP3)
430 NEW _ SALES3 = max(0,OLD _ SALES3*(1 + FRAC _ CHANGE3))
440 OLD _ SALES3 = NEW _ SALES3
442 If NEW _ SALES3 _ .00001 THEN NEW _ SALES3 = 0

500 Rem THIS SECTION COMPUTES MARKET SHARES
510 TOTAL = NEW _ SALES1 + NEW _ SALES2 + NEW _ SALES3
520 HIST1(I%,1) + = NEW _ SALES1/TOTAL
530 HIST2(I%,1) + = NEW _ SALES2/TOTAL
540 HIST3(I%,1) + = NEW _ SALES3/TOTAL
550 HIST1(I%,2) + = (NEW _ SALES1/TOTAL)^2
560 HIST2(I%,2) + = (NEW _ SALES2/TOTAL)^2
570 HIST3(I%,2) + = (NEW _ SALES3/TOTAL)^2
572 AVG _ FIT = (NEW _ SALES1*FIT1 + NEW _ SALES2*FIT2
+ NEW _ SALES3*FIT3)/TOTAL
574 AVG _ FITNESS(I%) = AVG _ FIT
580 Next I%
590 Next J%

600 Rem THIS SECTION COMPUTES MEANS AND STANDARD DEVIATIONS
610 For I% = 1 to NUM _ TIMES%
620 HIST1(I%,1) = HIST1(I%,1)/NUM _ RUNS%
```

```
630 HIST1(I%,2) = sqrt(HIST1(I%,2)/NUM_RUNS%-HIST1(I%,1)^2)
640 HIST2(I%,1) = HIST2(I%,1)/NUM_RUNS%
650 HIST2(I%,2) = sqrt(HIST2(I%,2)/NUM_RUNS%-HIST2(I%,1)^2)
660 HIST3(I%,1) = HIST3(I%,1)/NUM_RUNS%
670 HIST3(I%,2) = sqrt(HIST3(I%,2)/NUM_RUNS%-HIST3(I%,1)^2)
680 Next I%

690 Rem THIS SECTION PRINTS THE HISTORIES TO A FILE
700 FILE1$ = "TEK1HIST.DAT"
710 FILE2$ = "TEK2HIST.DAT"
720 FILE3$ = "TEK3HIST.DAT" 722 FILE4$ = "AVGFIT.DAT"
730 Open output #3,FILE1$
740 Open output #4,FILE2$
750 Open output #5,FILE3$
752 Open output #6,FILE4$
760 For I% = 0 to NUM_TIMES%
770 A = HIST1(I%,1):B = max(0,A-HIST1(I%,2)):C = min(1,A + HIST1(I%,2))
780 D = HIST2(I%,1):E = max(0,D-HIST2(I%,2)):F = min(1,D + HIST2(I%,2))
790 G = HIST3(I%,1):H = max(0,G-HIST3(I%,2)):I = min(1,G + HIST3(I%,2))
800 Print #3,I%,A,B,C
810 Print #4,I%,D,E,F
820 Print #5,I%,G,H,I
822 Print #6,I%,AVG_FITNESS(I%)
830 Next I%
990 End

10000 Def func RAN_NUM(Z1,Z2)
10010 Local Z1,Z2
10020 Return Z1 + rnd*(Z2-Z1)
10030 Func end
```

Appendix 9B. The Nelder-Mead Algorithm

The algorithm is presented here in terms of finding a set of discrete probabilities having maximum entropy while being constrained to a prescribed mean and variance. The presentation is based on that of Nash, except that inequalities have been reversed to find a maximum instead of a minimum.

The basic idea is that we wish to find n parameters that maximize a function subject to a set of constraints. In particular, we wish to find n probabilities associated with years 0 through $n - 1$. The Nelder-Mead algorithm does this by starting with $n + 1$ points in an n-dimensional space. These points must be noncoplanar, i.e., at least one point must be out of the plane formed by the rest (if $n > 3$, then one point must be out of the hyperplane formed by the remaining n points). The *simplex* formed by the $n + 1$ points is moved through the n-dimensional space and is contracted until the $n + 1$ points cluster around a "hilltop" that is a local maximum of a specified function defined in the n dimensions. A *simplex* is a set of $n + 1$ linearly independent points in a Euclidean space of dimension n or greater, and all the points formed by linear combinations of the $n + 1$ points which define the simplex, i.e., the vertices and all interior points. (Note that this algorithm suffers the same defect as all "hill-climbing" algorithms, i.e., it may halt at a local maximum that is not the global maximum.)

The algorithm seeks to maximize a function defined in the n-dimensional space. The object is to find the point in the space (i.e., the set of n probabilities) for which the entropy is a maximum while the mean and variance are equal to prescribed values. Thus, the function to be maximized is the entropy minus the penalty terms related to the deviation of the sum of probabilities from 1.0, the deviation from the desired mean, and the deviation from the desired variance. In addition, to assure that the solution stays in the positive unit hypercube, a penalty should be subtracted for any negative probabilities. One possible function that meets these requirements is:

$$\begin{aligned}\text{FUNCTION} = \text{ENTROPY} &- 100 \times \text{ABS(PSUMPROBS} - 1)\\ &- 10 \times \text{ABS(PMEAN/MEAN} - 1)\\ &- 10 \times \text{ABS(PVARIANCE/VARIANCE} - 1)\\ &- \text{BIGNUM} \times \text{NUM\_NEGATIVE\_PROBS}\end{aligned}$$

where PMEAN and PVARIANCE are the mean and variance at the point at which the function is being evaluated, MEAN and VARIANCE are the desired mean and variance, respectively, PSUMPROBS is the sum of the probabilities (e.g., vector components) at the point being evaluated, BIGNUM is a large number such as 10^{20}, and NUM_NEGATIVE_PROBS is the number of probabilities less than 0. The coefficients 100 and 10 were selected to weight the deviation of probabilities from 1.0 more heavily than deviations from the desired mean and variance. The choice of coefficients was subjective and arbitrary, but was found satisfactory in several trials. This function must be evaluated at several points in the algorithm. Hence, it should be defined as a subroutine that can be called from anywhere in the program.

Nash recommends the following values for the parameters used in the algorithm (their use will be seen in the following explanation):

$$\alpha = 1 \text{ (a reflection factor)}$$

$$\beta = 0.5 \text{ (a contraction factor)}$$

$$\Gamma = 2 \text{ (an extension factor)}$$

These values were found to give satisfactory results in several trials.

The steps in carrying out the Nelder-Mead algorithm are as follows.

1. Establish a simplex of $n + 1$ points in the n-dimensional space of probabilities. This is an array whose row index ranges from 0 to $n - 1$ and whose column index ranges from 0 to n. The array is filled with 0s except as follows: Cells a_{ii} are filled with 1 for i ranging from 0 to $n - 1$; and cells a_{in} are filled with 1 for all i. The result is a square submatrix of n rows with the first n columns having 1s on the diagonal and 0s elsewhere, and the final column is all 1s. Note that the final column is not a legitimate probability but is guaranteed to be out of the plane formed by the other columns of the array. The

penalty function for the sum of probabilities deviating from 1.0 will assure that this point is not selected as the optimal point.

2. Evaluate and rank the $n + 1$ points (columns) of the array. This is done by calling the evaluation function for each column. Three values are kept, LARGE, SMALL, and NEXTSMALL, as well as their indexes, LARGECOL, SMALLCOL, and NEXTSMALLCOL. Here LARGE is the evaluation function value for the point with the largest value, SMALL is the value for the point with the smallest value, and NEXTSMALL is the value for the point with the next-to-smallest value.

3. Compute the mean absolute difference between the largest and smallest points. This is done by subtracting the rows of LARGECOL from those of SMALLCOL, taking the sum of the absolute differences and dividing by n. If this distance is less than some tolerance (such as 0.001), go to step 11.

4. Compute the CENTROID of all the points except LARGECOL. CENTROID is a vector whose components are the mean values of the corresponding components of all the points in the simplex except the point with the largest value.

5. Compute the evaluation function at the CENTROID. If this gives a value larger than LARGE, replace LARGECOL with CENTROID and go back to step 3, otherwise go on to step 6.

6. Define a new point by reflecting SMALLCOL through CENTROID. This is done by computing

$$\text{REFLECT\_POINT} = (1 + \alpha) \times \text{CENTROID} - \alpha \times \text{SMALLCOL}$$

If the value of the evaluation function at REFLECT_POINT is greater than LARGE, go to step 7. If the value at REFLECT_POINT is between NEXT_SMALL and LARGE, go to step 8. If the value is less than NEXT_SMALL, go to step 9.

7. Extend in the same direction. This is done by computing:

$$\text{EXTEND\_POINT} = \Gamma \times \text{REFLECT\_POINT} + (1 - \Gamma) \times \text{CENTROID}$$

Replace LARGECOL with whichever is larger, REFLECT_POINT or EXTEND_POINT. Go back to step 3.

8. Replace SMALLCOL by REFLECT_POINT. Go back to step 2.

9. Replace SMALLCOL with a point between REFLECT_POINT and CENTROID. This is done by computing:

$$\text{NEW\_SMALLCOL} = \beta \times \text{SMALLCOL} + (1 - \beta) \times \text{CENTROID}$$

$$\text{SMALLCOLL} = \text{NEW\_SMALLCOL}$$

If the value of the evaluation function at the new SMALLCOL is greater than that at either the former SMALLCOL or at REFLECT_POINT, go back to step 2. If the value at the new SMALLCOL is less than that at the larger of the former SMALLCOL and REFLECT _ POINT, go to step 10.

10. Contract the simplex about the point having the largest value. This is done by replacing each point except LARGECOL with

$$\text{NEWPOINT} = \beta \times (\text{OLDPOINT}) + (1 - \beta) \times \text{LARGECOL}$$

(Including LARGECOL in the computation would simply replace it with itself; hence, time can be saved by omitting it from the computation.) Go back to step 2.

11. Print out the results—the probabilities at LARGECOL, the value of the evaluation function, and the mean and variance of LARGECOL.

Appendix 9C. Approximating Probability Distributions

In Monte Carlo simulations of technological progress or diffusion of innovations, it is necessary to draw random numbers from specified distributions. Many scientific calculators, as well as computer programming languages, have functions that will produce uniformly distributed pseudorandom numbers. In general, however, these sources do not directly provide random numbers drawn from other distributions. It is necessary to transform random numbers from a uniform distribution to the desired distribution.

Knuth presented algorithms for transforming uniformly distributed random numbers into random numbers having other distributions such as normal, exponential, beta, gamma, etc. These algorithms can be used only if the desired distribution of random numbers is one of these standard distributions.

The technological forecaster wishing to prepare a probabilistic forecast will most likely be dealing with an empirical distribution that does not exactly fit any of the standard distributions. For instance, suppose the forecaster wishes to perform a stochastic projection such as that described in Sec. 9.4 above. The distribution of time intervals between successive historical "record setting" events emphatically does not have a normal distribution, since it is skewed to the right (there are no negative intervals). Moreover, it may not be a good fit to any of the other standard distributions. Somehow, the forecaster must obtain a set of random numbers that has a distribution "close to" that of the empirical distribution. Several ways to do this will now be examined.

First, however, we need to examine some definitions. A probability distribution can be described either by a probability density function or by a cumulative distribution.

The probability density function $f(x)$ is defined as

$$f(x) = P(x < X \le x + dx) \tag{9C.1}$$

that is, $f(x)$ is the probability that a random variable X will fall in the infinitesimal interval between x and $x + dx$. As dx goes to zero, this becomes the probability per unit length, or probability *density,* in the vicinity of x.

The cumulative distribution function F(x) is defined as

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = P(X \le x) \tag{9C.2}$$

where y is a dummy variable of integration. F(x) is the probability that a random variable X will be equal to or less than x. F(x) is of course monotonically nondecreasing since f(x) can never be negative. Since f(x) can be 0, F(x) may have intervals in which it is constant.

Suppose a set of lags were described by a uniform distribution, with the minimum possible lag equal to 0 and the maximum possible lag equal to L. Then we would have

$$f(x) = \frac{1}{L} \tag{9C.3}$$

in the interval (0, L). We would also have

$$F(x) = \frac{x}{L} \tag{9C.4}$$

in the interval (0, L) and F(x) = 1 for $x > L$. More generally, however, F(x) is S-shaped rather than the straight line of Eq. (9C.4).

We will now look at a concrete example of approximating an empirical distribution. The first column of Table 9C.1 shows the lags from commercialization to first flight of aluminum alloys, arranged in ascending order. It is possible to generate an empirical cumulative distribution function from these lags. To begin with, we recognize that lag values may occur outside the range of those actually observed. Thus, the actual values are to be in the interior of the cumulative distribution. We, therefore, generate a cumulative probability distribution by dividing the number of data points equal to or smaller than a particular value by *one more* than the total number of data points. There are 13 data points in the list (counting duplicates). Thus, the cumulative proba-

TABLE 9C.1 Empirical Distribution of Lag of First Flight after Commercialization of Aluminum Alloys

Lag, years	Cumulative probability
1	0.0714
2	0.11428
3 (four times)	0.4286
4 (twice)	0.5
5 (twice)	0.5714
6	0.6429
8	0.8571
11	0.9286
0	0
17	1.0

bility associated with the first one is 1/14 or 0.0714. The remaining probability values are obtained in the same way, i.e., assigning a cumulative probability of $i/14$ to the i-th data point. Duplicates are included as many times as they appear.

Having put the actual values in the interior of the distribution, we still need to determine the end points of the distribution. Since the data are lags, there cannot be any negative values. Therefore the shortest possible lag is 0, and the probability is equal to 0 that any lag will be equal to or less than that. This is shown in the first row below the dashed line in the table. How do we determine the upper bound on possible values? This can be done using the formula:

$$x_u = \left(\frac{S}{\sqrt{(2 \ln n)}}\right)(4.6 - \sqrt{(2 \ln n)}) + M \tag{9C.5}$$

where x_u is the upper bound
n is the number of data points
M is the mean value of the data points
S is the standard deviation of the data points

This formula is derived from Extreme Value Theory (Johnson and Knotz, Chap. 21) and selects a value for the upper limit x_u such that there is at most a 1 percent chance that an actual observation could exceed x_u. Applying this formula to our data on aluminum lags, we obtain the value 17.143, which has been rounded to the value 17 and is shown in the last row of Table 9C.1. This cumulative distribution is plotted in Fig. 9C.1, where the actual values have been connected with straight lines.[1]

Suppose now that we draw a random number from a uniform distribution with the range (0,1), and that we obtain the value 0.54. The dashed line in Fig. 9C.1 shows how we obtain the X value corresponding to this probability. We trace a horizontal line from 0.54 on the vertical axis to the cumulative distribution, then drop vertically to the horizontal axis, where we find a lag of 4.6 years. Thus, with probability 0.54, the lag will not exceed 4.6 years. Using this same approach, we can find the X value corresponding to any probability value in the range (0, 1). In practice, of course, a computer program would be written to draw a random number from a uniform distribution, identify the interval into which it falls (i.e., between which pair of empirical data points it is found), then interpolate between those two points to get the corresponding lag value.

If there is a large number of data points, this method can become tedious. We might further approximate the empirical distribution by replacing sections of it with straight lines. In Fig. 9C.1, for instance, we might draw a

[1]Note that this distribution differs from the one in Fig. 9.6. This is because the empirical distribution developed here is not a *maximum entropy* distribution. Moreover, because of the way the lower and upper bounds on the distribution were determined, its mean and variance may differ from the mean and variance of the sample on which it is based.

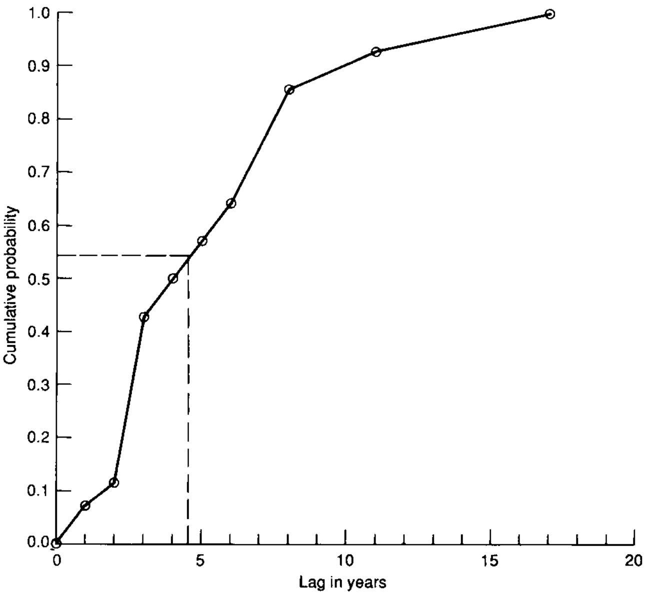

Figure 9C.1 Empirical cumulative probability distribution of aluminum alloy lags.

straight line from the point for a lag of 3 years to the point for a lag of 8 years. Likewise the points for 0 and 2 years and the points for 8 and 17 years might be connected by straight lines. By judicious choice of where to put the straight lines, we can get a good approximation to the empirical distribution while still keeping the number of line segments small.

In some cases, the empirical cumulative distribution has too many curved portions to allow straight-line approximations. In this case we must resort to another approach.

If there are many data points, we can work with the probability density function rather than the cumulative distribution function. We divide the range of the data into convenient intervals (they need not all have the same width) and plot a histogram for the data. In such a histogram, the height of a bar indicates the fraction of the data points falling into that interval. We convert this into a density function by dividing the height of the bar by the width of the interval. The result is a "histogram" in which bar height indicates the *density* of data points in an interval rather than the total number of data points in that interval.

The next step is to approximate this density function with a series of straight lines. These straight lines can then be integrated, to obtain a cumulative distribution function. Once this is obtained, random variables drawn from a uniform probability distribution can be converted to random variables drawn from the empirical distribution, just as in Fig. 9C.1.

The following example illustrates this approach. In Sec. 9.4, speed ratios were obtained in a three-step procedure. First, random numbers were drawn from a distribution for intervals between speed records. Second, the correlated portion of the speed ratio was obtained from a regression equation in which the interval was the independent variable. Third, the uncorrelated portion of the speed ratio was obtained by drawing a random number from the distribution of residuals from the regression.

It was found that the distribution of residuals was a narrow peak, symmetrical about 0. Therefore it was decided to approximate it as a triangle, i.e., approximate its skirts with straight lines. To illustrate how this was done, we compute the cumulative distribution corresponding to a triangular probability density.

Figure 9C.2 shows a triangular probability density centered on 0. The length of the base is b, extending from $-b/2$ to $b/2$. It is known that the area of a triangle is $\frac{1}{2} \times b \times h$, where h is the height. Since the triangle is a probability density, this area must be equal to 1.0. Hence, we can solve for h, which is found to be $2/b$.

From $-b/2$ to 0, $f(x)$ is given by:

$$f(x) = \frac{2hx}{b} + h = \frac{4x}{b^2} + \frac{2}{b} \tag{9C.6}$$

We integrate this to obtain $F(x)$:

$$F(x) = \int_{-b/2}^{x} \left[\frac{4z}{b^2} + \frac{2}{b}\right] dz = \frac{2x^2}{b^2} + \frac{2x}{b} + \frac{1}{2} \tag{9C.7}$$

(Here z is a dummy variable of integration. A check of this integral shows that it is actually equal to 0 at $x = -b/2$ and to 1/2 at $x = 0$, as it should be.) From 0 to $b/2$, $f(x)$ is given by:

$$f(x) = \frac{2}{b} - \frac{4x}{b^2} \tag{9C.8}$$

Integrating this we obtain:

$$F(x) = \int_{0}^{x} \left[\frac{2}{b} - \frac{4z}{b^2}\right] dz = \frac{2x}{b} - \frac{2x^2}{b^2} \tag{9C.9}$$

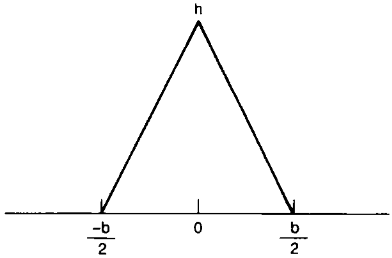

Figure 9C.2 Triangular probability density.

We can now solve these equations algebraically for x, given a value for $F(x)$. Let y be a probability drawn from a uniform distribution with the range (0, 1). If the random number is less than 0.5, we use the expression for $F(x)$ to the left of the origin and obtain:

$$x = \frac{b\sqrt{(2y)} - 1}{2} \tag{9C.10}$$

If y is greater than 0.5, we use the expression for $F(x)$ to the right of the origin and obtain:

$$x = \frac{b}{2} - \frac{b\sqrt{(2)}\sqrt{(y - 1)}}{2} \tag{9C.11}$$

Thus, given a probability, we can obtain the corresponding value of x. The x will then have been drawn from the empirical distribution of residuals.

While this example utilized a triangular probability density function, the method can be extended to any probability density function that can be approximated by a series of straight lines. The approach is to write the equation for each straight line segment, then integrate it from the left end to the right end, taking care that the constant of integration for each segment is equal to the value of the integral for the preceding segment, and that the value of the integral for the entire density function is 1.0. The resulting cumulative distribution function will consist of sections of parabolas. The cumulative distribution function gives a probability y for each value of x. The equation for each parabolic section must be solved to give the value of x corresponding to a probability y.

Whenever it is necessary to draw a random variable from an empirical probability distribution, one of the approximation methods in this appendix can be used to transform a number drawn from a uniform probability distribution to one drawn from the desired distribution. The approximations are somewhat crude, but considering the kind of data technological forecasters usually have to work with, the approximations given here will usually be found to be no worse then the original data.

Problems

1 Assume that some technology has annual increases in performance. Assume that the increase in performance in each year is proportional to the performance the preceding year but that the actual value of the increase is randomly distributed between 0 and 1. Perform a simulation of the growth of this technology as follows: Start it with an initial value of 1 in year 0. Obtain a single random digit, either from a table of random numbers such as App. E or use a random number function on a calculator. Multiply this by 0.1 and by the preceding value of the technology. Add the result to the preceding value to get the new value. Proceed forward to the end of year 10. What is the final value of the technology at the end of the tenth year? Plot your year-by-year results on semilog paper. Does the plot appear to be an exponential trend?

2 Carry out a manual simulation of the probabilistic biological analogy model of Sec. 9.3. At each time step, compute the average fitness of all three technologies, the deviation from average fitness of each technology, and the market share of each technology. Compute the change in sales for each technology as follows: For each technology, compute the mean fractional change in sales as the product of its market share and its deviation from mean fitness. For each technology, draw a random number (at least 3-digits long) from App. E (or use a calculator). Convert this to a number between 0 and 1 by placing a decimal point in front of it. Subtract 0.5 and divide the result by 10 (this converts the random number to one centered on zero and with limits of ± 0.1). For each technology, add its transformed random number to its computed mean fractional change in market share. Multiply the result by the sales of that technology for the most recent period. This is the change in sales. Add this algebraically (taking sign into account) to the sales for the preceding time period. Repeat this for 10 time steps. Plot your results and explain them.

3 Using the data on lags from invention to innovation from Table A.55 and the method of maximum entropy, compute the probability distribution for lags between invention and innovation.

Chapter 10

Environmental Monitoring

"Figuring the nature of the times deceas'd, the which observed, a man may prophesy, with a near aim, of the main chance of things, as yet come to life, which in their seeds and weak beginnings lie intreasured. Such things become the hatch and brood of time."

WILLIAM SHAKESPEARE
King Henry IV, Pt. 2
Act III, Scene 1

10.1 Introduction

Forecasts based on trends or growth curves seem to require continuity between the past and the future. Forecasts based on causal models likewise seem to require the consistent operation of the causal factors. These methods, therefore, seem inherently incapable of predicting "breakthroughs," whose very name seems to imply surprise and lack of advanced warning. Thus, is it true that breakthroughs are inherently unpredictable?

Webster defines a breakthrough as "a sensational advance in scientific knowledge in which some baffling major problem is solved." This definition does not inherently require surprise or unpredictability in the breakthrough. However, it is not precise, since it is not clear how great an advance must be to qualify as "sensational," nor how bad a problem must be to qualify as "baffling."

For the sake of precision, we define a *technological breakthrough* as an advance in the level of performance of some class of devices or techniques, perhaps based on previously unutilized principles, that significantly transcends the limits of prior devices or techniques. This definition has several implications. One is that simply moving up the growth curve of an established technical approach does not qualify as a breakthrough. Another is that the predictable limits of a specific technique must be surpassed by the breakthrough. This may come from finding that the previously accepted limit was in error, or it may come from the use of some new technique or discovery to perform a particular function. In the latter case, it is not necessary that the discovery be new in an absolute sense; it may have been widely used for some other purpose before and has been adapted to the new function. Finally, the definition implies that adopting a successor technique that has a level of inherent capability higher than that of the prior technique represents a breakthrough, i.e., a breakthrough means a move to a new growth curve with an upper limit higher than that of the old growth curve.

This last point is particularly important. A breakthrough may be identified as such simply because it is on a new growth curve, even though the current level of performance is inferior to existing devices on the old growth curve. This clearly implies that at least some breakthroughs are predictable, since the appearance of the first low-performance device on a new growth curve is a precursor event for the device that will later surpass the limits of the old growth curve. Such a situation was depicted in Fig. 5.1, which presented the growth curves for propeller-driven and jet-propelled aircraft. Even though the early jets were inferior to their propeller-driven contemporaries, the jet engine was clearly a breakthrough because it had the inherent capability to surpass the performance of the older technical approach.

10.2 Examples of Breakthroughs

Almost any list of major "breakthroughs" of the current century will include atomic energy, the transistor, and penicillin. According to technological folklore, each of these burst unheralded on a started world. However, even a cursory glance at the history of each of these breakthroughs shows many precursor events. To illustrate this point, a review of the history of each is worthwhile.

Atomic energy. Table 10.1 shows some major events in the history of atomic energy. The earliest event was a scientific breakthrough and was totally unpredictable; the equivalence of mass and energy was a violation of accepted scientific theory. Once this equivalence was recognized, however, there was at least the theoretical possibility of converting matter into energy.

The next major event was the discovery that not all chemically identical elements were physically identical. Before this it was believed that mass was the most important characteristic of an atom. Two atoms with the same chemical properties must necessary have the same mass; two atoms with the same mass must have the same chemical properties. However, Boltwood in 1906 and McCoy and Ross in 1907 showed that the radioactive elements ionium and radiothorium were chemically identical with the known element thorium, although they had different masses (and, therefore, had been assumed to be three distinct elements). Soddy coined the name "isotopes" for chemically identical elements with different masses. In 1913, J. J. Thompson found that neon also had two isotopes, demonstrating that nonradioactive elements as well as radioactive elements could have isotopes.

Prior to 1911, it had been assumed that the matter in an atom was uniformly distributed throughout the space it occupied (the "billiard ball" model). However, Rutherford showed that virtually all the mass of an atom was concentrated in a small nucleus; the rest of the volume of the atom was mostly

TABLE 10.1 Major Scientific Events in the History of Atomic Energy

Year	Event
1905	Mass-energy equivalence. Publication of a paper by Einstein establishing the equivalence of mass and energy.
1906	Isotopes of radioactive elements. Discovery of chemically identical elements with different radioactive properties.
1911	Atomic structure. Experiments by Rutherford showed that the mass of an atom is concentrated in a positively charged nucleus.
1913	Isotopes of nonradioactive elements. Discovery of isotopes through differences in physical properties.
1919	Ejection of protons from nitrogen. First artificially induced nuclear reaction.
1919	Mass spectroscopy. Accurate determination of the masses of isotopes.
1920s	Mass defect (packing fraction). Discovery that the mass of a nucleus is less than the sum of the masses of the constituent particles.
1932	Discovery of the neutron. New particle, same mass as the proton, but sharing no electric charge.
1938	Fission of uranium nucleus. Uranium atoms split into roughly equal halves.
1939	Chain reaction hypothesized. If neutrons are emitted during fission, further fissions can take place.
1942	Chain reaction produced. Acutal demonstration of fission by neutrons emitted from earlier fissions.
1945	Atomic bombs. First use in warfare.
1956	Commercial nuclear power generation. Actual power plant generating electricity from nuclear energy.

empty space. As a result of his experiments, it became clear that atoms had equal numbers of positively and negatively charged particles, and the positively charged particles made up most of the mass of the atom.

Rutherford continued his study of the structure of the atom. In 1919, he bombarded nitrogen atoms with alpha particles and found that a nuclear reaction took place. The nitrogen emitted a proton (nucleus of a hydrogen atom) and turned into an oxygen atom. This was actual transmutation of elements, the unfulfilled dream of the medieval alchemists. Moreover, the emitted proton had more energy than did the alpha particle, indicating that somehow energy was being obtained from the nucleus of the atom.

The source of this energy was determined with the development of mass spectrometers, which allowed researchers to determine the masses of atoms. It had been customary to quote atomic masses in terms of multiples of the mass of a hydrogen atom. However, most elements had computed masses that were nonintegral multiples of the mass of hydrogen. With the discovery of isotopes, it was assumed that these nonintegral masses were simply the result of mixtures of isotopes. However, better mass spectrometers showed that individual isotopes still had nonintegral masses. In all cases, the mass was less than it "should have been" if the atom were made up of protons and electrons. This "mass defect" was explained as mass that had been converted into "binding energy" to hold the atomic nucleus together. It was discovered that the average binding energy per particle in the nucleus increased with increasing atomic mass, up to a maximum for atomic masses around 90; above this the average binding energy per particle declined again. This clearly meant that energy could be obtained from a nucleus either by combining two light atoms to make a moderately heavy one or by splitting a heavy atom into two moderately heavy ones. However, there appeared to be no way to split atoms on a practical scale. Even atomic particle accelerators ("atom smashers") missed their target atoms too often. The efficiency of any process using a particle accelerator would be incredibly low. Thus, any consideration of obtaining energy from atoms usually focused on combining (fusing) light atoms.

In 1932, Chadwick discovered the neutron. It was quickly discovered that neutrons were much more efficient at causing atomic transmutations than were any other types of particles. All during the 1930s, many scientists throughout the world conducted experiments in which they bombarded various elements, including uranium, with neutrons.

In 1938, Hahn and Strassman conducted such an experiment with uranium and discovered that they had split the uranium atom into smaller fragments that gave off a great deal of energy. When they published their results, many other researchers realized they had just missed discovering uranium fission themselves because they failed to interpret the results of their experiments properly.

At this point, it became clear to many people that it was possible to obtain energy from the fission of uranium, even if many details still needed to be resolved. At the very least, one could imagine a power plant containing some uranium and a neutron source. The neutrons would fission the uranium, which would release energy. This energy could be extracted and used to perform work. Such a power plant was described by science fiction author Robert A. Heinlein in "Blowups Happen," a story published in 1940. The technical feasibility of such a plant was certain; only the economic feasibility remained in doubt.

In 1939, Enrico Fermi hypothesized that uranium atoms might give off neutrons when they were split by neutrons. If so, a "chain reaction" was possible, which would eliminate the need for a neutron source. This chain reaction was demonstrated in 1942. The remaining problems took another 14 years to solve, with the first large-scale commercial production of electrical power from atomic energy being achieved in 1956 at the Calder Hall Station in England. This was 18 years after the discovery of fission and 51 years after mass-energy equivalence was discovered.

The point of this recitation is that there were a great many precursor events between the unpredictable scientific breakthrough of 1905 and the eventual commercial use of atomic power in 1956. Not all these precursor events provided positive signals. Some, such as the impracticality of atomic energy plants using particle accelerators, were false negative signals. Some, such as the possibility of fusing light atoms into moderately heavy ones, pointed in the wrong direction. Nevertheless, atomic energy was not an unheralded event. Many writers, of both fiction and popularized science, discussed atomic energy from 1914 to 1944 on the basis of their knowledge of contemporary scientific findings.

The transistor. In 1945 the *Bell Telephone Laboratories* (BTL) established a group to conduct research on a solid-state amplifier. The intent was to draw on the research done on semiconductors during World War II and on Bell's prewar research in quantum mechanics. The group made a false start on a solid-state amplifier patterned after a vacuum tube; it then switched to what would now be called a *field effect transistor* (FET). Their attempts to build a FET were guided by a theory that A. H. Wilson had published in 1931 describing the theory of semiconductors. They built several devices on the basis of this theory. None of them worked. While attempting to determine the reasons for failure, the group discovered the point contact transistor. This was in 1947. With the additional insight into semiconductor behavior this provided, they developed the theory of the junction transistor and fabricated one in 1951. Finally, in 1952 they achieved the FET they had initially sought (and incidentally demonstrated that Wilson's theory was not valid).

It is important to note that the BTL group working on the solid-state amplifier were confident that they could build one, on the basis of existing theory. The major contribution they expected to make to the development was the improved semiconductor materials that had been developed during the war. The existence of the theory itself and the improvements in materials were both precursor events that gave advanced warning of the possibility of a solid-state amplifier.

However, there were other precursor events. In 1925, J. E. Lilienfeld applied for a Canadian patent on a solid-state amplifier. In 1926 he applied for a U.S. patent on the same device and was granted U.S. patent number 1,745,175 in 1930. In 1935, Oskar Heil of Berlin was granted British patent number 439,457 on a solid-state amplifier. In retrospect, the devices of both Lilienfeld and Heil are recognizable as FETs. Neither inventor could have built a working device, however, because the materials available to them were not adequate.

The transistor was clearly a breakthrough. However, it did not come unheralded. On the contrary, it was the result of a deliberate search by people who took seriously the advanced warning given by precursor events.

Penicillin. Penicillin was isolated by Chain and Florey in 1939. They traced their work to the observation by Fleming in 1928 that penicillium mold had an antibacterial effect. This observation was in itself an important precursor event.

The isolation of penicillin, however, was not the final step. Methods had to be developed to produce penicillin on a commercial scale before it could be used widely. Thus, the work of Chain and Florey was likewise a precursor event for the later commercialization of penicillin.

However, there was an earlier precursor event, which was completely forgotten. E. A. C. Duchesne entered the French Army Medical Academy in

1894, at the age of 20. He became interested in the possible antagonism between molds and bacteria and did his thesis research on this topic. He experimented with a penicillium mold, observing that under certain circumstances the growing mold would kill bacteria. Next he grew the mold on a nutrient solution. He infected guinea pigs with virulent bacteria and injected half the experimental animals with the nutrient broth on which he had grown the mold. Those receiving the "penicillin" injections survived; the remainder did not. He reported his results in his thesis prior to his graduation in 1897. After graduation he began his intended career as a doctor in the French army, serving at various posts where it was impossible to continue his research. In 1902 he contracted tuberculosis, and he died from the disease in 1912. His thesis was not published and was lost until a librarian accidentally discovered it around 1944. Once it was rediscovered, Fleming and others involved in the work on penicillin recognized its importance, particularly Duchesne's methodical research that led him directly to the application of his observations about the antagonism between molds and bacteria.

This example, as well as that of atomic energy, illustrates that often a breakthrough may really be a "rediscovery" of something that was discovered once before and lost, or discovered and laid aside because it wasn't fully appreciated.

10.3 The Process of Monitoring

The examination of breakthroughs in the preceding section shows that there were many precursor events that would have made it possible to forecast the eventual development of these technologies. The same is true of all other technological breakthroughs (though not necessarily true of scientific breakthroughs). Forecasting a technological breakthrough, then, requires that precursor events be identified and used to provide advanced warning. This process of identifying precursors of possible breakthroughs is referred to as *monitoring.*[1]

Monitoring for breakthroughs is a systematic means for identifying precursors and using them to provide advanced warning. The monitoring process is designed to help the forecaster answer two questions: Which events are precursors, and what do the precursors signify?

Monitoring involves four steps: collecting, screening, evaluating, and threshold setting. Each step has a specific purpose. Taken together, the steps provide a system that will help the forecaster extract the meaningful signals from a welter of noise and irrelevant events.

Each of these steps is described below. To help make the descriptions more concrete, an example, namely the development of the jet engine in the 1930s, is traced using the steps.

[1]The concept of monitoring for precursors was pioneered by James Bright (1970) and (independently but later) by Richard Davis (1973). Others may also have invented it independently. In principle, it is highly similar to the open-literature intelligence-gathering engaged in by military intelligence agencies.

Collection. Collection brings information into the system. The collection process must be designed to cover every likely source of information. Thus, the first step in the collection process is to decide which sources to monitor and to what depth. The approach should be flexible enough to permit the addition or deletion of sources as experience dictates, without disrupting the process. For instance, a forecaster planning to track legislative activity might limit initial coverage to include only the U.S. Congress. Later experience might indicate that one or more state legislatures should be included, either because the initial coverage was inadequate or because conditions have changed (e.g., the relevant activity has moved from Washington, D.C., to the states). The design of the collection activity should be flexible enough that changes such as this can be made easily.

Example To illustrate the idea of collection, consider an aircraft engine company in the year 1930. This company is concerned about possible threats to its business of reciprocating aircraft engines. The firm's technological forecaster decides to track patents granted in the field of aircraft propulsion. As part of the forecaster's routine monitoring of current patent grants, he or she finds in 1930 that a patent is granted to Flying Officer Frank Whittle, of the Royal Air Force (RAF), for an aircraft engine based on the jet principle. The patent describes an engine in which air is drawn in through a turbine compressor, fuel is burned in the compressed air, and the combustion gases are used to drive a turbine, which in turn drives the compressor. There is sufficient energy left in the combustion gases, after driving the turbine, to provide some net thrust for aircraft propulsion.

For a forecaster concerned with aircraft propulsion, such as one working for an aircraft engine firm, this is *prima facie* an important signal. It must, however, be screened for significance.

Screening. Once the forecaster has selected the sources to be monitored and begun to collect information about events, the next step is to screen the items found. It is a serious mistake to let everything into the monitoring system in the hope that it can later be located through some sophisticated retrieval procedure. Too much effort will be spent cataloging and coding items that will never be used and that must eventually be purged from the system. The only items that should be allowed to enter the system are those that are relevant to the concerns of the forecaster's organization; all others should be screened out. The forecaster will normally need the help of other members of the organization to carry out the screening. In an industrial organization, help might be solicited from people in *research and development* (R&D), finance, marketing, or production to determine whether an item should be allowed in or screened out.

Example The jet engine patent we are using as an example would clearly be significant to an aircraft engine company. If the jet engine were to become practical, it would pose a threat to the company's existing product line. It is a potentially competing product. Hence, the company would be interested in monitoring the threat from the jet engine.

A forecaster beginning such a monitoring program would first try to determine whether there were any past signals that had a bearing on the one just discovered. Such a forecaster would have found that the jet engine idea was not a new one. In 1910, Henri Coanda proposed a jet propulsion system in which the compressor would be driven by a reciprocating engine instead of by an exhaust turbine. In 1913, Rene Lorin, a French engineer, proposed a jet engine in which the compression is derived entirely from the aircraft's forward velocity, eliminating the need for a compressor. In 1921, Maxime Guillaume received a French patent on a jet engine with a turbine-driven compressor similar to Whittle's design. In 1929, Dr. A. A. Griffith of the *Royal Aircraft Establishment* (RAE) proposed that a turbine engine be used to drive a propeller for providing aircraft propulsion.

Thus, the forecaster, beginning with Whittle's patent, finds that the idea is not new; that it has considerable history behind it. Moreover, the forecaster has identified not just one potential threat but three: the ramjet, the turboprop, and the turbojet.

Evaluating. Once an item is allowed into the system, it must be evaluated for its significance. The evaluation seeks to answer questions such as, What does this mean to my organization? If it represents the start of a trend or pattern, would it affect our mission? Would it make a product obsolete? Would it alter a production process? Would it have an impact on a customer? A supplier? The evaluation must examine the item not only by itself but in the context of the other items already allowed into the system and evaluated. Does it tend to confirm a pattern suggested by earlier items? Does it indicate a shift away from a previous pattern? Does it suggest a new pattern? In evaluation, just as in screening, the forecaster should seek advice from others in the organization who are better informed about the possible significance of the item.

Evaluation does not stop with the review of each item as it enters the system. As additional confirming signals appear, evaluation includes sketching out the patterns that appear to be building up. Thus, the information in the system includes not only a collection of events, but a set of hypotheses about what they mean. These hypotheses, in turn, lead to a set of statements about what confirming or disconfirming evidence might be expected. That is, if a hypothesis is true, what else would the forecaster expect to find? If a hypothesis is false, what else would the forecaster expect to find?

Example To return to our jet engine example, Whittle had calculated that for his engine to work, a compression ratio of 4 to 1 and a compressor efficiency of 75 percent would be required. In addition, the turbine blades would have to withstand a temperature of 1500°F. Thus, a forecaster monitoring this potential breakthrough would be interested in seeking information on compressor efficiencies and compression ratios, and the temperature capabilities of alloys for turbine blades.

In addition, the forecaster would find that in 1922 the U.S. Army Air Service had become interested in jet engines and asked Dr. Edgar A. Buckingham, of the then National Bureau of Standards, to investigate their feasibility. Buckingham's report, published in 1923 by the *National Advisory Committee for Aeronautics* (NACA), had showed that at a speed of 250 mi/h the fuel consumption of a jet engine would be four times that of a piston engine. Only at speeds above 500 mi/h would the jet engine be competitive in fuel economy. Hence, the forecaster would also be interested in tracking aircraft speeds.

Once other items are identified as worth monitoring, the forecaster must return to the collection step and begin acquiring information about these items. This usually requires a deliberate search for specific kinds of information. As these items are found and screened, evaluation will include fitting them into the patterns already found in the previously acquired information.

Finally, the evaluation step requires that the forecaster purge the system of events and hypotheses that have been discredited by subsequent events. This purging is required to assure that the system contains only active and currently valuable material. Material that is of only historical interest can be retired to the archives. This historical material should not be thrown away. It should be used, from time to time, to review how well the monitoring activity has performed in the past and to identify means for improvement.

Threshold setting. As evaluation continues, the evidence for one or more hypotheses will become stronger and stronger. A hypothesis cannot be confirmed, of course, until the breakthrough actually occurs, but by then it is too late to issue a forecast that gives advanced warning. Hence, the forecast must be made before the hypothesis is confirmed but when it is "strong enough." This requires that the forecaster set a threshold for each hypothesis. When the confirming signals show that the hypothesis has exceeded its threshold, it is time to make the forecast. However, the hypothesis may still be disconfirmed after the forecast is given. Therefore, the threshold should be set to balance two sets of risks: the risk of acting too soon and the risk of acting too late. Here again, it is not the forecaster's role to select the proper balance. Others in the organization should be consulted to determine the balance of risks. These other individuals must include the managers responsible for the success of the organization. It may well be that several thresholds can be set. When a hypothesis passes one of the lower thresholds, it might trigger some low-cost hedging action that does not commit the organization too deeply; passing a higher threshold might trigger more significant actions.

Example To return to our jet engine example, when should our forecaster give warning that the jet engine is almost here? For instance, we had suggested the forecaster monitor turbine compression ratio and efficiency. In 1931, these were 2 to 1 and 62 percent, respectively. By 1935, these had reached 2.5 to 1 and 65 percent. They were not yet at the required values of 4 to 1 and 75 percent, but they were getting there.

As it turned out, Hans von Ohain obtained a German patent on a turbojet engine similar to Whittle's in 1935. In that same year, he received support from the Heinkel company for development of a jet engine. In 1936, Whittle founded Power Jets Ltd. to develop an engine according to his design. In 1938, the U.S. Army Air Corps laboratories at Wright Field (Dayton, OH) began a 5-year program of development of gas turbines for jet engines. NACA began a program of compressor development. The RAE began work on a turbocompressor based on Griffith's 1929 design.

Hans von Ohain's jet engine achieved flight in 1939. Whittle's flew in 1941. In 1940, the Caproni-Campini CC-2 flew with a Coanda-type jet engine. In 1942, a U.S. aircraft flew using a jet engine developed by General Electric. That same year saw the first flight of a German jet aircraft that was no longer experimental but a combat aircraft, which would appear in the air over Germany in significant numbers by 1944.

In retrospect, it is not possible to say with certainty what thresholds should have been set in the 1930s. Nevertheless, whatever the correct thresholds were, by 1938 it was clear to almost everyone that they had been passed. The jet engine was on its way. This historical example cannot tell us what the thresholds should be in other cases. They must be set on the basis of specific circumstances. Nevertheless, the jet engine example indicates that once the threshold has been passed, development may follow quickly. Hence, setting thresholds is an important aspect of monitoring for breakthroughs.

These four steps of the monitoring process cannot be conducted in a simple straight-through fashion. In practice, a great deal of looping back is required. For instance, evaluation may show that screening has been done improperly, allowing too much or too little material to enter the system. Thus, a review of the screening criteria might be in order. The jet engine example shows that as the hypothesis is tracked, additional parameters might have to be monitored, requiring a revision of the collection process. From time to time, it may be necessary to reset thresholds as additional evidence is evaluated. The monitoring system cannot be allowed to "run itself." It is supposed to be an aid to the forecaster, not a substitute for thought. The forecaster should be actively involved in operating the system and must revise it whenever circumstances indicate that changes are needed.

10.4 Where to Look for Clues

The examples given in Sec. 10.2 above show that breakthrough signals often occur in areas remote from the final application. The first signal indicating atomic energy, for instance, came from an esoteric branch of physics. Thus, a systematic monitoring program must cover a wide range of sources. How should the sources be selected? How can the forecaster minimize the chance of overlooking something?

The list of sectors of the environment, introduced in Chap. 3, provides one means of assuring that important sources are not overlooked. The forecaster should cover each sector of the environment to the depth that seems appropriate for the particular pattern being tracked. In certain cases particular sectors can be omitted. However, the decision to not cover a sector should be made deliberately, not simply by default. Some questions pertinent to each sector of the environment are given below.

The technological sector. What performance level has the technology already achieved? Will performance improve? At what stage of innovation is the technology? What else is needed to make it feasible? What would have to be changed to make it a threat or an opportunity? What changes are needed to make it compatible with complementary technologies? What changes in supporting technologies would make the technology easier to produce, operate, or maintain?

The economic sector. What is the market size? What are production and distribution costs? Will these change? Are licensing agreements possible? Are innovative financing methods possible?

The managerial sector. Have new managers been assigned to a project dealing with the technology? Have new management techniques been devised that extend the scope of the "manageable"? Can new management procedures reduce technical risk? Shorten development time? If new management techniques were in use, what external signs would be present? Have these signs been observed?

The political sector. Have new laws been passed or regulations issued? Is there agitation for changes in laws or regulations? Has an incumbent been defeated by someone with a significantly different attitude on matters affecting the technology? Has there been a court decision that will have a bearing on the technology? Has the funding of an agency dealing with the technology been increased or decreased? Has an agency been given expanded jurisdiction? Has an agency been established with jurisdiction that covers the technology? Are there signals from the economic, social, cultural, or intellectual sectors that might presage political change?

The social sector. What changes are taking place in the structure of society, such as shifts in age distribution, geographic distribution, level of education, marriage age, family size, and income level? Are there any shifts in custom or tradition? Do younger groups have significantly different customs or traditions than their elders? Would any of these changes affect the technology?

The cultural sector. Have any groups acquired new values or abandoned previous ones? Has the number of people who subscribe to a value increased or decreased? Is a particular value being emphasized or deemphasized? Have the relative rankings of some values changed? Is the acceptable or tolerable level of realization of some value changing? Is the range of cases or situations held to be within the purview of some value expanding or contracting? Is the desired level of implementation of some value being retargeted upward or downward? Is the cost or benefit of maintaining some value shifting under the impact of external influences? Would any of these changes affect the technology?

The intellectual sector. Are the values of the intellectual leadership of society shifting in any of the ways indicated in the preceding paragraph? Are these value shifts being reflected in television programs, movies, novels, plays, newspaper columns, or magazines of opinion? Would this affect the attitudes of intellectuals and opinion leaders toward the technology?

The religious-ethical sector. Have there been expansions or contractions of the application of some doctrine by one or more major religious bodies? Are there precursors of this expansion or contraction in discussions by theologians or prominent lay people? Have smaller denominations made expansions or contractions that might be precursors of similar moves by major denominations? Have there been changes in the professional standards of scientists and engineers? Have there been precursors of such changes in discussions by leaders in the profession? Would any of these changes affect the technology?

The ecological sector. Have significant but unintended effects of some technology been observed? Has new knowledge shown that the previously accepted effects of some technology may no longer be acceptable? Would this make the technology being tracked more or less acceptable from an ecological standpoint?

In each sector the forecaster will have to identify the questions to be answered. Once this has been done, the forecaster must select sources that are likely to provide signals should these signals occur in the sectors being monitored. The forecaster should take advantage of the help of others in the organization. People who regularly read some journal or monitor some source for their own reasons are often willing to provide the forecaster with information they gather. By drawing on the efforts that people in the organization are already undertaking for their own purposes, the additional workload imposed by systematic monitoring can be kept quite low.

10.5 Examples from the Automotive Industry

Appendix B contains tables listing chronologies of events in several automotive technologies. An examination of these chronologies will help to make more concrete the points made in preceding sections as well as illustrate some of the issues the forecaster must be alert for that arise in monitoring technological change.

Table B.1 lists events in the history of the application of fuel injection to passenger automobiles. Fuel injection had been used in diesel engines since their invention and had long been an established technology there. However, it was rarely used for gasoline engines on passenger vehicles, and the vast majority of such vehicles utilized carburetors that mixed fuel and air outside the cylinder. Nevertheless, there had been some interest in applying fuel injection to passenger vehicle gasoline engines. In the late 1960s, interest was heightened by the idea that fuel injection might reduce the emissions associated with gasoline engines.

The first entry in Table B.1 reports the demonstration, on a laboratory basis, that fuel injection could be controlled by fluidic means, i.e., by "computers" which utilized the flow of air through alternative channels as switching elements. However, the technology was not yet mature enough for deployment. The miniaturization of electronics, an important development of the

1960s, was soon harnessed to the task of controlling fuel injectors and displaced completely the idea of fluidic control. The feasibility of electronic control was soon demonstrated, and the identification of engine "response surfaces" and the development of "control laws" were necessary steps in adapting fuel injection to automobiles (i.e., given a set of measurements taken on an engine, in what direction should the engine controls be moved to improve engine performance?). A further development was the oxygen sensor in the exhaust, which provided a direct measure of the engine's departure from the optimal fuel to air mixture. With this sensor available, closed loop control became possible. In the next several years, electronically controlled fuel injection appeared on "luxury" and "prestige" automobiles. Cost-reducing measures were introduced in both the manufacturing and testing of fuel injection units (borrowing from aviation technology in the testing case), and finally by the 1987 model year, fuel injection became a standard item.

Table B.2 presents the history of a broader innovation, electronic engine control, which actually incorporates automotive fuel injection as a component. The technology involves measuring various engine parameters, and adjusting engine controls to achieve some overall optimum performance: minimum emissions, maximum fuel economy, maximum power output, or some weighted combination of these. However, this broader innovation does not depend upon fuel injection as such. Initially this broader innovation incorporated the conventional carburetor. The history of the innovation shows a sequence of indicators such as improved sensors, the introduction of limited-performance versions on prestige cars, reduction in manufacturing costs, and finally use in the entire line of cars.

Tables B.3 and B.4 present the histories of two other innovations that were also incorporated into electronic engine control: electronic control of ignition timing and the electronic (breakerless) distributor. Both of these antedate any effort to add "closed loop" control to optimize engine behavior. They began as attempts to replace mechanical devices with electronic devices to obtain the usual advantages: greater reliability, greater precision, reduced maintenance, and lower cost. Indeed, without the prior development of these electronic subsystems, it would have been more difficult to overlay electronic control on the entire engine. Again, these histories show progression from devices with limited performance or applicability to devices capable of being included in mass-produced automobiles.

In each of these instances, the succession of events from preliminary or theoretical concepts through hardware demonstration, to application, provide a time series of leading indicators that fuel injection, the electronic ignition timing, or the distributorless ignition is coming. Each successive favorable indicator enhances the strength of the hypothesis that the technology in question will be adopted.

Table B.5 shows another feature commonly found when a forecaster is using leading indicators for technological forecasting purposes: improvements in a technology that is under threat of replacement. As early as 1976, electronic control was added to standard carburetors to improve engine performance.

This reduced the competitive advantage that fuel injection possessed. The occurrence of this leading indicator serves to *reduce* the strength of the hypothesis that fuel injection is the coming thing. However, despite the addition of electronics, the conventional carburetor was unable to compete with fuel injection, and the negative leading indicators associated with carburetors simply indicated a delay in the eventual takeover by fuel injection.

Table B.6 shows events in the history of another category of automotive innovation: replacement of steel automobile body shells with plastic body shells. The advantages of plastic over steel are lighter weight (important in achieving fuel economy) and rust resistance. The history includes improvements in plastic materials, the use of plastics for other parts of automobiles (which may be considered a favorable leading indicator), and ultimately the production of a "prestige" car with a plastic body shell. The history of this innovation is not yet complete; it is uncertain whether plastic will eventually displace steel.

Tables B.7 and B.8 present the history of steel's "counterattack." To overcome one advantage of plastic, corrosion-resistant steel was needed. An abortive attempt was made to develop it in 1980 using vitreous enamel, but the later development of steel sheet that was galvanized (zinc coated) on only one side solved the problem and overcame plastic's competitive advantage with regard to corrosion resistance. One approach to overcoming the weight problem was the use of finite-element analysis in the design of steel panels. This design approach, long utilized in the aircraft industry, allowed the design of steel panels that utilized material more efficiently and provided adequate strength at minimum weight. Another counterattack was the use of *high-strength low-alloy* (HSLA) steels. These steels, produced by special formulations, heat treatment, and rolling, allowed the use of thinner panels with a strength equivalent to thicker panels of the older steels. This reduced total weight, thereby offsetting one of the competitive advantages of plastic over steel. The occurrence of these leading indicators weakens the hypothesis that plastic body shells will replace steel.

Table B.9 presents the history of plastic as a structural material in automobiles. Initially, plastic structural parts were not equivalent in strength to the steel parts in operational use. However, with improvements in plastic materials and improved design procedures, it soon became possible to replace steel for a variety of structural parts including floors, bumpers, and fenders. Racing cars took advantage of plastic's lighter weight to achieve improved performance, since they could sacrifice appearance and durability. Finally, Ford and General Motors established their own plants for the manufacture of plastic structural components. While these events directly support the hypothesis that steel structural parts will be replaced by plastic, they can also be interpreted as indirect but favorable leading indicators for the hypothesis that plastic will replace steel in automobile body shells.

Table B.10 presents some events in the history of turbocharging for automobile engines. Superchargers driven by turbines in the exhaust were used in aircraft from about 1930. They have long been applied to diesel engines for trucks. However, they had not been applied to gasoline engines. The first few events in the history show addition of turbocharging to racing engines. Turbocharging was then incorporated in several prestige cars and finally was planned for widespread use in mass-produced automobiles.

Table B.11 presents some events in the history of diesel engines for passenger cars. The automotive pattern of first putting a new technology in prestige cars, then reducing the cost for low-end cars is apparent here. By now it is clear that automotive diesel engines will be strong competitors for gasoline engines in automobiles. Whether they will replace the gasoline engine completely has not yet been determined. It is interesting to note that fuel injection was transferred from diesels to gasoline engines, as shown in Table B.1, but electronic control was transferred in the opposite direction, its use in gasoline engines preceding its use in passenger diesels.

These histories of automotive technologies show several key features that technological forecasters should watch for in monitoring for precursors of technological change. These are:

1. The occurrence of "incomplete" inventions, which demonstrate feasibility of some technology but which need other elements before they can be deployed economically or in the actual operating environment. The fluidic-controlled fuel injector is an example of this.

2. The development of performance-improving supporting technologies needed by the basic technology. Examples in the automotive chronologies include engine control laws, instrumentation, and materials.

3. The development of cost-reducing supporting technologies needed by the basic technology. Examples in the automotive chronologies include manufacturing and testing technologies that reduce the cost of the basic technology.

4. The development of complementary technologies with which the basic technology must interact in order to be useful. Examples in the automotive chronologies include sensors for measuring automobile engine parameters.

5. The use of a technological advance in a prestige or high-performance application before its transition to general use. Examples in the chronologies include use in racing cars, in sports cars, and in prestige or luxury cars, and transition of testing technology from aviation to automotive uses. (Note that a similar transition was described in Chap. 7, in which composite materials were first used extensively in a performance-sensitive application, fighter aircraft, then appeared later in a cost-sensitive application, transport aircraft.)

6. The requirement of an incentive for using the technology, such as reduced cost or elimination of externalities. Examples in the chronologies include fuel economy and reduction of emissions. Fuel economy provides an incentive not only for more efficient engines but for lighter-weight vehicles. It is sometimes the case that a new incentive, such as fuel economy, brings about the adoption of innovations that were "waiting in the wings" but were not

needed before. Several of the weight-reducing innovations shown in the chronologies could have been adopted a decade or more earlier but had only a small payoff in an era of lower-priced fuel.

7. The existence of different origins for the leading indicators. Some of the indicators in the chronologies came from motor car manufacturers (e.g., Ford), while others came from suppliers to the manufacturers (e.g., Bosch, a supplier of fuel-injection systems). The former tend to be nearer-term indicators of coming change than do the latter, for obvious reasons.

In addition to these features, the chronologies also exhibit the transition of the actual technologies through some of the stages of innovation described in Chap. 1. This progression has utility not only for descriptive but for forecasting purposes, i.e., it may be useful not just to forecast the next stage of commercial introduction but also when a technology will achieve the next stage. These stages are:

1. *Proposal:* At this stage, the technology is described, and the architecture of the major components is outlined. No detailed calculations have been carried out. There is no assurance at this stage that the proposed technology does not violate some fundamental laws.

2. *Scientific findings:* At this stage, the technology is described, and detailed calculations have been carried out. At this stage, it has been determined that the technology does not violate any fundamental laws, whether or not it is within the current state of the art.

3. *Laboratory feasibility:* At this stage, a laboratory or experimental apparatus has been built that demonstrates the desired performance, at least under laboratory conditions, and that validates the design calculations.

4. *Operating prototype:* At this stage, a model intended not for laboratory demonstration but for use in the operational environment has been built. The prototype should demonstrate that the technology is producible and that it will perform as required in the hands of the expected users. Since the prototype may be a high-performance vehicle or prestige application, such as a space vehicle or a racing car, the "expected user" may be specially trained or skilled.

5. *Commercial introduction or operational use:* At this stage, the technology is being offered for sale or deployment in its intended use. This may be mass use, or it may be for a high-performance or prestige application.

These examples from automotive technology illustrate some typical patterns of technological change that can be utilized by technological forecasters to anticipate breakthroughs. (The material in this section is based on Martino (1987).)

10.6 Organizing for Monitoring

The process of monitoring for signals of coming technological change cannot be haphazard in nature. It must be organized so as to take into account the following three points. First, someone must be in charge. Second, the monitoring activity must be integrated into the overall organization. Third, the monitoring activity must make full use of the assets of the organization. Each of these points will be taken up in more detail.

A designated forecaster. Someone must be in charge of the forecasting and monitoring operation. There must be one person who is responsible for the activity. This need not be a full-time job. It may be merely one of several assignments carried out by a forecaster or staff planner. Nevertheless, there must be one person to whom people go either to obtain information or supply information. There must be one person who is responsible for either carrying out or supervising the steps in the monitoring process described earlier in this chapter. This advice to put someone in charge does not mean that a monitoring "empire" should be established. As discussed below, many of the people involved need not work for the person responsible for the activity. Nevertheless, there should be an organizational focal point for the activity, and there must be someone who knows what is going on and can answer questions about what hypotheses are being tracked, what data are being collected, and what additional information is being sought.

Integration with the organization. It is critical that the monitoring activity be integrated into the organization. In particular, the person in charge of monitoring must have close communication with the organization's top management. This is because the monitoring process must take into account the organization's objectives. If the forecaster does not know what the organization's objectives are, he or she cannot determine whether a particular event is relevant to those objectives, either as a threat or as an opportunity. If the forecaster is to provide a warning of possible technological threats and opportunities, he or she must know what would constitute a threat or an opportunity. Without knowledge of the organization's objectives, and without communication with top management, effective monitoring is not possible.

An additional feature of integration into the organization is the setting of thresholds. These cannot be set in isolation from the rest of the organization. In order to set thresholds, the forecaster must know what options are open to the organization. For instance, if there are low-cost hedging options available, the forecaster might be able to set up a series of thresholds. When one of the "lower" thresholds is crossed, some low-cost or long-lead-time action might be triggered. This would be done with full understanding that the warning may turn out to be a false alarm. The actions triggered by this low threshold should not represent a major commitment by the organization. A major commitment should be made only when a much higher threshold is crossed. Setting thresholds for warning must involve those executives in the organization who are responsible for taking various actions and who are aware of the costs and risks associated with taking or not taking particular actions. In some cases, an executive may set a threshold for providing further information. It is not always necessary to determine in advance what action should be taken when a particular threshold is crossed. In some cases, it would be sufficient to inform the proper executive, i.e., the issue hasn't gone away, in fact it's back in stronger form. The executive may determine the proper course of action at that time in the light of existing circumstances. The important thing is to set the threshold so the issue gets back on the agenda when another significant event takes place.

Making use of the organization's assets. The purpose of the monitoring function is to keep track of important events outside the organization. This does not mean that the people involved in the monitoring function must carry out that role. Most organizations have many people who are normally in contact with the "outside" as part of their jobs. It is important that the forecaster responsible for monitoring be able to utilize these people as part of the "information network" to provide signals about forthcoming technological change. For instance, salespeople are in regular contact with customers and can provide information about changed customer requirements and possible competitive threats. Purchasing agents are in contact with regular suppliers and can provide information about new raw materials and components. Scientists and engineers regularly attend conferences and read the technical literature. They can provide information about new discoveries, products, processes, or materials they learn about through their normal "technical awareness" activities.

The important point is to make effective use of the people in the organization who, in the normal course of their regular work, might expect to learn new and useful pieces of information. There are two aspects to this. First, the people in question must be motivated to seek information. Second, they must be motivated to pass it on to the forecaster who is responsible for monitoring.

The biggest problem with utilizing people such as sales representatives, purchasing agents, *research and development* (R&D) people, etc., is that the monitoring activity must cross organizational boundaries. The forecaster ordinarily does not have any authority over the people whom he or she needs to utilize. The forecaster cannot order them to do anything, and often cannot even provide rewards when they do turn over useful information. This is one more reason why the monitoring function must be closely integrated with the rest of the organization. It must be seen to have the support of top management. The monitoring function must be visibly important to those who gather and turn over the information, and to their immediate superiors. Without this visible importance, the people in contact with the "outside" will not provide information, and will not go out of their way to seek it because they will judge that no one cares anyway.

When people do an effective job of spotting and passing on important bits of information, the forecaster should reward them for it. This may be difficult, since the forecaster usually has no line authority over the people who have provided the information. Even so, the forecaster can provide some feedback to the source of the information. A memo of appreciation is certainly in order, even if nothing more can be done. Since very few pieces of information will actually trigger major actions, often not much can be said. Even so, the memo might state that "It makes such and such appear more likely," or "It shows we still have time before such and such becomes a problem." In either case it will be clear that the information was valuable and that the forecaster wants more as soon as the situation changes. In addition, when the forecaster submits a formal report on the basis of information provided by someone who is not a subordinate, the report can mention the name of the individual, as a means of rewarding him or her for making the effort to collect and transmit the information.

In short, people in the organization can be motivated to seek information by making it clear that the monitoring function is seen to be important by top management. They can be motivated to pass on the information they collect by rewarding them for doing so. Both these steps are important to a successful monitoring activity.

The logistics of monitoring. The forecaster conducting a monitoring program must successfully keep track of a wide range of pieces of information. Most of the pieces of information entering the system will be in the form of memos or letters and copies of articles or documents. For each such document, it is important to record (perhaps with the document itself) the name of the person who provided it, the original source of the information (e.g., what journal or the name of the contact from whom the provider got the information), the names of the people involved in the information (e.g., authors of an article), the company or agency in which or for which the activity being reported was carried out, the date of both the document itself and of the event described, and references to earlier items about the same event or about the same outside persons (this is especially important if the item is a follow-up on an earlier item).

Even more important then the items themselves, however, is information about the pattern(s) the items are believed to support or deny. An essential part of the screening process described earlier is the linking of new items to older items already received. An essential part of the evaluation process is the generation of hypotheses about what the items mean. The monitoring system should contain a description of each hypothesis being tracked. This description should include references to those items of information related to the hypothesis, the confirming and/or disconfirming information currently being sought, the current level of confirmation of the hypothesis, and the threshold(s) established for taking further action.

Many successful monitoring programs have been conducted using nothing more elaborate than a set of file folders into which new pieces of information are stored. Each folder will hold all those items relevant to a particular hypothesis. If an item is relevant to more than one hypothesis, it can be duplicated and stored in each appropriate folder. However, the low cost of personal computers means that they can be very helpful in automating many of the routine tasks associated with operating a monitoring system. The computer can store information that would have to be typed anyway. Bulky documents might have to be stored as physical documents. Short documents might even be scanned into a computer file. In either case, the computer can be used to index the documents. However, the computer should not be looked upon as an end in itself. It is merely a tool to replace the nonintellectual work load involved in maintaining a monitoring system. The computer possesses no magical "data retrieval" system that will eliminate the need for careful thought and analysis on the part of the forecaster.

By proper organization of the monitoring system, the forecaster's workload can be kept to a reasonable level. It is entirely feasible to operate an effective monitoring system as a part-time activity for one person, provided an efficient means of handling the logistics is developed and provided most of the information comes from other people in the organization.

10.7 Summary

The forecasting methods given in previous chapters seem to require continuity between the past and the future in order to forecast the future. The question then arises, "Can a breakthrough be forecast?" The answer is "Yes," but it depends on another kind of continuity between the past and the future. Breakthroughs in technology do not come as "bolts from the blue." On the contrary, they are the end result of a chain, or even a network, of precursor events, and these events give warning that a breakthrough is coming.

Forecasting a breakthrough, then, involves a systematic search for these precursors, coupled with an evaluation of the significance of the precursors found. The forecaster seeking advanced warning of a breakthrough must search all the relevant sectors of the environment in order not to miss important signals of coming breakthroughs. Moreover, the signals found must be synthesized into possible patterns of change, and the forecaster should continue to search for additional signals suggested by the hypothesized patterns. It is important to search for both confirming and disconfirming signals.

References

Bright, J. R., "Evaluating Signals of Technological Change", *Harvard Business Review* (Jan.–Feb., 1970).

Davis, R., "Organizing and Conducting Technological Forecasting in a Consumer Firm," in J. R. Bright and M. Schoeman (eds.), *A Guide to Practical Technological Forecasting,* Prentice Hall, Englewood Cliffs, NJ, 1973.

Martino, J. P., "Using Precursors as Leading Indicators of Technological Change," *Technological Forecasting & Social Change,* 32:341–360 (1987).

For Further Reference

Heiman, G., *Jet Pioneers,* Duell, Sloan and Pierce, New York, NY, 1963.

Neville, L. E., *Jet Propulsion Progress,* McGraw-Hill, New York, NY, 1948.

Whittle, F., *Jet: The Story of a Pioneer,* Frederick Miller, London, England, 1953.

Problems

1 Which of the following devices or techniques represent breakthroughs as defined in the text? What functional capability was advanced? What limitations of the prior techniques were surpassed by the breakthrough?

Bessemer steel-making process
Logarithms
Typewriter
Vaccination
Canning
Magnetic recording

2 Review the history of the invention of the airplane from the standpoint of its inevitability. What were the achievements and contributions of Sir George Cayley, Sir Hiram Maxim, Otto Lilienthal, Octave Chanute, Samuel P. Langley, and the Wright brothers?

3 Review the history of the self-propelled road vehicle, which culminated in the development of the automobile. What were the three major contending power sources during the period 1890 to 1915? What were the advantages and disadvantages of each for automobile propulsion? What precursor events were there that indicated the direction automotive technology would finally take?

Chapter

11

Combining Forecasts

11.1 Introduction

Each individual forecasting method has its own strengths and weaknesses. In many cases a forecast can be improved by combining separate forecasts obtained by different methods. By doing this, the forecaster avoids problems of trying to select the one best method for a particular application. In addition, the strengths of one method may help compensate for the weaknesses of another.

This chapter covers some of the more important methods for combining forecasts. It includes combined forecasts of a single technology and the means for combining forecasts of several different items into a composite whole. Particular emphasis is placed on the problem of consistency in combining forecasts of different items into a single overall forecast.

11.2 Trend and Growth Curves

The history of a technology will often show a sequence of different technical approaches. Each new approach usually represented a breakthrough, since it surpassed the upper limit of the previous approach. Thus, a plot of the performance of individual devices in the history of the technology might show a succession of growth curves, one for each technical approach, while the overall performance of the technology is given by an exponential trend. An illustration of this is shown in Fig. 11.1, which shows the history of aircraft speed records as a succession of growth curves on which is superimposed an exponential trend.

The first growth curve, for wood and fabric aircraft, had an upper limit of approximately 300 mi/h, set by the physical limitations of the structure. The second growth curve, for aircraft with metal skins and structure, had an upper limit of approximately 520 mi/h, set indirectly by the speed of sound. Even though the aircraft itself was flying at speeds below the speed of sound, the propeller added its rotational speed to the forward speed of the aircraft, and encountered "compressibility" problems. The third growth curve, for subsonic jets, had a top speed of approximately 750 mi/h, set directly by the speed of sound. In this case, the entire aircraft encountered compressibility problems. The fourth growth curve, for supersonic jets, was achieved by tailoring the aircraft's shape to reduce drag at transsonic speeds (the Whitcomb "area rule"). However, this growth curve had an upper limit of approximately 1750 mi/h, set by the heating effects of high-speed flight that weakened the aluminum structure.

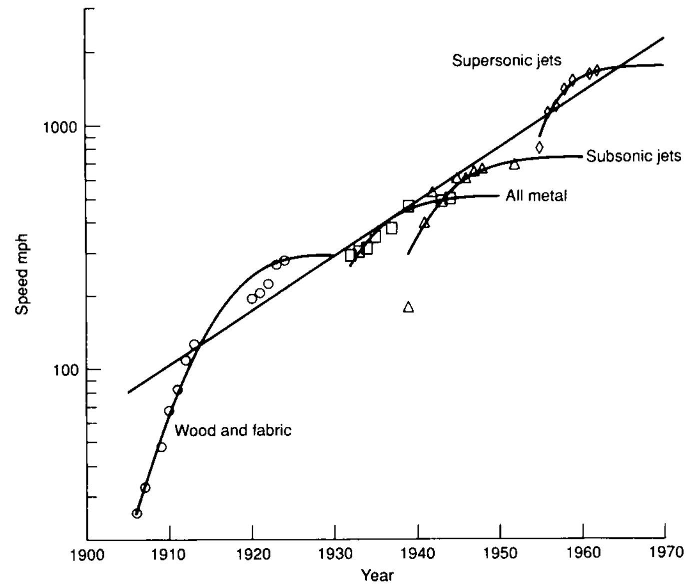

Figure 11.1 Aircraft speed, growth curves, and trend.

As the growth curve for each technical approach to aircraft design and construction began to fall behind the long-term trend, the technical approach was replaced by a successor technical approach that had a higher upper limit and was capable of "keeping up with the trend" for some additional period of time.

To forecast a technology in which a succession of growth curves produces a long-term trend, it is necessary to project both the overall trend and the relevant growth curves. Projecting the trend will give an estimate of future levels of performance to be expected. Projecting the growth curve for the current technical approach will provide information on whether the technical approach is capable of keeping up with the trend. If the technical approach is capable of keeping up with the trend for some time yet, then a near-term threat to the approach is unlikely (but not impossible). However, if the growth curve is nearing its upper limit and must soon fall below the trend (or has already begun to do so), then a successor technical approach will soon replace the current one. If such a successor technical approach can be identified, even if it is only in an early stage of development, projecting its growth curve can help predict the time when it will compete successfully with the current approach. In this connection, it is worth noting the intersecting growth curves in Fig. 5.1, which shows speeds of propeller-driven and jet aircraft. The earliest jet aircraft were not yet competitive with contemporary propeller-driven aircraft. Yet it was clear they were a potential successor approach, since they had an inherently higher upper limit than did propeller-driven aircraft. Thus, using both trend and growth curves can provide much more information about the future course of some technology than can the use of either by itself.

11.3 Trend and Analogy

External events can cause a deviation, temporary or permanent, from a long-term trend. Many of the technologies described in the data of App. A show the effects of external events such as the Great Depression and World War II.

Figure 11.2 shows the efficiency of coal-burning electric plants, in kilowatt · hours per pound of coal consumed. Also shown is a Gompertz curve fitted to the data from 1920 through 1975. The upper limit was taken as 1.33, corresponding to 35-percent efficiency from boiler to busbar. The effects of the Great Depression, starting in 1930, are clear. The curve is above the data. The hiatus in power plant construction during World War II continues this depression-era effect. In 1950, recovery is shown as having started. By 1955, new power plants and the retirement of old plants have moved the average efficiency above the curve. After 1965, however, we see the effects of another external event. Efficiency does not just stay level, as it did during the Great Depression, but actually declines. This is the result of environmental regulations, which required scrubbers on the exhaust stacks of power plants. These scrubbers consumed up to 5 percent of the plant's total output, reducing the net electrical production per pound of coal.

Figure 11.3 shows the utilization of inanimate automotive horsepower in the United States (essentially prime movers and factory machinery). From 1850 through 1930 the data follow an exponential trend with little deviation.

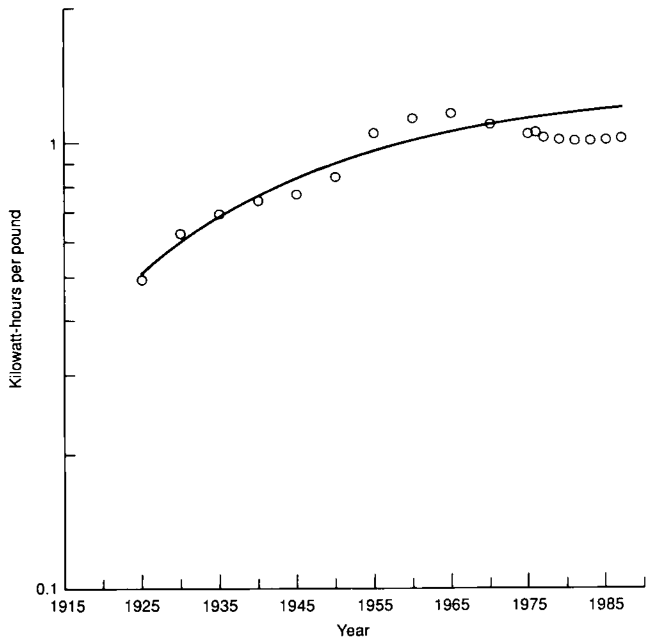

Figure 11.2 Efficiency of coal-burning electric plants.

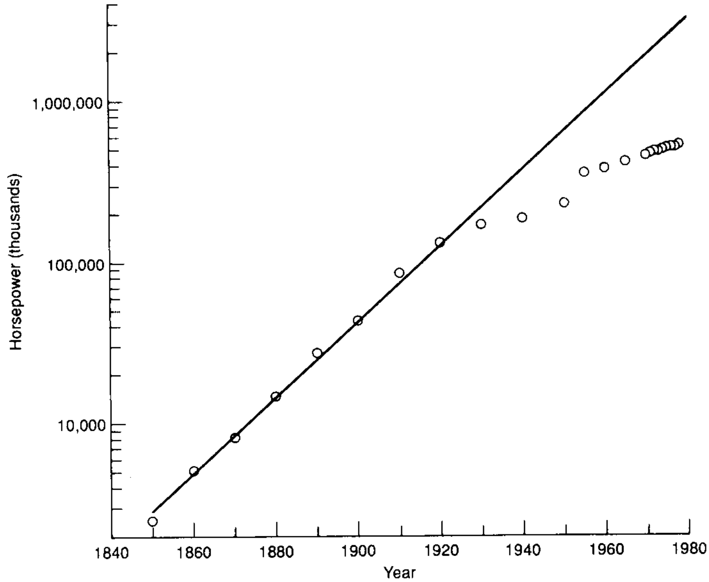

Figure 11.3 Inanimate nonautomotive horsepower.

From 1930 through 1945, the actual utilization falls well below the projection of the previous trend. As with the electric power plants, this is in large measure due to the Great Depression and World War II. Recovery was under way by 1950, but starting with 1955, the rate of growth slowed again. In part, the slow growth after 1970 is due to the rise in the price of energy. In part, unfortunately, the behavior of the plot after 1955 is an artifact resulting from changes in definitions by the Census Bureau. Hence, the slope after 1955 cannot be compared directly with the slope prior to 1930 to see if the United States will "get back on the trend." The new data does seem to show a trend, but it may be too soon to tell whether this is a long-term trend or simply the effect of post-1970 energy prices. In any case, the effects of the various external events are clear, and they might be used as analogies to forecast the effects of similar future events.

As these examples show, external events can have an impact on both trends and growth curves. However, growth curves tend to have shorter time scales than do trends; hence, the impacts may not be so easily observed. Nevertheless, keeping in mind the differences in time scale, analogies can be combined with both growth curves and trends.

The purpose of combining analogies and trends is to forecast the magnitude and timing of the deviation from a trend that will be caused by some external event. (In all this discussion it is assumed that the requirements of a formal analogy have been met.) The simplest case is that of an event analogous to some earlier event that has already had an impact on a trend. The results of the impact are then forecast as analogous to the results of the previous event.

A more complex situation involves analogies not only between events that affect each other but between technologies as well, i.e., the historical case may show the impact of an event on one technology, and the current case is the result of an impact of an analogous event on what is considered to be an analogous technology. The impact of the anticipated event is then taken to be analogous to the impact of the model event on the model technology.

The combination of trends and analogies can be applied in reverse as well. When a deviation from a trend has been observed in the history of some technology, the forecaster may try to identify the cause of the deviation. In some cases the cause will be fairly obvious, e.g., war, depression, or natural catastrophe. In other cases, the cause will be specific to the technology, e.g., pollution controls. In the latter case, it will be necessary to carry out a historical investigation. Once the cause of the deviation is determined, however, it can be used as the basis for an analogy to forecast future deviations from analogous causes.

11.4 Components and Aggregates

The forecaster often deals with an aggregate that can be decomposed into components. For instance, total U.S. energy production can be decomposed into production of coal, petroleum, natural gas, nuclear power, etc. This raises the question of whether the components should be forecast individually and added together to obtain the aggregate or whether the aggregate should be forecast by itself and the components derived from it in some manner. In some cases, the aggregate appears to have a "lawful" behavior, while the components are to some degree interchangeable or in competition with one another. In such cases, a "top-down" forecast is appropriate, with the aggregate being forecast first and the total allocated among the components. In other cases, the aggregate exists only in a statistical sense. It is the components that seem to have "lawful" behaviors or to be independently explainable. In such cases, a "bottom-up" approach is appropriate, with the components being forecast individually and added together to obtain the aggregate.

As an example of the top-down case, consider a forecast of the amount of food preserved by each of several technologies such as canning, freezing, and dehydration. The total amount of preserved food consumed in the United States is determined by population and income. Hence, this total should be forecast first. The total should then be allocated among the different preservation technologies on the basis of cost, taste, ease of preparation, and other competitive factors. It would make no sense to forecast each preserved food separately, then add them to get a total, as though people bought a particular kind of food for its own sake rather than because it nourished them.

As an example of the bottom-up case, consider a forecast of the number of aircraft engines manufactured in the United States. This total is made up of two largely independent components: engines for civil aircraft (airlines and general aviation) and engines for military aircraft. There is little or no connection between the market for civil engines and the demand for military engines. Each of these two components can change independently of the other. Hence, it is necessary to forecast each separately and add them to obtain the total.

In some cases, however, forecasts of components and aggregates can be combined fruitfully. Consider the problem of forecasting the number of households in the U.S. with *cable television* (CATV). The forecaster could attempt to forecast this total directly. However, finding a suitable "growth law" would be difficult. So long as the population continues to grow, there is no obvious upper limit. Hence growth curves are not suitable. Ultimately, the number of households with CATV cannot grow faster than the number of households, which in turn is linked to population growth. Once every household has CATV, the same growth law that describes the number of households will also describe CATV growth. However, the real value of a forecast is during the time when the number of households with CATV has not yet reached saturation but is coming into equilibrium with the population. Hence, forcasting during this period, when different growth laws describe population growth and CATV ownership, is the challenge.

This situation is typical of cases in which forecasts of components and aggregates can be combined. In Chap. 4, a forecast was made of the fraction of households with CATV. Since a fraction has an obvious upper limit, a growth curve was a suitable approach. If the total number of households is then forecast separately as an aggregate, the number of households with CATV can be calculated as the product of the two forecasts.

There are many cases in which we are interested in the total number of installations of some kind but a direct forecast does not seem feasible. In such cases, it is usually better to forecast the potential number of installations (e.g., total population, firms, households, or whatever) and to forecast separately the fraction that will utilize the technology. In these cases, combining forecasts of the total and of the fraction will be better than attempts to forecast directly the number using the technology. [Armstrong (Chapter 9) described the use of what he called "segmentation methods" of forecasting in many different topic areas. In particular, segmentation methods include producing separate forecasts for components and for aggregates, then combining the two.]

It should be noted that when a technological forecaster requires forecasts of populations, households, etc., the best source is the U.S. Bureau of the Census. First, the use of Census Bureau forecasts avoids the need to defend the forecasts; no one else has more credible forecasts. Second, population forecasting is a specialized topic with its own problems and techniques. Technological forecasters are well advised to stick to their own specialties. The expertise of other forecasters in their fields of specialization should be utilized.

11.5 Scenarios

Sometimes the forecaster has a set of forecasts that are related in some way or that all bear on a particular situation. Taken together, they provide an overall picture of an environment, as opposed to the small segment of the environment captured by each of the forecasts individually. Scenarios are used to combine these individual forecasts into a composite whole.

A scenario is a written description of a situation. Kahn and Wiener (1967) defined "scenario" as follows:

> Scenarios are hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision-points. They answer two types of questions: (a) Precisely how might some hypothetical situation come about, step by step? and (b) What alternatives exist, for each actor, at each step, for preventing, diverting, or facilitating the process?

Scenarios have three purposes:

1. To display the interactions among several trends and events in order to provide a holistic picture of the future
2. To help check the internal consistency of the set of forecasts on which they are based
3. To depict a future situation in a way readily understandable by the nonspecialist in the subject area

Technological forecasters frequently make use of scenarios for one or more of these purposes. Scenarios can be very effective in providing a composite picture of some future to be used as the background for a decision. The scenario can define the alternatives and the choices that must be made among them, as well as portray an environment into which a technological development must fit. In particular, the scenario can explore the effects of the technology on the environment and vice versa. A range of scenarios can be used to test the "robustness" of some technology strategy by depicting a variety of environments, from favorable to unfavorable, with which the strategy must cope.

Because of their varied uses, no single scenario can serve all purposes. Therefore, there is no single "correct" way to write one. Writing a scenario is to some degree a work of art, akin to writing a novel. Nevertheless, there are procedures that can help the forecaster go about his or her task of writing scenarios in a systematic manner. Following the procedure will not guarantee a good scenario; however, it will usually produce a better scenario than could be obtained without following the procedure.

The forecaster should follow these steps in writing a scenario:

1. *Develop a framework for the scenario:* What happens in each sector of the environment? What trends should be considered? Will they continue, or will they change? If there is change, when and in what way? Are there choices facing society that could make a big difference to the technology being forecast? If so, what are the critical decisions and at which points in the scenario must they be made? Who will be responsible for making the decisions?

2. *Forecast the technology (or technologies) to be considered:* When will it be deployed? On what scale will it be deployed? How rapidly will it be

adopted? What are the impacts from the technology on the trends and events in the framework and the impacts from the framework on the technology? Are there critical decisions to be made about the technology? When do these come in the scenario? Who are the decision makers?

3. *Plot the scenario(s):* For each scenario, select a sequence of events and decisions. Identify events that could trigger decisions. Check the sequences for consistency.

4. *Write the scenario(s):* Fill in the outline developed in the previous steps by a verbal narrative describing the events.

To illustrate this process, we develop a scenario for the resurgence of nuclear power in the United States. Each of the steps will be presented briefly.

1. *The framework:* The following background trends are assumed for the purposes of the scenario:
   a. Concern about protecting the environment, particularly acid rain and toxic emissions from coal-burning power plants, will continue or grow.
   b. Concern about the "greenhouse effect," created by carbon dioxide from burning fossil fuels, will continue or grow.
   c. Concerns about the security of America's energy supply will continue because of the threat of instability in the Middle East.
   d. Concern about the U.S. balance of payments, creating pressure to reduce energy imports, will continue.
2. *Technology developments:* The following developments in nuclear power are assumed for the purposes of the scenario:
   a. A breeder reactor will be developed successfully.
   b. The political ban on recycling uranium and plutonium from power reactors will end.
   c. With fuel recycling possible, nuclear power plant waste consists only of relatively short-lived fission products. The decision is made to store these aboveground in the New Mexico desert.

   The following developments and events in the U.S. energy industry are assumed for the purposes of the scenario:
   a. An earthquake ruptures the Alaska pipeline.
   b. Terrorists attack an oil tanker in an American harbor.
   c. Smog in Los Angeles causes numerous deaths and illnesses.
3. *The plot:* The following specific events (with dates) are assumed for the purposes of the scenario:

1995	Nuclear power accounts for 25 percent of U.S. electric power production.
1995	Coal accounts for 33 percent of the total U.S. energy production.
1996	The average temperature worldwide rises nearly one degree over the preceding decade; the greenhouse effect is apparently under way.

1997	The U.N. General Assembly votes to demand that industrialized nations "freeze" their level of use of fossil fuels; thus, developing countries can continue to industrialize.
1998	Terrorists, firing antitank rockets from a motorboat, attack an oil tanker unloading at Bayonne, NJ. The resulting oil spill and fire causes extensive damage to the harbor and contaminates the Hudson River as far south as lower New York Bay.
1999	An earthquake in Alaska ruptures the petroleum pipeline, resulting in massive contamination of the Alaskan tundra.
1999	A month-long temperature inversion over Los Angeles causes a high level of smog. Hundreds of people with respiratory problems die. Thousands are made ill.
1999	A Green Party is formed that advocates "small is beautiful," announces it will campaign on a platform of dismantling heavy industry and replacing all other power sources with windmills and solar energy.
2000	Republican and Democratic presidential platforms call for continuing industrialization, with a shift away from fossil fuels.
2000	The Green Party captures 15 percent of the Presidential vote nationwide and elects senators or representatives in California, Oregon, Colorado, and Alaska.
2000	The majority of voters support continued economic growth and vote down referenda calling for closing of fossil and nuclear power plants.
2001	The first executive order of a newly inaugurated president repeals the ban on nuclear fuel recycling.
2002	The nuclear industry develops a breeder reactor that is cost competitive with imported petroleum, thus greatly expanding the nuclear fuel supply.
2003	The first nuclear fuel recycling plant goes into operation; over 97 percent of spent reactor fuel can be recycled and less than 3 percent must be given long-term storage.
2005	Congress votes to subsidize New Mexico to accept long-term storage of nuclear reactor wastes; the justification is to compensate for tax loss from desert land taken off tax rolls.
2008	Nuclear power accounts for 75 percent of U.S. electric power production; most of the rest is accounted for by domestic natural gas.
2010	Coal use drops to 5 percent of the total U.S. energy production; Congress votes for retraining of and resettlement benefits for coal miners.

4. *The scenario:* The plot outline has been developed. The next step is to write a story around the bare bones of the plot outline. The purpose of the story is to portray the plot in dramatic fashion for someone who is not expert in the technology area. The story is intended to make the plot vivid, colorful, and memorable. It should appear plausible to the reader, even if it does not compel agreement.

There are several approaches to the writing of scenarios. Each has its advantages and disadvantages. These approaches are as follows.

a. Looking backward: The scenario is written from the perspective of a specific time in the future, after the events have all taken place. The "frame" for the scenario is often a speech, a historical essay, or simply the reminiscences of a participant in the action. The form of the story is "how we got to where we are." The person telling the story need not have been involved in every part of the action. He or she may simply be reporting what has been learned from others. This approach tends to be global or even Olympian in nature. It is a good means for presenting a "broad brush" treatment. However, it may fail to get the reader involved. The forecaster must insert drama into the scenario by describing the problems that were encountered along the way to "where we are," without giving away the solutions in advance. However, the solutions must be plausible once they are revealed to the reader. (Edward Bellamy's *Looking Backward* is of course the classic example of this type of scenario.)

b. Viewpoint character: The scenario is written from the viewpoint of someone who is seeing, undergoing, or taking part in the events as the scenario unfolds. All the reader learns is what the viewpoint character sees, hears, is told, etc., and only as the story actually unfolds. This approach is good for getting the reader involved and is good for presenting the personal impact of the events portrayed, especially if the viewpoint character is one of those "hurt" by the events described. It is almost impossible, using this approach, to present an overall perspective on the sequence of events. Moreover, this approach is hard to use if the sequence of events covers a long period of time. For instance, if the events take place over more than a decade, it will be hard to invent a character who would be "on the scene" for that whole time. (Most novels are written in this form.)

c. God's-eye view: The scenario is presented as events unfold, as in the "viewpoint character" approach. However, the perspective is global. No single person is seeing it all. This approach is good for presenting an overall perspective. However, it may be a dramatic failure, since there is no one person seeing or telling the story. This approach may fail to "bring the reader into the story." One effective way of handling this type of scenario is as a series of newspaper stories, stories by on-the-spot TV or radio reporters, extracts from letters, etc. The personal element can sometimes be brought in effectively through techniques such as presenting reporters interviewing actual participants or presenting conversations on radio call-in shows. This way of writing the scenario may turn out to be choppy, but it can add drama to an otherwise dull recitation of events. Although a scenario written in this way jumps around from one viewpoint character to another, it presents the story as a series of vignettes, each from the perspective of a human being who is seeing history unfolding and reacting to it. (Orson Welles' (in)famous radio dramatization of *War of the Worlds* is an example of this approach to scenario writing.)

d. Diary: In this approach the story appears as a series of diary or journal entries, or extracts from personal letters, written shortly after the events described. The advantage of this approach is that the diary writer can not only describe events as he or she saw them but must include information not available to him or her at the time of the events, i.e., what was happening to other people at the same time. This approach can also purport to use extracts from several people's diaries, thus, changing viewpoint from time to time. Each entry is extracted from the journal (or personal letters, etc.) of whoever was "where the action was" at the time in question. A major disadvantage of this approach is that diary entries tend to filter out the drama of the situation. This can be turned to an advantage if the forecaster can "underplay" some dramatic events. (Bram Stoker's novel *Dracula* is a well-known example of this type of scenario; the matter-of-fact diary entries lend verisimilitude to what is otherwise an incredible tale.)

To turn our plot outline into a scenario, we now select one of these approaches and actually write out the events in the form chosen. For the nuclear power scenario, this will be left as an exercise for the student (Prob. 6). The approach cannot be chosen arbitrarily. Different problem situations lend themselves better to one of the above approaches than they do to the other approaches. The forecaster must select the best scenario-writing approach on the basis of the plot itself, the audience, and the effect desired.

The following are the main problems the forecaster must solve in writing a scenario, regardless of the approach taken. The forecaster must take special pains to overcome them.

1. *Transition from the present:* The trajectory by which the future situation is reached must be developed carefully, especially if the scenario is to be a guide to action. The sequence of events and the specific choices that must be made must be portrayed clearly.

2. *Plausibility:* A scenario will be plausible only if the chain of events that leads to it also appears plausible. A chain of unlikely events ("acts of God") does not make for a plausible scenario. This does not mean, however, that every event in the scenario must be one that has a high likelihood of occurring. It is sufficient if events are members of plausible classes, such that some member of the class is sufficiently likely to occur. For instance, suppose a scenario involves weather-control technology. The probability of a hurricane striking a specific U.S. city is quite low. Nevertheless, there is a fairly high probability that a hurricane will come ashore *somewhere* in the United States in any given year. Hence, the fact that a specific city is unlikely to be hit does not mean that the forecaster cannot use a hurricane striking a specific city as a trigger event for a decision to develop (or test or deploy) the technology.

3. *Reversal of trends:* Some scenarios will require that historical trends be reversed. Major trends have in fact been reversed, but the scenario writer must provide the reader with some reason to believe that the reversal of a trend is plausible under the circumstances of the scenario.

4. *Convincing linkages among events:* Frequently the scenario writer has to provide a reason for why a particular event will take place rather than some other event or why one decision is made instead of another. These choices are supposed to be triggered by other events in the scenario. The linkages between causal events and subsequent events or decisions must be convincing. The reader must be satisfied that the causal event really would have the effect claimed.

5. *Motivation of actors:* Major actors will make certain decisions during the course of events depicted in the scenario. The writer must provide proper motivation for these decisions, i.e., the writer must make it plausible that the person in question would actually make that decision under those circumstances. ("My constituents demanded it." "I'll show them!" "My husband burned the toast that morning.") Providing proper motivation is particularly difficult in writing "best-case" or "worst-case" scenarios. With best-case scenarios, it is hard to avoid blandness if everyone does everything right without any struggle to avoid doing the wrong things. With worst-case scenarios, it is often hard to imagine someone knowingly and deliberately taking the actions needed to produce these extreme scenarios.

Despite these problems, the scenario is a popular means for combining forecasts into a composite whole. This popularity is due in no small measure to the power of a well-written scenario to make an otherwise dull forecast come to life and appear in vivid colors. The technological forecaster should be prepared to take advantage of this power of the scenario when it is appropriate.

11.6 Cross-Impact Models

In Chap. 9, we discussed probabilistic methods, in which we introduced the idea that a particular future event may not be certain to occur but may have a probability associated with it. More generally, we can say that two elements of every forecast are the time of the event and the probability of the event. (A third element, of course, is the identity of the event itself.) However, the forecast of timing and probability are often based on assumptions about some other events that might occur between the present and the time forecast for the event of interest. A particular forecast will assume certain outcomes for these intermediate events. If the intermediate events have different outcomes, the original forecast becomes irrelevant. It may be a correct deduction from its assumptions, but it no longer applies because its assumptions are no longer valid.

Cross-impact models are a means of taking into account the dependence of some forecasts on other forecasts. A cross-impact model consists of a set of events, each assigned a time and a probability. In addition, a cross-impact model includes a set of "cross impacts" among the events. If the forecast for event E_2 assumed a particular outcome for an earlier event E_1, then E_1 is said to have a cross impact on E_2. If E_1 has a different outcome than the one assumed in the forecast of E_2, the forecast for E_2 will change. The cross impact shows up as a change in timing, in probability, or in both for E_2. The impact may come from nonoccurrence if the earlier event was forecast to occur. Conversely, the impact may come from occurrence if the earlier event was forecast not to occur.

In principle, there may be cross impacts between every pair of events in a cross-impact model. For N events in the model, there may be $N(N - 1)$ cross impacts (each event may impact every event but itself). In practice, the number of cross impacts will be much smaller than this. If the outcome of E_1 will be determined before E_2 could possibly occur, then a cross impact of E_2 on E_1 is impossible and can be ignored. A typical cross-impact model may have 200 events. There would be a maximum of 39,800 possible cross impacts. In practice, however, the number of cross impacts is more likely to be in the range of 1000 to 2000.

In use, the cross-impact model simulates the future implied by the set of events and cross impacts. The event with the earliest forecast date is selected and its outcome is determined (for instance, by drawing a random number). Depending on the outcome, adjustments are made in the timing and probability of all the other events that receive impacts from this first event. The next event is selected, its outcome is determined, and adjustments made in the remaining events that receive impacts from it (and that have not, of course, had their outcome already determined). This process is completed when the outcome for all the events has been determined.

A cross-impact model, then, looks upon the future as a collection of interconnected events. As time passes, the outcome of certain events is determined. The cross impacts from these events, as they are determined, cause the forecast timing of the remaining events to be earlier or later, and the forecast probability to be higher or lower, than originally forecast. Eventually, the outcome of every event is decided, and it either does or does not occur.

The program MAXIM is a cross-impact model that can be used to carry out the procedure described above. The user supplies events with timing and probability, and cross impacts among the events. MAXIM can then be used to simulate one or more sequences of events, and provide statistical information about the sequences.

To illustrate the workings of a cross-impact model, we consider a forecast of space exploration using MAXIM. First, we identify some possible future events related to exploration and development of near-Earth and cis-lunar space. These events are as follows:

Heavy Lift Vehicle (HVY_LIFT_VEH): A launch vehicle intended to transport heavy loads into *Low-Earth Orbit* (LEO) (approximately 300-mi altitude). The vehicle may be piloted but is primarily intended for transportation of cargo rather than passengers.

Second generation shuttle (2ND _ GEN _ SHUTTLE): A follow-on vehicle to the current Space Shuttle. This vehicle is intended to transport both passengers and cargo into LEO.

Single Stage to Orbit (1 _ STAGE _ TO _ ORBIT): A vehicle that can achieve orbit without the need for boosters or multiple stages. The entire vehicle achieves orbit and is recovered after deorbit.

Return to the moon (RETURN _ TO _ MOON): A manned flight to the moon, not as a one-shot effort but as the prelude to developing a lunar habitat.

Commercial space station (COM'L _ SPACE _ STN): A space station established in LEO; the station is financed privately and intended for commercial activities.

High-value space manufacturing (HI _ VAL _ SPC _ MFG): The manufacture in space of small quantities of high-value materials such as pharmaceuticals, high-purity alloys, and defect-free semiconductors. The manufacture of these high-value products will take advantage of the vacuum, of the high-intensity solar radiation, the absence of gravity, or other space conditions. These products will be things that either cannot be made at all on Earth or can be made only at much higher cost than in space, even taking launch and recovery costs into account.

Permanent moon base (PERM _ MOON _ BASE): The establishment of a permanently staffed base on the moon. Activities at the base might include scientific research, lunar exploration, and extraction of lunar materials.

Orbital transfer vehicle (ORBITAL _ TRANS _ VEH): The development and deployment of a class of vehicles that operate only in space and never return to Earth once they are launched. Ultimately they might be assembled or even manufactured in space. Their function is to transfer cargo and passengers from one orbit to another.

Lunar materials used in geosynchronous orbit (LUNAR _ MTLS _ GEO): The use of materials extracted from the moon for construction, manufacturing, or other purposes in *geosynchronous orbit* (GEO), approximately 22,000 miles above the Earth's surface.

Lunar materials used in Low Earth Orbit (LUNAR _ MTLS _ LEO): The use of materials extracted from the moon for construction, manufacturing, or other purposes in LEO.

Orbital factory (ORBITAL _ FACTORY): The establishment in LEO of satellites for manufacturing other than strictly high-value items. The products manufactured would be intended for use in space; manufacturing "on-site" would be cheaper than bringing them from Earth. This manufacturing might include the fabrication of Orbital Transfer Vehicles and the manufacture of components for satellites and even comparatively low-value items that could still be sold competitively on Earth.

Solar power satellite (SOLAR _ POWER _ SAT): A satellite in GEO that collects solar energy, converts it to microwaves or directed laser energy, and beams it to a receiving station on Earth, where it is distributed as electricity. The satellite might also transmit power to lower-altitude space vehicles or installations that cannot, for one reason or another, deploy large solar collectors.

Each of these events has a specific identity. However, they need not all be forecast in the same way. There are several possible types of forecasts, which will now be illustrated with examples from the above list.

Single-time event. This event is forecast to occur at a specific time, with a specific probability, e.g.,

Event	Heavy lift vehicle
Year of occurrence	1998
Probability of ever occurring	0.6

Event with a span of time. This event is forecast not for a single date but as falling within a span of time. It is necessary to forecast the first year of the span, the length of the span, the probability of its ever occurring, and the conditional probability within the time span (i.e., on condition that the event does occur, what is the probability of it occurring in a given year), e.g., for a commercial space station with a probability of ever occurring of 0.6:

Year	Conditional probability
2002	0.3
2003	0.2
2004	0.2
2005	0.2
2006	0.1

Note the distinction between the probability of ever getting a commercial space station (0.6) and the conditional probabilities that if there is one, it will occur in specific years. Note also that the conditional probabilities must sum to 1.0.

Set of mutually exclusive events. There may be sets of events such that at most one member of the set can occur. This may happen because the events are contradictory, because they represent alternative choices, or for other reasons. It is necessary to forecast the date for each event, the probability that one member of the set will actually occur, and the conditional probability for each event within the set, e.g.,

Event*	Year	Conditional probability
2ND _ GEN _ SHUTTLE	1999	0.4
1 _ STAGE _ TO _ ORBIT	2004	0.6

*Probability of set: 0.5.

The idea here is that a choice will be made between building a second generation shuttle or building a single stage to orbit vehicle. The second generation shuttle would be an upgraded and improved version of the Space Shuttle of the 1970s and 1980s. However, it would still require boosters. The single stage to orbit vehicle, however, will take longer for the technology to be developed. Note the distinction between the probability of getting one member of the set (0.5) (i.e., the probability of getting *either* the second generation shuttle or the single stage to orbit vehicle) and the conditional probabilities that if a member of the set does occur, which event it will be. Note also that the conditional probabilities must sum to 1.0.

Table 11.1 lists a complete set of events for this model. This table is in MAXIM's event-data worksheet format provided on the program disk. Each event is shown with the first year identified, the length of the span, and the conditional probabilities for the years within the span. Note that MAXIM allows time spans up to 10 years, not just the 5 years shown in the table. The two events making up a set are designated as a set by giving them the same set number. The probability for any event of the set is shown as well as the conditional probabilities within the set. All the rest of the events are single-time events.

These events are not independent of one another. They cannot happen in isolation. For instance, if we never return to the moon, clearly there is no possibility of using lunar materials for construction in either GEO or LEO. The dependencies among these events are illustrated in Fig. 11.4.

The forecasts given in Table 11.1 are made using various assumptions about earlier events. In most cases, if an event is dependent upon another event, it is assumed that the earlier event will actually happen, i.e., establishment of a moon base is forecast for 2005 with probability 0.6 on the assumption that manned return to the moon will actually occur in 2000. Hence, if any of the prior events fail to occur, the events dependent upon them will be affected. The impact of the failure to occur will be a change in the timing, prob-

TABLE 11.1 Events for Space Cross-Impact Model

				Probability		Conditional probability for year within span				
Event name	Set	Year	Span	Of occurring	In set	1	2	3	4	5
HVY_LIFT_VEH		1998	1	0.6						
2ND_GEN_SHUTTLE	1	1999	1	0.7	0.4					
RETURN_TO_MOON		2000	1	0.85						
COM'L_SPACE_STN		2000	5	0.6		0.3	0.2	0.2	0.2	0.1
HI_VAL_SPC_MFG		2003	1	0.8						
1_STAGE_TO_ORBIT	1	2004	1	0.7	0.6					
PERM_MOON_BASE		2005	1	0.6						
ORBITAL_TRANS_VEH		2010	1	0.5						
LUNAR_MTLS_GEO		2010	1	0.6						
LUNAR_MTLS_LEO		2015	1	0.5						
ORBITAL_FACTORY		2018	1	0.7						
SOLAR_POWER_SAT		2020	1	0.6						

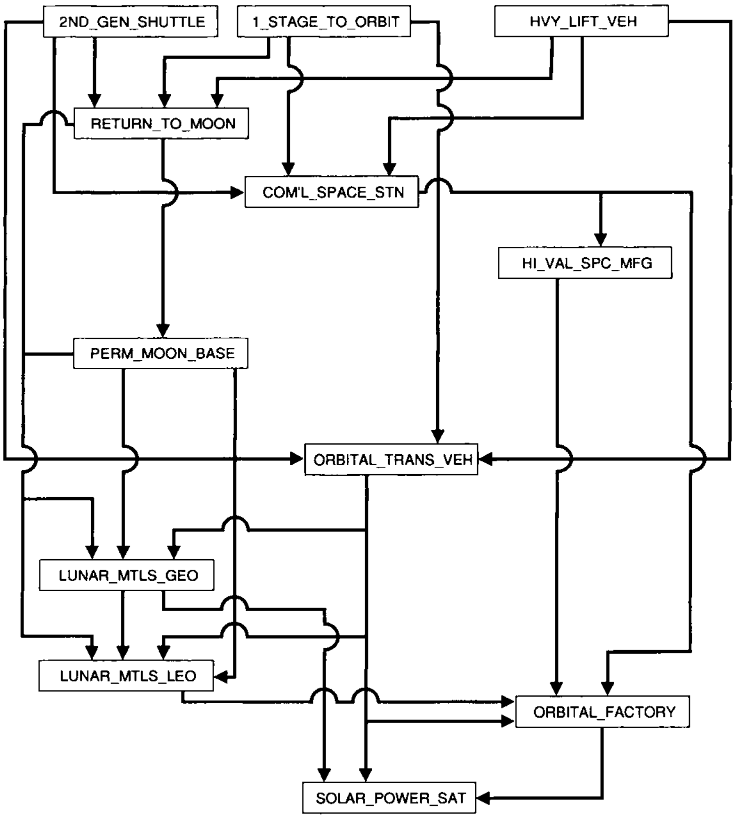

Figure 11.4 Dependency diagram for space exploration events.

ability, or both. Since in this case all the prior events are enabling events, failure to occur must delay the impacted event, reduce its probability, or both.

However, when the prior event is a member of a set, it is known that at most only one member of the set will occur. Hence, the forecasts of events dependent upon the events in the set must be made on the assumption that the set events *will not* occur. The cross impacts, then, will all be expressed in terms of the impact from occurrence. Since in this example all the set members are enabling events, occurrence will advance the date, increase the probability, or both of the impacted events.

Note that in all the cases in the table the earlier event represents an enabling or favorable event for the later ones. In other models, of course, nonoccurrence of one event may hasten another or make it more likely. Conversely, occurrence of one event may delay another event or make it less likely. The forecaster has complete freedom to specify any logical relationship between pairs of events in the model.

The cross impacts for this model are shown in Table 11.2. This table is in the MAXIM cross-impact data worksheet provided on the program disk. This

TABLE 11.2 Cross-Impacts for Space Model

		Occurring		Not occurring	
Impacting event	Impacted event	New T	New P	New T	New P
HVY_LIFT_VEH	RETURN_TO_MOON				0.7
HVY_LIFT_VEH	COM'L_SPACE_STN				0.5
HVY_LIFT_VEH	ORBITAL_TRANS_VEH			2015	0.3
2ND_GEN_SHUTTLE	RETURN_TO_MOON		0.9		
2ND_GEN_SHUTTLE	COM'L_SPACE_STN		0.7		
2ND_GEN_SHUTTLE	ORBITAL_TRANS_VEH	2008	0.7		
1_STAGE_TO_ORBIT	RETURN_TO_MOON		0.9		
1_STAGE_TO_ORBIT	COM'L_SPACE_STN		0.8		
1_STAGE_TO_ORBIT	ORBITAL_TRANS_VEH	2008	0.7		
RETURN_TO_MOON	PERM_MOON_BASE			−1	−1
RETURN_TO_MOON	LUNAR_MTLS_GEO			−1	−1
RETURN_TO_MOON	LUNAR_MTLS_LEO			−1	−1
PERM_MOON_BASE	LUNAR_MTLS_GEO			2020	0.3
PERM_MOON_BASE	LUNAR_MTLS_LEO			2025	0.2
PERM_MOON_BASE	SOLAR_POWER_SAT				0.2
HI_VAL_SPC_MFG	ORBITAL_FACTORY			2025	0.3
ORBITAL_TRANS_VEH	LUNAR_MTLS_GEO				0.4
ORBITAL_TRANS_VEH	LUNAR_MTLS_LEO				0.3
ORBITAL_TRANS_VEH	SOLAR_POWER_SAT				0.4
ORBITAL_TRANS_VEH	ORBITAL_FACTORY				0.6
LUNAR_MTLS_GEO	LUNAR_MTLS_LEO			2025	0.1
LUNAR_MTLS_GEO	SOLAR_POWER_SAT			2025	0.3
LUNAR_MTLS_LEO	ORBITAL_FACTORY			2020	0.5
ORBITAL_FACTORY	SOLAR_POWER_SAT			2025	0.4
COM'L_SPACE_STN	HI_VAL_SPC_MFG			2008	0.4

table lists the "impacting" event and the "impacted" event, and has columns for the new year (New T) and the new probability (New P) if the impacting event occurs. Similarly, there are columns for the new year and new probability if the impacting event does not occur. No entry means that no change will occur. A brief explanation of the logic behind the impacts in the table is given below.

The heavy lift vehicle would make return to the moon easier. However, it is not essential, since the original moon trips were made with a specialized vehicle. Hence, failure of this advanced vehicle reduces the probability of return to the moon but does not eliminate it completely. In the same way, the second generation shuttle or the single stage to orbit vehicle would make return to the moon easier, but they are not essential. As explained above, anything dependent on these two mutually exclusive events must be based on the assumption that they will not occur. Hence, the occurrence of either will advance the time or increase the probability of the impacted event. Note that the single stage to orbit vehicle may not occur in time to have any impact on the return to the moon. Only if this latter event is delayed past the deployment of the single stage to orbit vehicle could there be an impact. The impact is shown, but barring a delay of return to the moon, that event will be decided before the single stage to orbit event can occur.

The heavy lift vehicle, the second generation shuttle, and the single stage to orbit vehicle would make it easier to deploy the orbital transfer vehicle. Failure to obtain the first of these would delay the appearance of the vehicle and reduce the likelihood of deploying it. Occurrence of either of the other two would advance the appearance of the vehicle and enhance its probability.

Failure to return to the moon completely eliminates the moon base and use of lunar materials. The probability of these latter events is reduced to zero.

Establishment of a permanent moon base makes it easier to obtain lunar materials for use in GEO, LEO, and for power satellite construction. However, it would be possible to make individual expeditions to obtain lunar materials. Hence, failure to establish a moon base hinders but does not eliminate the events dependent upon it.

A commercial space station makes it easier to undertake manufacturing in space. However, even in its absence, manufacturing would be possible. It would require batch processing on individual flights, however, which would be more expensive. Hence, failure to establish a commercial space station delays space manufacturing and reduces its likelihood.

The use of lunar materials would be cheaper than the use of materials brought up from Earth for space manufacturing and for construction of a solar power satellite. However, these latter events could still take place in the absence of lunar materials. Hence, failure to obtain lunar materials delays these events and reduces the probability of the events but does not eliminate them entirely.

Availability of an orbital transfer vehicle would reduce the cost of space manufacturing, of using lunar materials, and of constructing solar power satellites. However, these activities could still be undertaken, at higher cost, even if the orbital transfer vehicle were not available. Hence, failure to deploy this vehicle reduces the probability of the events dependent upon it, but does not eliminate them.

The commercial space station is not essential to deploying solar power satellites. However, its availability and its existence as a precedent would enhance the likelihood of the satellites. Moreover, its existence might provide a market for power obtained from the satellites. Hence, failure to obtain the commercial space station delays the appearance of the satellites and reduces their likelihood.

Although the dependencies among the events are shown in Fig. 11.4, they can also be displayed usefully in tabular form. Table 11.3 shows those events that are impacted by specific events. Table 11.4 shows those events that impacted upon specific events. These two tables are actually outputs from MAXIM and amount to a sorting of the input data.

From Table 11.3, we can see that returning to the moon and the orbital transfer vehicle are critical events since many other events are dependent upon them. The heavy lift vehicle, the second generation shuttle, and the single stage to orbit vehicle all impact upon the same events. Any one of the three events would make further development easier, hence, individually they are not as critical as the two mentioned above.

TABLE 11.3 Impacts from Specific Events

Impacting event	Impacted events
HVY_LIFT_VEH	RETURN_TO_MOON COM'L_SPACE_STN ORBITAL_TRANS_VEH
2ND_GEN_SHUTTLE	RETURN_TO_MOON COM'L_SPACE_STN ORBITAL_TRANS_VEH
RETURN_TO_MOON	PERM_MOON_BASE LUNAR_MTLS_GEO LUNAR_MTLS_LEO
COM'L_SPACE_STN	HI_VAL_SPC_MFG
HI_VAL_SPC_MFG	ORBITAL_FACTORY
1_STAGE_TO_ORBIT	RETURN_TO_MOON COM'L_SPACE_STN ORBITAL_TRANS_VEH
PERM_MOON_BASE	LUNAR_MTLS_GEO LUNAR_MTLS_LEO SOLAR_POWER_SAT
ORBITAL_TRANS_VEH	LUNAR_MTLS_GEO LUNAR_MTLS_LEO SOLAR_POWER_SAT ORBITAL_FACTORY
LUNAR_MTLS_GEO	LUNAR_MTLS_LEO SOLAR_POWER_SAT
LUNAR_MTLS_LEO	ORBITAL_FACTORY
ORBITAL_FACTORY	SOLAR_POWER_SAT
SOLAR_POWER_SAT	NONE

From Table 11.4, we can see that the later events are the ones most dependent upon earlier events. In particular, the use of lunar materials and the establishment of the solar power satellite are strongly affected by earlier events.

Given these events and cross impacts, how is the model run? We start with the event that is first in chronological order, the heavy lift vehicle. It is forecast to have probability of 0.6 of ever occurring. Hence, we test the occurrence of this event by drawing a random number in the range of 0.00 to 0.99. If the number that is drawn is strictly less than 0.6, the heavy lift vehicle is taken to occur. If the number drawn is equal to or greater than 0.6, the event does not occur. The rule "strictly less than" is required for occurrence because the value 0.00 counts as "less than"; the value 1.00 cannot occur. Depending upon the outcome of the test (occurrence or nonoccurrence), we adjust the timing and probability of all those events impacted by the first one. We then test the event that is second in chronological order in the same way. Note that the chronological order of the events may change as a result of impacts from an event whose outcome is determined by the random test. This process is con-

TABLE 11.4 Impacts on Specific Events

Impacted event	Impacting events
HVY_LIFT_VEH	NONE
2ND_GEN_SHUTTLE	NONE
RETURN_TO_MOON	HVY_LIFT_VEH 2ND_GEN_SHUTTLE 1_STAGE_TO_ORBIT
COM'L_SPACE_STN	HVY_LIFT_VEH 2ND_GEN_SHUTTLE 1_STAGE_TO_ORBIT
HI_VAL_SPC_MFG	COM'L_SPACE_STN
1_STAGE_TO_ORBIT	NONE
PERM_MOON_BASE	RETURN_TO_MOON
ORBITAL_TRANS_VEH	HVY_LIFT_VEH 2ND_GEN_SHUTTLE 1_STAGE_TO_ORBIT
LUNAR_MTLS_GEO	RETURN_TO_MOON PERM_MOON_BASE ORBITAL_TRANS_VEH
LUNAR_MTLS_LEO	RETURN_TO_MOON PERM_MOON_BASE ORBITAL_TRANS_VEH LUNAR_MTLS_GEO
ORBITAL_FACTORY	HI_VAL_SPC_MFG ORBITAL_TRANS_VEH LUNAR_MTLS_LEO
SOLAR_POWER_SAT	PERM_MOON_BASE ORBITAL_TRANS_VEH LUNAR_MTLS_GEO ORBITAL_FACTORY

tinued until the outcome for all the events has been determined. This completes one simulation run.

The result of a single simulation run of the model, using MAXIM, is shown in Table 11.5. Some events occurred, while others did not. In particular, the heavy lift vehicle and the single stage to orbit vehicle were both developed. In addition, the commercial space station and the orbital transfer vehicle were both deployed. However, no return was made to the moon. With low-cost materials from the moon not available, orbital manufacturing and the solar power satellite were not deployed.

However, getting only one simulation run does not tell us much about the future. It produces one sequence of possible outcomes (in effect, a single scenario). The strength of the cross-impact model is its ability to obtain the distribution of future events over time, as they are advanced or delayed by the occurrence or nonoccurrence of previous events. This distribution is obtained

TABLE 11.5 Results of Single Run of Space Model

Event	Year of occurrence
HVY_LIFT_VEH	1998
2ND_GEN_SHUTTLE	N/O*
RETURN_TO_MOON	N/O
COM'L_SPACE_STN	2001
HI_VAL_SPC_MFG	2003
1_STAGE_TO_ORBIT	2004
PERM_MOON_BASE	N/O
ORBITAL_TRANS_VEH	2008
LUNAR_MTLS_GEO	N/O
LUNAR_MTLS_LEO	N/O
ORBITAL_FACTORY	N/O
SOLAR_POWER_SAT	N/O

*N/O, not occurring.

by making repeated simulation runs and observing the variation in outcomes. Of course, after each run the probabilities and times are reset to their initial values, so that all runs start from the same set of events.

Table 11.6 shows the results of 50 runs of this model, using MAXIM. (This is a reformatting of the actual MAXIM output, to fit the page more conveniently.) For each event, the table gives the year(s) in which it occurred during the simulation runs, the frequency of occurrence in those years (fraction of the total number of runs in which the event occurred), predicted frequency of occurrence, and the difference between the two. A positive difference means the event occurred more frequently than forecast; a negative difference means it occurred less frequently than forecast. Note that for events involving a set or a span of time, the predicted frequency for a given year is the product of the probability of ever occurring and the conditional probability for that year.

For instance, the heavy lift vehicle is shown as having occurred in 1998 in 54 percent of the runs (27 out of 50), while it was forecast to occur in 60 percent of the runs (30 out of 50). The difference is a negative 6 percent. This difference is due solely to random fluctuations. In another set of 50 runs, the frequency of occurrence might be higher than the forecast, rather than lower. In a larger number of runs, one would expect the percentage fluctuation to be smaller, although the actual difference might be even greater than that obtained here (3 out of 50). The orbital transfer vehicle occurred in 2008 in 32 percent of the runs, in 2010 in 8 percent of the runs, in 2011 in 8 percent of the runs, and in 2012 in 10 percent of the runs. The delays were caused by a failure to occur of various combinations of events that affect the appearance of the vehicle. Note that in total, the orbital transfer vehicle occurred in 58 percent of the runs, whereas it was forecast to occur in 50 percent of the runs.

TABLE 11.6 Result of 50 Simulation Runs

		Frequency		
Event	Years	Actual	Predicted	Difference
HVY _ LIFT _ VEH	1998	0.54	0.60	0.04
2ND _ GEN _ SHUTTLE	1999	0.34	0.28	0.06
RETURN _ TO _ MOON	2000	0.86	0.85	0.01
COM'L _ SPACE _ STN	2000	0.18	0.18	0.00
	2001	0.16	0.12	0.04
	2002	0.08	0.12	- 0.04
	2003	0.26	0.12	0.14
	2004	0.04	0.06	- 0.02
Total		0.64	0.60	0.04
HI _ VAL _ SPC _ MFG	2003	0.64	0.80	- 0.16
	2008	0.06	0.00	0.06
Total		0.70	0.80	- 0.10
1 _ STAGE _ TO _ ORBIT	2004	0.44	0.42	0.02
PERM _ MOON _ BASE	2005	0.52	0.60	- 0.08
ORBITAL _ TRANS _ VEH	2008	0.32	0.00	0.32
	2010	0.08	0.50	- 0.42
	2011	0.08	0.00	0.08
	2012	0.10	0.00	0.10
Total		0.58	0.50	0.08
LUNAR _ MTLS _ GEO	2010	0.30	0.60	- 0.30
	2020	0.04	0.00	0.04
Total		0.34	0.60	- 0.26
LUNAR _ MTLS _ LEO	2015	0.06	0.50	- 0.44
	2030	0.02	0.00	0.02
Total		0.08	0.50	- 0.42
ORBITAL _ FACTORY	2018	0.28	0.70	- 0.42
	2020	0.14	0.00	0.14
	2025	0.04	0.00	0.04
Total		0.46	0.70	- 0.24
SOLAR _ POWER _ SAT	2020	0.18	0.60	- 0.42
	2028	0.02	0.00	0.02
Total		0.20	0.60	- 0.40

This difference is due not only to statistical fluctuation but to the impacts from earlier events that altered the timing and probability for the vehicle. In general, if the difference between the forecast and actual frequencies of occurrence exceeds what might be explained by statistical fluctuation, there is some connection with other events through the cross impacts specified in the model.

Note that, in general, the actual frequency of occurrence is different from the forecast frequency of occurrence, and for the later events the differences are large negative values. This is because the forecast probabilities were based on the assumption that the prior enabling events had occurred. When these did not occur, the later events were either reduced in probability or were canceled. What we want in a forecast, of course, is these actual frequencies of occurrence. However, the forecaster should not attempt to estimate those frequencies directly, which amounts to trying to solve the model in your head. Instead, the forecaster is better advised to estimate the conditional probabilities for each event (conditional on specified outcomes for prior impacting events) and use a computer program such as MAXIM to simulate the behavior of the system.

The key point to be observed here is that the cross-impact model allows the forecaster to combine a set of individual forecasts when these are related or interconnected. The final result is to show the composite result of the entire set of forecasts. Individual events are shown in the context of all the other events to which they are related. The output of a cross-impact model is a frequency distribution of events over time, based on their forecast timing and probability and modified by other events that have an impact on them. The cross-impact model provides the forecaster with a convenient tool for taking into account the interactions among probabilistic forecasts of individual events.

11.7 Model Construction

Both scenarios and cross-impact models have multiple "output" variables. This is to be expected, since both involve the combination of separate and distinct forecasts. In this they differ from such combinations as growth curves and trends, and the combinations appearing in top-down and bottom-up forecasts. In developing such a multiple-output model, the forecaster must proceed backward, just as is done with models such as KSIM. The starting point for either a scenario or a cross-impact model is the nature of the desired information. This determines the events that go into either the scenario or cross-impact model. The forecaster should ask the following questions, in the order given:

1. What decision needs to be made? Specifically, what change in resource allocation or organizational structure is being considered?

2. What information is needed to make that decision? Specifically, what does the decision maker need to know in order to make the best choice? How is "best" determined?

3. What "marker events" provide this information? These events might be the completion of some project, the achievement of some goal, the reversal of a trend, or the crossing of some threshold by a trend. The determination of the marker events depends upon the nature of the decision. These will be the events which affect the decision directly, without any other intervening events. In most cases, only part of the decision information comes from a tech-

nological forecast; hence, the marker events in the scenario or cross-impact model need to provide only that portion of the total-decision information.

4. What events trigger or influence the marker events? What events in turn influence them? The chain of events must be traced backward, from the marker events to all the other events that affect or influence them, directly or indirectly.

This sequence of questions produces a list of events that must be included in the scenario or the cross-impact model. Constructing the scenario or model is then a matter of determining the linkages among the events, and expressing them in the appropriate form (e.g., changes in timing and probability for a cross-impact model). The scenario or cross-impact model will involve a network of events, i.e., the various chains of events from the starting point to the end point will be cross-linked. This complexity is, of course, the reason that a scenario or a cross-impact model was selected as the forecasting tool.

The critical issue is that the starting point for construction of a scenario or a cross-impact model is the decision to be made and the information needed to make that decision. The scenario or the cross-impact model must be designed to provide that information, if they are to be useful.

11.8 Summary

In many cases the forecaster has alternate ways of forecasting the same event, or has several forecasts of related events. A forecast can often be made more useful by combining several forecasts. The use of trend and growth curves together, and the use of analogies with trend and growth curves can improve the forecast of a single technology. Where an aggregate and its components can be forecast separately, more information can often be gained by combining the separate forecasts. When a large number of related forecasts must be combined into a composite whole, either scenarios or cross-impact models may be used. In all the cases, the forecaster should consider the possibility of combining two or more forecasts to offset the weaknesses of some with the strengths of others.

References

Armstrong, J. S., *Long-Range Forecasting,* John Wiley & Sons, New York, NY, 1985.
Kahn, H. and A. Wiener, *The Year 2000,* Macmillan, New York, NY, 1967.

Problems

1 What are the advantages and disadvantages of combining separate forecasts?

2 Use the history of U.S. electrical power production from Table A.1. Plot the data on semilog paper. Is there a trend in some or all of the data? Are there any deviations from what you see as a trend? Fit a trend to whatever subset of the data you consider appropriate. Explain any major departures from the trend. How could you use this as an analogy to forecast future departures from a trend?

3 Using MAXIM, make 50 simulation runs of the space model in the chapter. What is the frequency of occurrence for each event in your simulations? Do they differ significantly from the results given in the chapter? Are there any events for which the frequency of occurrence differs significantly from the original forecast? Select what you consider to be a critical event and modify its probability of occurrence. Make another 50 simulation runs. Is there a big difference in the outcome?

4 Take one possible sequence of outcomes from the cross-impact model in the chapter and write a scenario describing it. What additional information do you need to provide the background? What actors should be involved?

5 Refer to Table A.28, the U.S. total gross consumption of energy resources. Fit an exponential curve to the historical consumption of each energy source listed (anthracite, bituminous, etc.), and project each one to 1980 (ignore the projections made by the Bureau of Mines in the last three rows). Does the total of your projections for the components add up to your projection of the total? Should the projection for the total or the projection of the components be adjusted? If the latter, which components? What is the shift in the percentage of the total for each component between 1965 and 1980?

6 Expand the nuclear energy scenario given in the chapter. Add whatever background trends, technology events, trigger events, etc., you deem necessary or appropriate. Select one of the approaches listed in the chapter. Write the scenario.

7 Revise the scenario given in the chapter so that the critical environmental problem is global cooling ("new ice age") instead of global warming ("greenhouse effect"). Add whatever background trends, technology events, trigger events, etc., you deem necessary or appropriate. Select one of the approaches listed in the chapter and write the scenario.

Chapter

12

Normative Methods

12.1 Introduction

The forecasting methods discussed in the previous chapters are "exploratory" methods, i.e., they all start with past and present conditions and attempt to project these to estimate future conditions. They explore the possible futures implicit in past and present conditions.

A complementary approach is taken by "normative" methods. These methods have their foundations in the methods of systems analysis. They start with future needs and identify the technological performance required to meet those needs. In essence, they forecast the capabilities that will be available on the assumption that needs will be met.

Normative and exploratory methods are customarily used together. An exploratory forecast has implicit within it the idea that the capability will be desired when it becomes available. A normative forecast has implicit within it the idea that the required performance can be achieved by a reasonable extension of past technological progress. The complementary use of normative and exploratory forecasts will be taken up in later chapters on applications. In this chapter, we discuss the normative methods.

Three common methods of normative forecasting will be described. These are relevance trees, morphological models, and mission flow diagrams.

12.2 Relevance Trees

Relevance trees are used to analyze situations in which distinct levels of complexity or hierarchy can be identified. Each successively lower level involves finer distinctions or subdivisions.

Figure 12.1 shows a relevance tree. At the top of the tree is an automobile. (The "head end" of the tree will hereinafter be referred to as "level zero." Level one is the first full level below the head end.) At level one, we have three elements of the automobile. One of these elements is further subdivided at the second and lower levels. The other two elements could be subdivided in the same way.

Each item on the tree is referred to as a "branch." The point from which several branches "depend" is a "node." Thus, except for those branches at the top and bottom of the tree, each branch depends from a node and each has a node from which several other branches depend. There is no requirement that each node have the same number of branches.

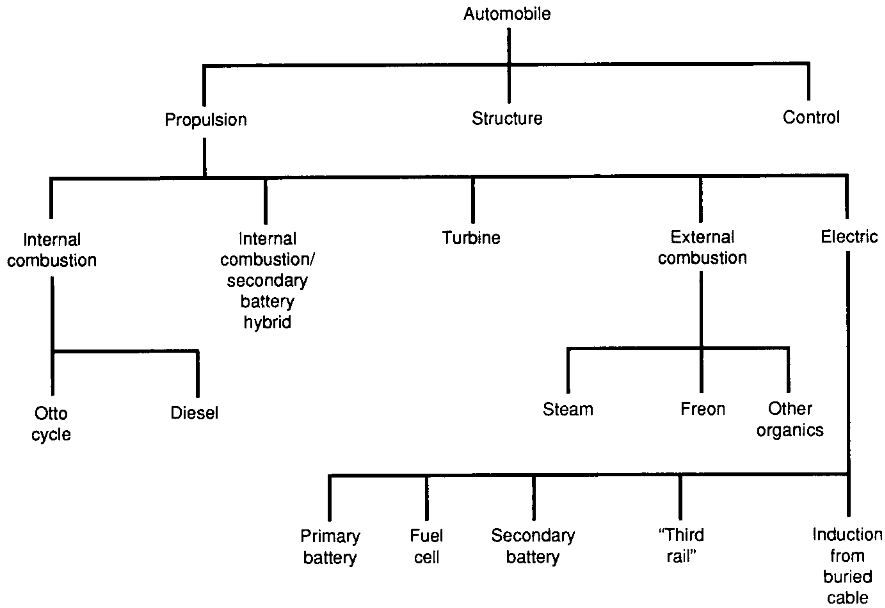

Figure 12.1 Relevance tree showing the major components of an automobile.

Figure 12.1 clearly shows the hierarchical structure of a relevance tree. A relevance tree also has several other characteristics that are not obvious from Fig. 12.1, but which must be stated. First, the branches depending from a node must be a closed set. They must be an exhaustive listing of all the possibilities at that node. In some cases, this means simply listing all the members of a finite set. In other cases, however, the set is closed by the agreement that while there may be other items, they are not relevant or important by comparison with those listed. Second, the branches depending from a node must be mutually exclusive; they must have no overlap among them. Third, for a normative relevance tree, the branches must be viewed as goals and subgoals. Each node is a goal for all branches depending from it; each goal is satisfied by the satisfaction of all the nodes below it, and, in turn, each node derives its validity as a goal from the sequence of branches linking it to the top of the tree.

The branches of relevance trees used for normative forecasting will be either problems or solutions. Thus, two extreme or "pure" types of trees are the solution tree and the problem tree.

The solution tree has both "and" and "or" nodes. The or nodes are the most common. At each node, the tree has two or more alternate solutions or answers to the question of how to achieve the solution at the next higher level; achievement of any single solution is sufficient. In some cases, there will be and nodes; each of two or more partial solutions must be implemented if the higher-level solution is to be achieved.

Figure 12.1 is a solution tree. The branches at each node represent more refined or detailed means of achieving the solution at the next higher level. At each node, there are two or more alternative solutions.

A problem tree has only and nodes. At each node, there are several problems that must be solved if the problem at the next higher level is to be solved. Figure 12.2 shows a problem tree. One segment of the relevance tree for an automobile is isolated here. At the top of the tree is the problem of electrical propulsion. At the first level there are three problems, each of which must be solved if electrical propulsion is to be achieved. At the second level there are several more problems, each of which must be solved if the problems at level one are to be solved.

Normally the technological forecaster will use mixed trees, that is, trees containing both problems and solutions. The exact nature of the tree developed for a specific situation will depend upon the purposes it is to serve. There is no such thing as a universal or "all-purpose" relevance tree, even for a particular technology. The forecaster has considerable flexibility in designing or shaping a relevance tree. This flexibility should be utilized to assure that the tree satisfies the purpose for which it was prepared.

It is possible for a relevance tree to be internally inconsistent, in that branches in one portion of a tree are incompatible with branches in another portion. For instance, a branch included as a solution to one problem may be incompatible with a branch included as a solution to some other problem. The same may occur with problems. A problem in one portion of the tree may exist only in a particular environment or set of circumstances. The environment or circumstances may rule out a problem found in some other portion of the tree,

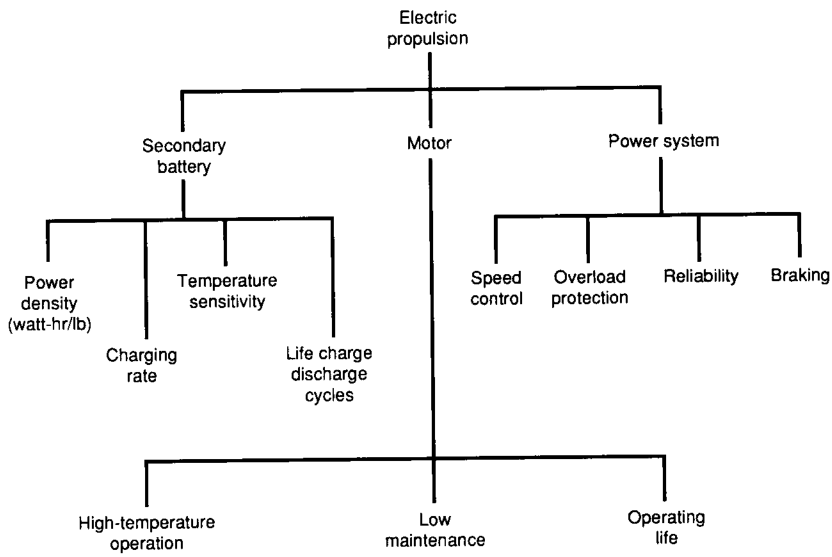

Figure 12.2 Problem tree for an electric automobile.

a problem which can exist only in another environment or under some other circumstances. This incompatibility is not really a matter for concern; it simply means that some combinations of solutions are impossible (for incompatible solutions) or unnecessary (for incompatible problems). For instance, we do not have to solve simultaneously the problems of operating a machine in a sandy desert and in a muddy swamp. After a tree is developed, the forecaster may need to identify incompatibilities before using it to draw conclusions. However, the existence of incompatibilities does not negate the utility of a tree.

Relevance trees can be used to identify problems and solutions and deduce the performance requirements of specific technologies. However, they can also be used to determine the relative importance of efforts to increase technological performance. Suppose an objective is to be achieved, and suppose that there are three tasks that must be carried out to achieve this objective. We can then determine how much technological progress is needed in order to carry out each task. Assume that the first task needs twice as much progress as the second task, which in turn needs three times as much progress as the third task. We can then assign these tasks numerical weights or "relevance numbers" of 0.6, 0.3, and 0.1. The relevance numbers must reflect the relative importance assigned to each of the tasks. In addition, the relevance numbers at a node must sum to 1.0. This requirement is referred to as "normalization" of the relevance numbers. It is an essential feature of the application of relevance numbers to relevance trees. Figure 12.3 shows a relevance tree with three tasks given the weights we have just assigned. Each of the tasks, in turn, may be achieved by two or more approaches, as shown. Relevance numbers have been assigned to each approach as well. The relevance numbers assigned to each task reflect the degree of improvement needed in the approach if the task is to be accomplished. Only the approaches for a particular node (task) are compared with one another. As stated before, the assigned relevance numbers must sum to 1.0 at each node.

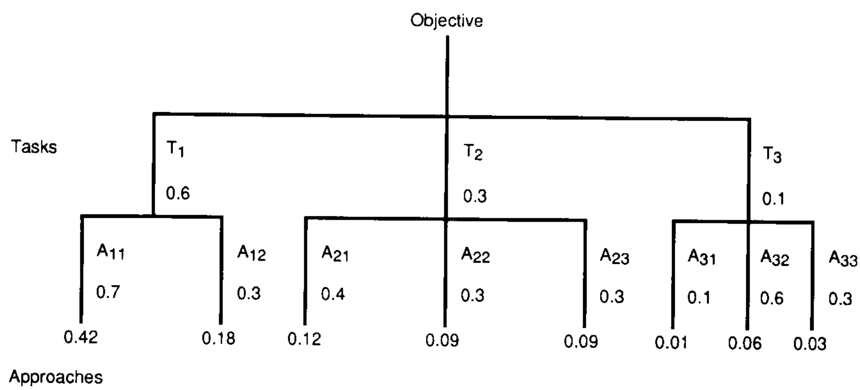

Figure 12.3 Relevance tree with relevance numbers.

We can now determine the relative importance of the approaches with regard to the overall objective, not just with regard to the tasks to which they apply. For instance, A_{11} has a relevance of 0.7 to task T_1, which in turn has a relevance of 0.6 to the overall objective. Therefore, the relevance of approach A_{11} to the overall objective is $0.6 \times 0.7 = 0.42$. The relevance of each of the approaches is shown at the bottom line of Fig. 12.3. The relative amount of effort to be allocated to each approach can now be determined from the relevance numbers. The advantage of this technique is that approaches can be compared with one another even though they do not apply to the same task, since their relative importance to the overall objective can be determined.

A word of caution is in order. The relevance numbers assigned to individual branches in this example are precise to only one significant figure. When multiplied together, however, they produce numbers of two or more significant figures. This represents a spurious precision. For instance, it is reasonable to say that A_{11} (with a relevance number of 0.42) is about twice as important as A_{12} (with a relevance number of 0.18). To argue that A_{12} is precisely 50 percent more important than A_{21} is reading more into the numbers than is really there.

This use of relevance numbers was originally given the name "PATTERN" (Alderson and Sproull, 1972). PATTERN was developed as a means of allocating *research and development* (R&D) resources to the problems most needing solution.

Most relevance trees do not use relevance numbers. They are utilized to derive performance requirements for individual technologies in order to achieve some overall objective. Relevance trees can be extremely useful for this normative purpose. Their utility can be enhanced, however, by incorporating relevance numbers where it is appropriate. The use of relevance numbers increases the power of the relevance tree as a tool for identifying needed changes in technology.

12.3 Morphological Models

The word "morphology" is used in many sciences to describe the study of the form or structure of something (e.g., plants in biology and rocks in geology). The morphological model was invented by Fritz Zwicky (1957 and 1969), who used it to analyze the structure of problems. The morphological model is a scheme for breaking a problem down into parallel parts, as distinguished from the hierarchical breakdown of the relevance tree. The parts are then treated independently. If one or more solutions can be found for each part, then the total number of solutions to the entire problem is the product of the number of solutions to individual parts. For instance, if a problem can be decomposed into 4 independent parts and there are 2 solutions to each, then there are a total of 16 different solutions for the whole problem. Each solution for the whole problem is a combination of solutions for the individual parts. Some of these solutions will, of course, be impossible if a solution to one part is incompatible with a solution to another.

To make this more concrete, Fig. 12.4 shows a morphological model of automobile propulsion, which can be contrasted with the relevance tree of Fig. 12.1. The model has six elements: the number of wheels, the number of driven wheels, the number of engines and/or motors, the type of transmission, the engine type, and the power source. Each of these elements has two or more alternative components, as shown. This morphological model provides a total of $2 \times 4 \times 4 \times 3 \times 4 \times 5 = 1920$ distinct solutions to the problem of automobile propulsion. Some of the solutions are not really possible. For instance, the number of driven wheels obviously cannot exceed the number of wheels; some of the power sources shown are compatible with only one of the engine types shown; etc. Once the impossible solutions are deleted, however, there is still a significant number of different solutions to the problem of automobile propulsion.

Once the impossible solutions are eliminated from consideration, the morphological model can be used to derive the performance requirements for each of the elements of the remaining solutions. These then become the normative forecasts of the performance if the overall function is to be performed. In addition, the derived performance requirements can be used to estimate when the solution might be feasible if exploratory forecasts of the performance of each possible solution are prepared. Finally, if the required performance levels appear to be attainable by the time they are required, they can be used as objectives for an R&D program.

In general, morphological models and relevance trees are suited to different types of problems. However, some problems can be treated by either method. Forecasters will ordinarily select the method that is easiest to tailor to their particular needs. A comparison of the application of the two different methods on the same problem is worthwhile. Comparing Figs. 12.1 and 12.4, we see similarities and differences. Some items appear in both models of automobile propulsion. The various types of nonelectric engines, which show up as branches in the relevance tree, appear as alternate components under the "engine type" element of the morphological model. The sources of electrical energy that appear as branches in the relevance tree are also found as components under the "power source" element of the morphological model. This is

P_1	Wheels	3	4		
P_2	Driven wheels	1	2	3	4
P_3	Engines	1	2	3	4
P_4	Transmission	None	Mechanical	Fluid	
P_5	Engine type	Internal combustion	External combustion	Turbine	Electric
P_6	Power source	Hydrocarbon fuel	Primary battery	Secondary battery	Third rail

Figure 12.4 Morphological model of automobile propulsion.

actually a fairly general result. If the same system is modeled by a relevance tree and by a morphological model, the elements of the morphological model will correspond to major connected sections of the relevance tree, and the branches appearing at the bottom level of such sections in the relevance tree will be the components of the corresponding elements of the morphological model.

When a problem is obviously suited to one or the other approach, the forecaster has little difficulty in choosing between them. However, in those cases where either could be applied, the one that best suits the forecaster's purposes must be selected. This may require a careful study of the situation. There is no hard and fast rule that forces the selection of one over the other.

12.4 Mission Flow Diagrams

The mission flow diagram was originally devised by Harold Linstone (1968) as a means of analyzing military missions, a fact which accounts for the name of the method. However, it can be used to analyze any sequential process. It involves mapping all the alternative routes or sequences by which some task can be accomplished. All the significant steps on each route must be identified. The analyst can then determine the difficulties and costs associated with each route. In addition, it is possible to create new routes and identify the difficulties and costs associated with these. Once these difficulties and costs are identified, the performance requirements can be derived for the technologies involved and then used as normative forecasts.

Figure 12.5 shows a mission flow diagram for sending a message. The figure includes one path for which there is no physical movement at all, which illustrates that a mission flow diagram is not limited to cases involving physical movement. Figure 12.5 also includes cases of two paths merging into one and

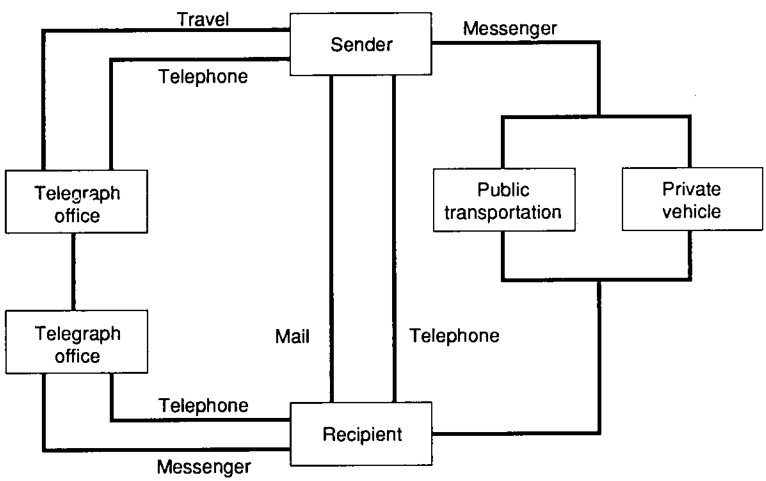

Figure 12.5 Mission flow diagram for sending a message.

of one path branching into two. This illustrates that different paths may have common segments and that these may occur at the beginning, middle, or end of the paths.

Like the relevance tree and the morphological model, the mission flow diagram is of little value if it is restricted to simply describing some situation. The merit of the mission flow diagram is in allowing the user to identify the problems and to derive the performance required of some technology to overcome the problems. Like the other two normative methods, the mission flow diagram permits the user to include new technologies and alternatives that do not yet exist.

Numerical weights can be placed on the alternative paths of a mission flow diagram, similar to the use of relevance numbers in relevance trees. These numbers could then be used to determine the relative importance of the technological solutions to each of the bottlenecks or difficulties on the paths. However, little work has been done in carrying out this approach. The most common use of mission flow diagrams is to identify difficulties and to compare existing paths with alternatives that might be developed if desired.

12.5 Goals

Throughout the preceding chapters, we have stressed the idea that forecasts are not prepared for their own sake. They are prepared as inputs to decisions. The purpose of a forecast is to improve the quality of the decision that incorporates it.

This point is particularly relevant with regard to normative forecasting methods. These are goal-oriented methods. They cannot even be used unless some goal has been specified, a goal that the technology being forecast is to help accomplish.

Thus, in developing a normative forecast, the first issue a forecaster must address is, "What question am I trying to answer?" Elements of this issue include: Who "owns" the question I'm trying to answer? What is that person trying to accomplish? Why is that person trying to accomplish it? What does that person need to know in order to make a good decision? What part of that decision information is supposed to come from my forecast?

Only when these questions are answered is it possible to prepare a forecast that will be useful and effective in improving the quality of a decision. Hence, before undertaking a normative forecast, the forecaster must make sure these questions are answered. Then he or she must assure that the forecast includes the information implied by the answers.

12.6 Summary

In this chapter, we have discussed normative, or goal-setting, methods of forecasting. They are used to determine the level of functional capability that must be achieved to solve a problem or overcome a difficulty. Relevance trees are used to carry out analyses of hierarchical structures; morphological models are used to analyze parallel structures; and mission flow diagrams are used to analyze processes with steps in a spatial, temporal, or logical sequence. Often the same system can be analyzed by either a relevance tree or a morphological model. In such cases, the forecaster should use whichever method is most convenient or appropriate for the problem at hand.

The strength of normative models is that they tend to organize and structure the problem. They can help assure completeness, so that solutions holding some promise are not overlooked. By systematically laying out the structure of a problem, they can assist in the generation of new alternatives that may be superior to those currently in use. Even if all the alternatives uncovered with the use of normative methods prove inferior to those already known, this fact alone provides additional confidence in the choice of the superior method.

Normative methods also have a disadvantage, which they share with all systematic methods of problem solving. One is that these methods may tend to impose rigidity on the solutions proposed. The problem itself may be distorted when it is fitted to a normative structure. Even if this does not happen, there may be a tendency to look only at the solutions that can be expressed easily within the formal structure of the model. Another difficulty arises when numerical weights are applied to the branches or elements of the model, as with PATTERN. The problem lies in the fact that numbers tend to create their own validity. No matter how much subjectivity, and perhaps even sheer guesswork, went into the assignment of the numbers, they become very impressive once they have been written down. They can become even more impressive if they have been subjected to considerable manipulation in a computer. It often becomes very difficult to challenge the numbers under these circumstances, as though the computerization had somehow overcome any deficiencies in their origin. These disadvantages, however, are not sufficient to rule out the use of normative methods. With adequate care, the disadvantages can be overcome completely, so that normative methods can, indeed, be very valuable. It is imperative, however, that the forecaster be aware of these difficulties.

Since one of the claimed advantages of normative methods is their completeness, it might be asked whether there is any guarantee that a normative model is complete. The answer, of course, is "No." There is no formula or procedure that will guarantee that the constructor of a relevance tree or morphological model has not omitted something. Is this, then, not a weakness? Of course it is if the forecaster uses a normative model in such a way as to be vulnerable to surprise. For instance, the forecaster might assert that there is only one way in which a problem can be solved, since all the other paths on the mission flow diagram are either infeasible or inferior to the path chosen. In such a situation, the existence of an alternative not shown in the model can completely alter the forecaster's conclusions. If, however, the forecaster says only that there is *at least* one feasible solution (or however many are found) and that it places certain requirements on specific technologies, any alternatives that were missed will not overthrow the entire forecast.

Finally, it should be noted that the systematic and orderly structure of any of the normative models we have discussed makes it easy to determine whether a specific solution has been omitted or not. It is not necessary to search through the set of all the configurations of a morphological model, for instance, to determine whether that set includes a particular, well-defined solution. It is necessary only to check whether each of its components appears in the appropriate elements of the model. Hence, even though there is no way to guarantee the completeness of a normative model, it is always possible to tell whether a specific solution has been missed. In addition, it is also usually possible to alter the structure easily to include solutions previously omitted.

In conclusion, it is essential that the normative methods be kept in proper perspective. They are in no sense a substitute for creativity or imagination. They merely provide a systematic means for examining technological requirements. When properly used, they provide a framework within which creativity and imagination can work effectively. However, it must be remembered that they are only sophisticated ways of making lists of alternatives. To revise the complaint of the author of Ecclesiastes, "of the making of lists there is no end." Nevertheless, even though normative methods are not in themselves a substitute for thought, they can be a very useful tool when applied properly.

References

Alderson, R. C. and W. C. Sproull, "Requirements Analysis: Need Forecasting and Technology Planning Using the Honeywell PATTERN Technique," *Technological Forecasting & Social Change,* 3(2):255–265 (1972).

Linstone, H. A., "The Mirage Studies and Some Implications," in J. P. Martino (ed.), *Long Range Forecasting Methodology,* National Technical Information Service, Springfield, VA, 22–51, 1968.

Zwicky, F., *Morphological Astronomy,* Springer-Verlag, Berlin, 1957.

Zwicky, F., *Discovery, Invention, Research,* Macmillan, New York, 1969.

For Further Reference

Blackman, A. W., "Normex Forecasting of Jet Engine Characteristics," *Technological Forecasting & Social Change,* 2(1):61–76 (1970).

Fischer, M., "Toward a Mathematical Theory of Relevance Trees," *Technological Forecasting & Social Change,* 1(4):381–390 (1970).

Martin, W. T. and J. M. Sharp, "Reverse Factor Analysis: A Modification of Relevance Tree Techniques," *Technological Forecasting & Social Change,* 4(4):355–373 (1973).

Swager, W. L., "Strategic Planning I: The Roles of Technological Forecasting," *Technological Forecasting & Social Change,* 4(1):85–109 (1972).

Swager, W. L., "Strategic Planning II: Policy Options," *Technological Forecasting & Social Change,* 4(2):151–172 (1972).

Swager, W. L., "Strategic Planning III: Objectives and Program Options," *Technological Forecasting & Social Change,* 4(3):283–300 (1973).

Problems

1 What is the difference between normative and exploratory methods of forecasting? In what situations is each appropriate? How do they interact?

2 Construct levels one and two of a relevance tree describing the materials used to fabricate an automobile.

3 Construct levels one and two of a relevance tree describing the habitability requirements of a moon base.

4 Construct levels one, two, and three of a relevance tree of improvements or changes needed in urban public transit. This may be a mixed-type tree if necessary. Place relevance numbers on each branch of the tree. (If this problem is worked on by a class or group, the Delphi procedure may be used to reach a consensus on the relevance numbers. In this case, the panel estimate should be the mean instead of the median of the individual estimates to preserve the property that the relevance numbers must sum to unity at each node.)

5 Develop a morphological model of the possible forms of wrenches.

6 Develop a morphological model of the problem of extracting food from the sea. For one configuration of your model, prepare a normative forecast of the technology involved. When might your configuration be needed? Will it be feasible by then?

7 Develop a mission flow diagram showing alternate paths for making the energy in unmined coal available in the home as electricity. For one path on your diagram that is not currently in use, prepare a normative forecast of the technologies involved. What are the advantages of your chosen path over those currently in use? When might your path be feasible?

8 Devise an alternate path for Fig. 12.5 that has the following properties: The recipient need not be in any specified physical location to receive the message, the message transmission (but not its composition or preparation) must take essentially no time, the method must permit a dialogue between the sender and the recipient, and a written text may be provided if desired. Develop a normative forecast of the technologies required if your alternative path is to be technically and economically feasible.

Chapter

13

Planning and Decision Making

13.1 Introduction

The preceding chapters have dealt with the methods of making technological forecasts. Their emphasis was largely methodological, and while some applications were mentioned, these were secondary to the discussion of the methods.

The viewpoint taken in this book is that technological forecasting is an aid to decision making. Technological forecasts are not produced in a vacuum, nor are they produced for their own sake; if a forecast is not useful for making decisions, it is worthless. This is true regardless of the technical elegance of a forecast or the quality of the data behind it.

Each of the following four chapters will be devoted to a specific application of technological forecasting. The areas covered are those in which technological forecasting has been most widely used, although these are not the only areas in which technological forecasting is possible and desirable. The discussions will tend to focus on the unique problems of each application area. However, there is enough similarity among application areas that much of what is said can be applied, with the proper adjustments, to other areas as well.

In the remainder of this chapter, we discuss the features of planning and decision making that are common to all application areas. It, thus, provides an introduction to the chapters on specific applications areas by placing them in a broader context. It also provides a basis for extending applications of technological forecasting to application areas not specifically covered in the subsequent chapters.

13.2 The Purpose of Planning

One of the best-kept secrets of the planning profession is that planning has nothing to do with actions to be taken in the future. Instead, planning deals with actions to be taken in the present, to achieve a desired future. Put another way, planning does not deal with future actions but with the futurity of present actions.

Nevertheless, the emphasis of planning is on the future. The entire purpose of planning is to achieve a preferred future, which may be a good future or simply the "least bad" future. In either case, planning is an attempt to apply rationality to the task of shaping the future.

Shaping the future is something that is done in the present. By selecting one action over another, we shape the future to some degree. Shaping the future is, thus, inevitable. Even if we choose our actions at random, we cannot avoid shaping the future. We plan in order to select those activities in the present that will lead to a preferred future. Planning provides the link between a preferred future and a present action. Ultimately, it provides a reason for choosing one action over another. The chosen action leads to the preferred future.

How far in the future can this "preferred future" lie? Clearly, there is no fixed answer. We may engage in meal planning by buying a package of breakfast rolls today in order to eat them tomorrow morning and in estate planning by establishing an annuity today that will be payable to our children at a time perhaps half a century in the future. Clearly, planning may deal with almost any span of time.

Much discussion of planning involves "short-range" and "long-range" planning. These are often distinguished in terms of years. For instance, it might be asserted that short-range planning deals with a future of 2 years, 5 years, or some such number of years into the future, whereas long-range planning deals with a future 20 years away. Distinguishing between short-range and long-range planning on the basis of time span, however, is completely incorrect. The true distinction involves a combination of the lead time for actions to have their effects and the length of time until the "preferred future" comes into existence.

Consider a situation in which some preferred future is desired 5 years hence, and suppose that the action that must be taken to achieve this future requires a lead time of 5 years to produce its results. The planner must take action right away if the desired future is to be achieved. Any delay in taking that action is tantamount to not choosing the preferred future. In this case, planning for a time 5 years hence is a short-range plan.

Consider another situation in which some preferred future is desired 3 years hence, and suppose that the action that must be taken to achieve this future requires a lead time of 1 year to produce its results. The action, therefore, cannot even be taken for 2 years yet. In this case, planning for a time 3 years hence is a long-range plan.

Is there any purpose to long-range planning? If action need not or cannot yet be taken to achieve the preferred future, why even be concerned with it? The reason is that when the time comes to take the necessary action, it may not be possible to do so. Other events may have intruded to make the action impossible. Worse yet, the planner may have inadvertently taken some action, in carrying out a short-range plan, that precludes the action needed for the long-range plan.

Long-range planning, just as any other kind of planning, deals with actions to be taken in the present; it links these current actions with some preferred future. The essential feature of long-range planning is that it deals with present actions to *preserve* or *create* the *option* to take decisive goal-oriented action at a time in the intermediate future when that action is necessary to achieve the preferred long-range future.

Another way to look at the distinction between short- and long-range planning is that in short-range planning the planner cannot wait for additional information; any decisions taken must be based on the situation as it already exists. Even though additional information might be available one-tenth of a "lead time" into the future, the planner cannot wait for it because the short-range plan requires action now. The long-range planner, by contrast, not only can but should wait for additional information. A good long-range plan is one that has some explicit, built-in process for gathering the additional information that can be available before decisive action needs to be taken.

An implication of the possibility of additional information is that the planner in a long-range situation may not, and even need not, be completely certain of the specific action that will be taken when the time for it comes. There may be a need to create or preserve options for several alternate actions when the time for action arrives. The lead times for all these actions might not be the same or they may not yet be precisely known. Thus, another aspect of a long-range plan is the need to gather information about the option that should be taken and to have this information available before the option with the longest lead time might have to be chosen.

There is, of course, no sharp boundary between short- and long-range planning. If the action needed to achieve some preferred future can be postponed by 24 hours, literally from today until tomorrow, this does not convert the situation from short- to long-range planning. From the perspective of the overall goal, if no new information can be obtained in the intervening 24 hours, the delay is simply procrastination. Thus, a short-range plan need not be acted upon right away; it is only necessary that no new information, change in resources, or change in external conditions be possible before the deadline for a decision imposed by the lead time for the action. Under such circumstances, a plan is short range, and the decision is effectively made "right away." So long as something can change before the deadline for the decision, even if that deadline is only a few minutes off, the planning is effectively long range.

It is sometimes useful to speak of intermediate-range planning as well. Suppose that to achieve some preferred future, we will need to exercise one of a limited set of options in 10 years; however, to exercise these options, we will need some physical facilities that take 5 years to construct. Unfortunately, we do not yet know how to construct them. Thus, our short-range plan may be to establish a *research and development* (R&D) program to learn how to construct the facilities. The construction itself then becomes an intermediate-range plan. It is not really a long-range plan, because it does not of itself bring about the preferred future. It is instrumental in nature rather than being an end in itself. It derives its justification from the long-range plan, which is to lead to a preferred future. Not all situations require intermediate-range planning. In some situations, however, it can be useful to recognize plans that do not require immediate action but, on the other hand, do not lead directly to a preferred future. These are intermediate-range plans.

One issue that can arise with plans of any degree of futurity is the level of detail involved. Some planners equate the goodness of a plan with the level of detail of the plan, i.e., with the degree to which the plan constrains future actions. As Samuelson (1990) has pointed out in an amusing but perceptive essay, this is simply wrong. The best plan is not the one that takes the most into account but the plan that *has* to take the *least* into account. As Samuelson puts it, "the best plan is the one which gives you the best chance to improvise in response to the unexpected." Such a plan is one that creates and preserves options, i.e., one that does *not,* by early actions, preclude the decision maker from responding to changed conditions.

In summary, the purpose of planning is to choose actions in the present that will lead to a preferred future. A short-range plan is one that deals with actions that can no longer be delayed if they are to have their desired effect. A long-range plan is concerned with actions taken in the present to assure that at some later time it will be possible to choose a preferred future. This involves the preservation and creation of options for later choice. However, the distinction between short- and long-range planning does not alter the fact that both provide a link between actions taken now and some preferred future.

13.3 The Role of the Forecast

In the first chapter, a list of purposes for forecasts was presented. These can be summarized in terms of taking actions to maximize the gain or minimize the loss from events beyond the control of the person using the forecast. This concept can now be placed within the context of planning.

A plan is analogous to a route chosen from a map. Where does the map come from? The forecast is analogous to the map; it describes the various possible events and the routes by which these events might be achieved. Thus, the role of the forecast is to describe the alternatives open to the planner. The forecast informs the planner of the possible destinations, the routes to those destinations, and the relative distance or difficulty of each route. The forecast does not impose any specific choices on the planner; instead it merely defines the choices.

However, the analogy between a route chosen from a map and a plan based on a forecast is not perfect. A traveler ordinarily expects to find towns and road intersections where a road map designates them and to find that the distances shown on the map are accurate. Even the best of forecasts, however, may show towns and intersections where none exist and may fail to show others that do exist. In addition, the distances predicted by a forecast should not be expected to be as accurate as those shown on a road map.

Of course this has some correspondence to traveling as well. The experienced traveler will be aware that even the latest road map may fail to show all the roads under construction and all the closed bridges that will require detours, not to mention the accidents that have temporarily blocked traffic. Nevertheless, we ordinarily expect a map to be much more accurate than a forecast.

Despite the potential inaccuracy, however, the role of the forecast is to define the possible alternatives. What possible destinations might we select?

What are the routes to get there? We expect a forecast to provide answers to these questions, i.e., a forecast reduces the uncertainty about what lies ahead.

This, of course, is true of any forecast, be it economic, demographic, or technological. The important point about a technological forecast is that it provides information about the technological alternatives, options, and consequences.

A technological forecast is important in the development of those plans that involve the development or deployment of new technology or the creation of technological options. It is also important in those plans that can be affected by technological change. Plans can be affected by technological change in several ways:

1. A technological change may provide new ways of achieving objectives.
2. A technological change may render certain means of achieving the objectives obsolete.
3. A technological change may render certain objectives obsolete.

The technological forecast then provides the background for selecting plans of any time horizon, short range or long range. It helps the planner identify those actions that must be taken in the short term either to achieve a particular technological capability by the time horizon of the plan or to create options that will be needed at future decision points.

13.4 Decision Making

Decision making can be defined as the act of choosing from among a set of feasible course of actions. This definition contains several elements, all of which are important.

First, decision making involves action. The action is either a change in the present situation or a deliberate choice to retain the present situation.

Second, the courses of action must be feasible, i.e., the decision maker must be able to carry them out. If there are no feasible courses of action, no decision is possible.

Third, there must be more than one course of action available. If there is only one course of action possible, there is no decision to be made.

Fourth, decision making involves selecting one of several courses of action. This usually implies some limitation on the resources available. If all the available courses of action can be pursued simultaneously, no decision is needed. A decision becomes necessary only when more than one course of action exists and not all can be pursued simultaneously.

Finally, decision making is an act. This means that the choice among courses of action must not be passive or allowed to happen by default. If the situation is allowed to drift until all courses of action but one are foreclosed or until circumstances force the choice of a particular course, then there is no need for a decision maker.

The purpose of decision making is to change things or at least to make a deliberate choice not to change. But what is there to change? We can think of decision making in the context of an organization, although the same ideas can be applied to the situations of individuals. A decision may change the objectives of an organization, its structure, the allocation of its resources, or the assignment of its personnel.

The objectives of an organization deal with the following questions. What function do we perform in society? What kind of an organization do we want to be? To the extent that we are able to shape society, what kind of a society do we want? Changing the objectives of an organization changes the answers to these questions.

The structure of an organization deals with questions such as these. Should we be organized by function, by product and/or service, by customer, or by location? Should we be more centralized or more decentralized? Should we integrate backward into the production of some of the raw materials or supplies we use? Should we integrate forward into closer dealings with the final user? Should we integrate horizontally by expanding our present activities or merging with other organizations with similar purposes?

The allocation of resources within an organization involves questions such as the following: Are there elements of the organization that need a greater share of the available resources? Are there elements of the organization that are receiving a share of the resources out of proportion to their contribution? Are there elements of the organization that are bottlenecks, where additional resources would benefit the whole organization?

The assignment of personnel within an organization involves questions such as these: Do we need to move people with certain skills from one element of the organization to another? Do we need more people with a particular skill than we now have? Are we going to have more people with a particular skill than we can effectively use?

These questions always arise when an organization is faced with change. Our particular interest, however, is the way in which these questions arise in organizations faced with technological change. The role of technological forecasting is to allow the managers of such organizations to do a better job of answering these questions. Technological forecasts provide the decision maker with the information needed to make wise choices among the possible changes in an organization while there is still an opportunity to do so. Forecasts do not force particular choices.

13.5 Summary

The purpose of planning is to take into account the future consequences of present actions. In particular, planning helps to select actions in the present in order to achieve a preferred future.

Forecasting provides the planner with an estimate of the kinds of futures possible and the specific actions that might lead to each of the alternatives. Technological forecasting in particular provides an estimate of the technologies that will become available so that the planner can take the greatest possible advantage of them and minimize their adverse impact.

Any decision will change some aspect of the organization for which the decision is made. Decision making provides the ultimate justification for planning and forecasting, since the purpose of a plan is to lead to a set of good decisions.

References

Samuelson, D. A., "The Travel Agent's Parable," *OR/MS Today*, **17**(3):6 (June 1990).

Problems

1 What are the advantages of using a formal written forecast as the basis for a plan?

2 Consider a paper company that grows its own trees. These trees take 20 years to reach usable size. Does the company's planting of trees this year imply a firm commitment to use them for making paper in 20 years? Why or why not?

3 Describe how the following features can be used to increase the flexibility of a plan.
- *a* Inclusion of multiple options at future decision points
- *b* Inclusion of steps to collect additional information prior to making needed future decisions
- *c* Inclusion of intermediate or lesser goals short of completion of the full plan

Is the inclusion of these features likely to make a plan more or less acceptable to a decision maker? Why?

Chapter

14

Technological Forecasting for Research and Development Planning

14.1 Introduction

The term "research and development" (R&D) is often used as though it described a homogeneous set of activities. This is not the case at all, and the differences among the activities lumped under this label are important. These activities do have some important similarities, however, and it is worth discussing the similarities before taking up the differences.

All R&D activities share the characteristic of something being done for the first time. They involve doing pioneering work in understanding new concepts as well as experimentation and doing trial studies coupled with attempts to predict or explain the results of experiments on the basis of scientific laws. The experiments often entail precise measurements and extensive calculations. In these characteristics, they differ from most of the activities customarily excluded from the category of R&D.

However, one of the most significant common characteristics of R&D activities is uncertainty. The uncertainties of R&D programs can be classified into three categories: technical uncertainty, target uncertainty, and process uncertainty. Technical uncertainty involves the question of whether technical problems can be solved in order to achieve a desired outcome. There is no way to resolve this question except by carrying out the program. Target uncertainty involves the question of whether the technical characteristics specified for an R&D program are the correct ones, i.e., even if a project meets its technical goals, will it perform satisfactorily in use? Process uncertainty involves the question of whether the human or organizational process of development will lead to a desired outcome. Even if technical goals are correctly chosen and are achievable, a project may fail for organizational or managerial reasons. Since the project is intended to do something that was never done before, there can be no certainty that the organization for carrying it out is correct.

The inherent uncertainty of R&D frequently gives rise to arguments that R&D cannot be planned. However, this argument must be rejected. As with any other activity, present actions must be linked with preferred futures through planning. Nevertheless, one important point must be kept in mind: The specific outcomes of R&D projects cannot be prescribed in advance. The greater the technical uncertainty in a project, the less possibility there is of prescribing the outcome in advance; the greater the target uncertainty in a project, the less possibility there is of knowing what outcome to plan for; and the greater the process uncertainty, the less possibility there is of knowing how to obtain the planned outcome.

Technological forecasting can play two important roles in helping the R&D planner cope with these uncertainties. The first is to help set goals for R&D programs; the second is to identify opportunities that might be exploited. Before these roles can be discussed in detail, however, we must look at the differences among the various kinds of activities lumped together under the heading of R&D.

R&D can conveniently be divided into four categories: research, technology advancement, product development, and test and evaluation. The uncertainties in each of these are somewhat different, and their respective goals are quite different. Hence, the role of technological forecasting can be discussed separately for each of the four categories. The remaining sections of this chapter will take up these categories separately.

14.2 Research

Research is a phenomenon-oriented activity. The researcher engages in research to gain new knowledge about some phenomenon in the universe. Research inherently involves a great deal of technical uncertainty. It is almost impossible to forecast the outcome of a research program except in general terms: "We will learn more about phenomenon X." Scientific breakthroughs are almost by definition unpredictable.

What, then, does technological forecasting have to offer? First, information about what phenomena are worthy of investigation, and second, information about what kinds of investigation will be possible. We will take these up in that order.

Phenomena will be investigated for a variety of reasons ranging from intellectual curiosity to the practical concerns of the organization sponsoring the research. It is primarily this latter group—the commercial and governmental organizations sponsoring research—that needs to know what phenomena should be investigated.

Such groups are involved with one or more technologies. In order to improve their functioning or to remain competitive in the marketplace, they must push their technologies to ever-higher levels of performance. This means that the devices embodying the technologies must face increasingly severe environments. These organizations should then be interested in knowing more about the phenomena that will be encountered in these environments. For instance, an aircraft company might be concerned with learning more about aerodynamics at higher speeds or altitudes. An automobile company might be interested in learning more about the behavior of fuels in internal combustion engines. A tire company might be interested in learning about the behavior of the materials it uses under conditions of high temperature and stress.

In all these cases, the obvious motivation of the company is to learn how to build a product that will work in the expected environment. But the research *as such* will not lead directly to the desired product or product improvement.

The output of the research effort is simply knowledge about phenomena. The company's designers can use the knowledge to design a better product, but the research effort itself may appear to have little relationship at all to specific products, especially to the company's present line of products.

Some classic examples of this kind of research, aimed at understanding phenomena rather than developing a new product, will illustrate the point. One instance is Irving Langmuir's work at General Electric (GE) on the fundamental physics and chemistry of metal lamp filaments, which won him the Nobel Prize in 1932. It also gave GE an edge over its competitors in knowing how to make efficient lamps. Another example, also from GE, is the work of Saul Dushman on electron emission from vacuum tube cathodes. This understanding of the basic physics of vacuum tubes gave GE a dominant position in the market for radio equipment. Yet another example is that of Wallace Carothers, who did basic research on the structure of polymers for DuPont, which led to his election to the National Academy of Sciences in 1936. Once the structure of polymers was well understood, the knowledge could be used to design polymers with specific characteristics. The knowledge led to commercialization of nylon, a very profitable product for DuPont. The issue here is that none of these researchers was working on a product per se. Each was working to gain fundamental knowledge that would be useful to his employer in improving products or developing new products.

If a company or some other organization is to remain competitive, it must have new knowledge to be used in the design of its products. But not just any kind of new knowledge will be helpful. The most useful new knowledge will be about a phenomena that the company's products will encounter or utilize as they deliver improved performance for the user. Here is an important role for technological forecasting. A forecast of the technology the organization employs or will employ can provide information about the knowledge that must be gained from research. A forecast of aircraft speeds and altitudes, for instance, can identify the range of phenomena that researchers for an aircraft company should be investigating. Similarly, a forecast of engine performance (e.g., horsepower-to-weight ratio or fuel consumption) can identify the range of phenomena that the researchers for an automobile company should be investigating.

This use of technological forecasting applies to all organizations doing research that is intended to lead to improvements in their products or services. The technological forecasts cannot predict what will be learned by the research. However, they can identify research needs by identifying ranges of phenomena that will be encountered but for which knowledge is lacking. These knowledge "gaps" then provide the basis for research programs to make the improved products possible.

Some organizations, such as universities, really have no need to identify research "gaps" as the basis for planning research programs. Even so, however, they need to plan the research they will undertake in order to assure that they will be able to carry it out. At the very least they need to plan for the acquisition of new instruments and other research equipment. They may also need new buildings or other facilities for their research. If these are to be available when needed, planning must be done ahead of time. Furthermore, this argument applies to companies and government organizations as well. If they are to continue to do research, they must plan to have the equipment they will need when they will need it.

The technology of research instrumentation and equipment is just as susceptible to prediction as is any other technology. Forecasts of instrument technology, then, can be used by research planners to assure that their equipment will be adequate to serve the needs of their research staffs.

Figure 14.1 shows a plot of the accuracy of astronomical angular measure derived from the data in Table A.38. The regression fit includes only the data points starting with 1580; this is because the two earlier data points were made with simple quadrants and without the use of a clock. The data points from 1580 represent the progress of modern astronomical instruments. As can be seen, the progress followed a regular pattern of exponential improvement from 1580 to 1935, the end of the recorded data series.

Appendix A contains data on the improvement of other scientific measurement technologies, each of which also shows considerable regularity in improvement over time. This regularity allows the performance of the technology to be forecast with considerable confidence.

These examples are not merely isolated cases. As with most technologies, there are patterns of regularity in the improvement of measurement and instrumentation technologies. An important role of technological forecasting is

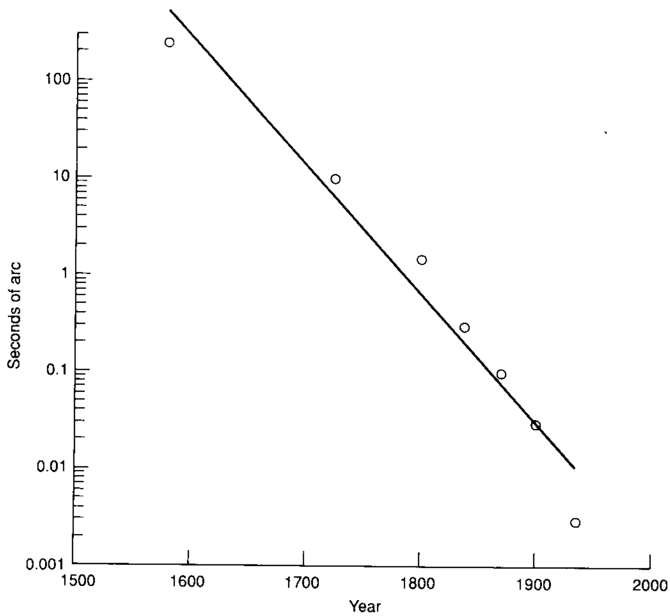

Figure 14.1 Accuracy of astronomical angular measure.

to predict the future capability of instruments and equipment for scientific research. These forecasts can be used by research planners to determine the needs of their laboratories and to plan for the availability of equipment with the best available performance.

Research planners can utilize technological forecasting to help them link present actions with desired outcomes from their research programs. Technological forecasts can provide information about the specific phenomena that are worth investigating in order to provide knowledge needed for the improvement of particular technologies. In addition, forecasts of instrumentation technology can help research planners determine what kinds of research will be possible and assure that their laboratories are equipped to carry it out.

14.3 Technology Advancement

Unlike research, technology advancement is a problem-oriented activity. Research is concerned with learning more about a particular phenomenon. Technology advancement, by contrast, is concerned with learning how to solve a particular problem or perform a particular function. Persons engaged in technology advancement are not limited to the use of only one particular phenomenon or to the use of only certain kinds of knowledge. Their concern is with solving a problem, and they may draw upon any knowledge or phenomena that they find useful.

Like research, however, technology advancement does not result in a specific product ready for use. The purpose of a technology-advancement project is to reduce technical uncertainty. It is intended to answer the question, "Can it be done at all?" This means that a significant portion of all technology advancement projects will reach negative answers. They may conclude that a particular problem cannot be solved at all or that at least it cannot be solved with the currently available techniques or knowledge. Such a result may not be popular with the sponsor, but it is a legitimate resolution to the technical uncertainty existing at the outset of a project.

Technology advancement programs are not carried out for their own sake. The information about whether a problem can be solved by any of the available means is eventually intended to lead to a product for some user. There are two major issues facing the planner of a technology advancement program. First, what level of performance will be required when the results of the program are ready to enter product development? Second, will the current technical approach be capable of delivering the required level of performance? The first question can also be turned around to ask when will a specific level of performance be available for incorporation into a product? The second can be rephrased as, "When will a new technical approach be required?" Both these questions about the overall level of performance of a technology and the technical approach that will be utilized to achieve it can be answered by technological forecasts.

A forecast of the overall performance of some technology can help answer the first question. Here, it is important to recognize that it is the overall technology, and not a specific technical approach, that is important. This forecast can be used either to identify the performance that can be expected by a particular time or to determine when a particular level of performance can be achieved.

Once the overall technology has been forecast, the technology planner can then look to forecasts of individual technical approaches. How long can the current technical approach keep up with the overall technology? If it will begin to reach its limits soon, is there a successor technical approach available? If so, when will its performance begin to compete with the current technical approach? If not, how much time is there to look for a successor?

Information about technical approaches is particularly important to technology planners. They need to know when current laboratory facilities will have to be updated or replaced, the kinds of skills that will be required in the laboratory staff, as well as the goals their specific projects should have. This way, they are neither striving after some level of functional capability that will be nearly impossible to achieve in the desired time nor pursuing something that is so easy that other competitive groups are likely to surpass it.

Figure 14.2 illustrates the kind of information that can be provided by a forecast. It displays the overall trend of aircraft speed, based on aircraft speed records from 1906 to 1939. Both an exponential trend and a Gompertz curve have been fitted to the data. By the late 1930s, it is clear that the growth curve is dropping below the trend. At the time, the limitation was known to be the speed of sound. As the aircraft approached the speed of sound, first the propeller became less and less effective (propeller tip speed will exceed the

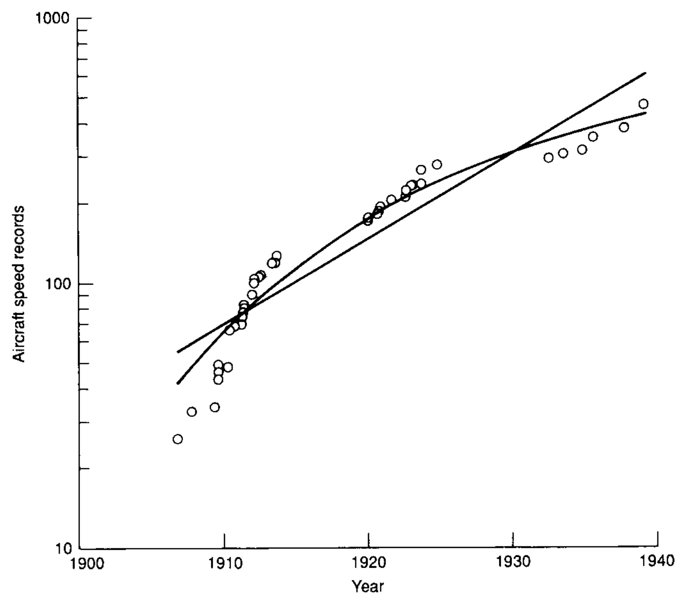

Figure 14.2 Trend and growth curve for aircraft speed.

speed of sound before the aircraft even reaches the speed of sound), and finally a shock wave would build up in front of the aircraft which kept it from accelerating to any higher speeds.

However, the speed of sound was not a fundamental limitation. Artillery shells and rockets had already exceeded the speed of sound by that time. Thus, it was clear that aircraft speed might likewise exceed the speed of sound if the proper propulsion were available. Moreover, as related in Chap. 10, the jet engine was already under development.

Hence, the discrepancy between the growth curve for propeller-driven aircraft and the long-term trend for aircraft speed could be used to indicate how soon a replacement for the piston engine would be needed, and what performance it would have to achieve in order to be competitive with the current and projected performance of piston engines.

In addition to a replacement for the piston engine and propeller, higher aircraft speeds would require new materials for aircraft structures. Aircraft instruments that would operate correctly at higher speeds and altitudes (e.g., airspeed indicators, altimeters, and attitude indicators) would also be needed. In each case, the issue to be examined is whether the current technical approach is capable of achieving the desired performance, or whether a successor technical approach will be needed.

This example illustrates the application of technological forecasting to technology advancement programs. Forecasts enable the planner to identify the performance required of different technologies on the basis of the demands that will be placed on them.

The technology planner needs to use forecasts of both the overall technology and of the current and potential successor technical approaches. These forecasts, both individually and in combination, indicate what performance to expect and whether it is time to switch to a successor technical approach. Once the appropriate technical approach is selected, the planner can determine what laboratory facilities and technical skills will be required to carry out a technology advancement program.

14.4 Product Development

Like technology advancement, product development is a problem-oriented activity. However, its concern is not with learning how to solve a problem but, rather, with how to embody that knowledge in something a customer or user will find satisfactory, i.e., technical uncertainty should already have been resolved by a technology advancement program. Once the stage of product development is reached, there should be little technical uncertainty left. The primary concern at this point is target uncertainty.

There are two aspects to target uncertainty. The first is what the customers or users will want. The second is what competitors will be offering for performing the same function. Technological forecasting can provide little or no information about customer or user desires. Its main function in product development is to provide estimates of the performance of competing devices or techniques that might be offered at the same time or shortly after a new product is to be marketed.

The selection of a design goal for a product development project must represent a compromise between two conflicting objectives. On the one hand, the project should represent as small a technical challenge as possible. Aiming too high increases the cost of the project and increases the risk of technical failure. On the other hand, the project should be as ambitious as possible. Aiming too low increases the risk that a competitor will bring out a superior product shortly after the project is completed. The goal set for the project must, therefore, balance the risks of technical failure of the project and early obsolescence of the product.

In Chap. 6, we regressed time on the technical performance of a series of devices to estimate the degree of technology that was obtained. This technology measure can be used to estimate the likelihood that a particular level of functional capability will be reached by a particular time.

Use of the method can be presented most easily if the technology is described by only a single parameter. Ordinarily a regression fit of the data for that single parameter is made using the following equation:

$$Y_i = A + BT_i + e_i$$

where Y is functional capability, T is time, e is a random error, i refers to the ith data value, and the coefficients A and B are obtained from the fitting procedure. Confidence intervals can be placed on the fitted equation, giving the range of Y values that might be achieved at a specific time T for a given probability. The technology measure reverses this process, using instead the equation:

$$T_i = A' + B'Y_i + e_i$$

where all the terms are as above except that A' and B' are the coefficients obtained by regressing T on Y. Placing confidence intervals on this equation gives the range of T values by which a given value of Y might be achieved.

Figure 14.3 illustrates such a "reverse" regression, i.e., T as a function of Y. However, Fig. 14.3 is plotted in the conventional form, with T on the abscissa and Y on the ordinate. The 50 and 90 percent confidence intervals have been plotted. However, they must now be read horizontally instead of vertically. The 90 percent confidence interval, for instance, gives the range of T within which it is 90 percent likely that some target level of Y will be achieved. Since the probability distribution is assumed to be symmetrical about the fitted line, the early 90 percent probability contour gives the time by which it is 5 percent likely that the value of Y will have been achieved. The late 90 percent probability contour gives the time by which it is 95 percent likely that the value of Y will have been achieved. The fitted trend itself gives the time by which it is 50 percent likely that the value of Y will have been achieved. Other likelihood values can be plotted if they are desired. The figure also shows the time interval within which it is 90 percent likely that the performance target level will have been achieved. The

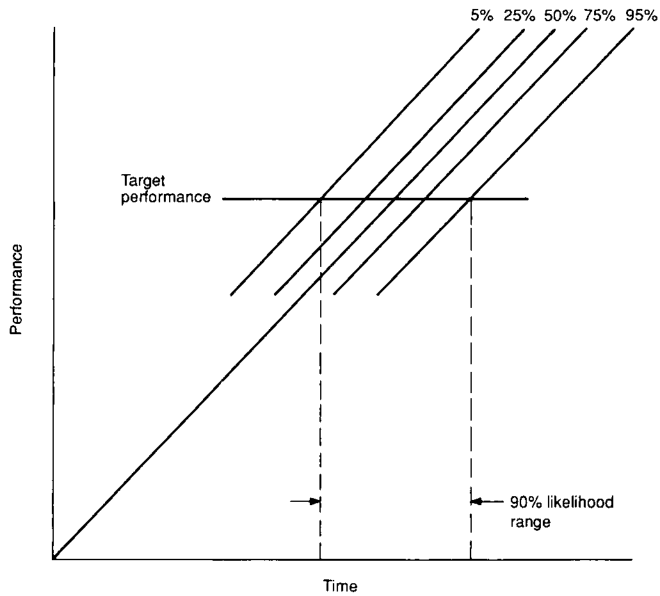

Figure 14.3 Reverse regression of time on performance.

50 percent probability contours likewise give the times by which it is 25 and 75 percent likely that the value Y will have been achieved.

Note that this reverse regression can be carried out by regressing time on time the logarithm of Y if functional capability has been growing exponentially. If the functional capability has been following a growth curve, the value can be linearized as described in Chap. 4, and time regressed on the linearized value. Therefore, this approach is not limited to technologies that grow linearly.

When functional capability involves two or more technical parameters, the same procedure can be applied. However, it cannot be illustrated as easily as for the case of only one technical parameter. The principle remains the same. The standard error of the forecast is computed, and confidence bounds for T are calculated. These give the probability that a given value of the measure of technology will be achieved by a particular date. Put another way, the confidence bounds give the probability that a particular trade-off surface for the technical parameters will have been reached by a particular date.

The forecast of the likelihood of a particular level of functional capability being achieved by a particular date gives the risk that some competitor will have a product with that performance on the market. Once this estimate of the risk is obtained from the forecast, it can be balanced against the risk of technical failure of the development project, and a goal that properly balances these risks can be chosen. As usual, the forecast provides information on which to base a decision, but it does not force a particular decision. The R&D planner is still free to select the goal on the basis of estimates of the cost of failure and the cost of early obsolescence.

14.5 Testing and Evaluation

Testing and evaluation are also problem-oriented activities. They are concerned with determining whether some new device will actually perform as it should in the environment in which customers or users will employ it. There are three different groups that might be concerned with testing and evaluation: The developer of a device is interested in validating a design and determining whether changes are needed before the device is marketed or deployed for operation; the customer or user is interested in determining whether a device meets its performance specifications and whether it is suitable for use; and finally, a regulatory agency may be interested in determining the effectiveness or potential hazards of some newly developed device, material, or technique. Each of these three groups may then subject the item to testing and evaluation.

Testing and evaluation may be carried out by either of two means or by a combination of both. First, the new item may be tested in the actual environment in which it is to be used. Second, the new item may be tested in a simulated environment.

A simulated environment may be something like a wind tunnel, an altitude chamber, or a water bath. The intent is to create an artificial environment that matches the one in which the item will actually be used. A simulated environment may be cheaper to achieve than the real thing, especially if the environment is an exotic one such as outer space. It is usually easier to attach an instrument for taking measurements in a simulated environment than in a real environment. Finally, a simulated environment makes it possible to repeat tests under identical conditions. However, if an environment is to be simulated, it is necessary to specify it in detail, usually in terms of its physical and chemical properties. Even then, there may be some question about the extent to which the "test chamber" actually matches the environment in which the item will be used.

Despite the problems of instrumentation and the difficulty of precisely repeating sets of conditions, testing in an actual environment rather than a simulated environment is done frequently. The major advantage of this is that it is not necessary to specify the environment in detail. In fact, if the environment cannot be specified in detail, there is no substitute for testing in the actual environment.

In dealing with testing and evaluation, the R&D planner is faced with two different problems. The first is determining what will need to be tested. It is necessary to know what will have to be tested in order to plan for adequate testing capability. The second is determining what testing will be possible. It is necessary to know what kinds of instrumentation or other test equipment will be available in order to determine how the required testing can be carried out.

A forecast of the technology to be tested can provide much useful information to the planner responsible for testing and evaluation. Such a forecast can be used to determine by how much an existing test facility must be upgraded or extended, or even whether an existing facility can be upgraded to the degree required. If a new test facility is to be constructed, such a forecast can help avoid building one that will be outgrown before it is physically worn out. A forecast can also help avoid making a major investment in a facility to test an obsolescing technical approach and direct investment toward a facility suitable for a successor technical approach.

A forecast of testing technology, including both simulation technology and instrument technology, can help the planner determine whether testing should be done in a real or a simulated environment, and how it can best be done in either environment. For instance, it may be desirable to conduct certain tests in a simulated environment because repeatability is needed or because the risk of catastrophic failure of the test item is intolerable in the real environment. A forecast of simulation technology will allow the planner to determine whether it will be possible to conduct the testing in a simulated environment and, if so, the range of environments that may be adequately simulated. If testing is desired in the actual environment, forecasts of instrument technology will allow the planner to determine the extent and nature of the tests possible.

In some cases, forecasts of both simulation technology and instrument technology will aid in selecting between the two or in selecting the proper mix of the two. For instance, if advances in simulation technology make it possible to reproduce accurately and cheaply an environment that can be reached only with difficulty in nature, a switch from testing in the real environment to testing in a test chamber might be appropriate. Conversely, if a forecast of instrument technology shows that instruments will be lighter, more rugged, or have remote sensing capability, it may be possible to switch from simulation testing to testing in the actual environment.

Whether the testing is done by simulation or in the real environment, there may be several suitable alternative instrument technologies. A forecast of these technologies will allow the planner to maintain the option of using whichever turns out to be the best at the time. Otherwise, certain actions might be taken that inadvertently preclude the use of one or more technologies.

Testing and evaluation is an important "last step" in the R&D process. Forecasts of the technology to be tested can be used to assure that the testing and evaluation capability will be available when needed. Forecasts of the technology available for test and evaluation, including simulation technology and instrument technology, can be used to assure that the testing and evaluation activities are carried out in a technically effective manner and at the lowest cost.

14.6 Summary

The R&D planner is faced with planning for something that is to be done for the first time and with no certainty that it can be done at all. In fact, the reason for attempting to do it is to determine whether it can be done. The R&D planner, like all other planners, must decide what to do in the present in order to enhance the likelihood of the future success of the R&D program.

Technological forecasts can be helpful to the R&D planner by identifying reasonable goals for projects, likely levels of the performance of products offered by competitors, and the capabilities of instruments used in R&D. Technological forecasts can also be helpful in identifying when specific technical approaches may have to be abandoned in favor of successor approaches.

The role of the technological forecast is to provide information for decision making. The forecasts do not force particular choices or courses of action on the R&D planner. However, the information they do provide can be combined with other information about the costs and benefits of various outcomes to determine what actions should be chosen in the present and what options should be preserved for a possible future choice.

Problems

1 The staff at the Lawrence Radiation Laboratory, University of California at Berkeley, pioneered the development of liquid hydrogen bubble chambers for research in particle physics. The sizes of the bubble chambers built by this group are given in the following table:

Year	Size	Remarks
1954	1.5 in diameter and 0.5 in deep	First liquid hydrogen chamber to show tracks
1954	2.5 in diameter and 1.5 in deep	First metal-wall chamber
1955	4 in diameter and 2 in deep	
1956	10 in diameter and 6.5 in deep	
1957	15 in diameter	Replaced 10-in chamber
1959	72 × 20 × 15 in	Held 300 liter of liquid hydrogen

SOURCE: Adapted from Shutt, R. P., *Bubble and Spark Chambers,* Academic Press, New York, 1967.

The functional capability of these bubble chambers is determined by the area (i.e., either by the area of the circular face of the cylindrical chambers or by the product of length times width for the rectangular chamber). Calculate the area for each of the above chambers and plot them on an appropriate scale. Should the first chamber, which was built solely to demonstrate the feasibility of liquid hydrogen bubble chambers, be included in the determination of a growth law for bubble chamber area? Why or why not? What growth law seems to describe bubble chamber area?

2 Assume the year is 1940. You have prepared the forecast of the divergence between the growth curve and trend for aircraft speed given in this chapter. What are the implications of your forecast in each of the following areas? Research projects to be undertaken (phenomena to be investigated)? Technical approach to be pursued for propulsion? Test facilities to be required, including wind tunnels and engine test stands? Perform a regression of aircraft speed records on time for the range of years used in this chapter. Compute the 90 percent confidence limits. By when is it 5 percent likely that a speed of 700 mi/h will have been achieved? 95 percent likely? How does this likelihood range compare with actual experience?

3 Using the table of electrical transmission voltages from Table A.35, prepare a forecast of the range of transmission voltages likely during the period 1975 to 1985. What voltage range would you recommend for testing facilities to be constructed by electrical equipment manufacturers, with construction starting in the period 1970 to 1975? In retrospect, how well would your forecast have turned out? What were the actual voltages used?

Chapter

15

Technological Forecasting in Business Decisions

15.1 Introduction

This chapter deals with the use of technological forecasting by business firms, i.e., by organizations that are privately owned, produce goods or services for sale in a market, and are in competition with other firms providing alternative goods or services. The essential characteristic of a firm is that it must obtain income from the sale of goods and services in excess of the cost of providing them. Moreover, it must do this in the face of competition from other firms that are also trying to attract customers.

From the standpoint of society, an important aspect of the business firm is its expendability. If a business firm fails to satisfy its customers, it goes out of business. This is in contrast with an organization such as a post office, a municipal hospital, or a state university, which remain in business no matter how poorly they satisfy their customers. Thus, the set of business firms in existence at any time consists primarily of those that are doing a good job of satisfying their customers.

From the standpoint of the individual firm, this expendability is viewed with much less equanimity. The failure of a firm results in direct losses to the owners and workers. Losses are also experienced by the suppliers who must find new customers, and by the customers who were still satisfied and who must now find new suppliers. Thus, the failure of a firm is something the people involved with the firm wish to avoid.

If a firm is to stay in business, it must continue to satisfy its customers. It must anticipate changes in their wants and needs, as well as changes in the ways these wants and needs can be satisfied. There are a great many aspects to this anticipation of change. Technological forecasting is one tool the business manager can use to help anticipate such changes.

15.2 What Business Are We In?

An important element of business success is to identify just what business one's firm is in. All too often, firms identify the nature of their business only in the most superficial terms: "We manufacture product X," or "We provide service Y." This identification is often too narrow, leaving the firm vulnerable to shifts in technology (as well as other shifts such as consumer tastes). A firm's vulnerability can often be reduced by recognizing that the nature of its business can be identified more broadly than by simply a good or a service.

How can the business of a firm be identified? A positive answer to one of the following questions will identify the nature of the firm's business:

Do we perform a specific function?

Do we make a specific product or provide a specific service?

Do we utilize a specific process?

Do we utilize a specific distribution system?

Do we possess a particular set of skills?

Do we use a specific raw material?

At any point in time, of course, a firm makes one or more specific products, uses a particular distribution system, and so forth. The important issue is which of these items characterizes the firm? Which is the last item the firm would give up if it had to change? Which is the one that gives the firm whatever unique advantages it has in satisfying the customers? The single feature that most characterizes the firm identifies the firm's business.

Some firms perform a specific function. Banks, for instance, are in the business of borrowing money from depositors and lending it to borrowers. They may adapt to technological and social change by using automated teller machines, providing bank credit cards, accepting electronic deposits from employers and electronic withdrawals from depositors, etc. The exact form of the business may change, but the function remains the same.

Some firms make a specific product. Steel mills, for instance, sell steel to a variety of customers and over the past century have made steel by a succession of different technologies. A steel company may change its technology, its sources of ore, or the way in which it ships products. The product itself, however, remains much the same.

Some firms utilize a specific process. Telegraph companies, for instance, send messages by using electrical signals that encode the information presented by the sender. Originally these companies transmitted text messages between individuals; later the bulk of their business was data of various kinds, such as stock prices and weather reports. Now the bulk of their business is becoming the direct transmission of data between computers instead of between people. While the nature of the data and the type of customer that is served have changed, the process has remained the same.

Some firms are identified with specific distribution systems. Mail-order firms such as Sears are perhaps the prime example of this type of firm. A firm identified with a specific distribution system may change the nature of the products it sells; it may change the type of customer it seeks; and it may change the geographical area it covers. The thing that remains constant is the use of a particular distribution system.

A firm may be identified with a particular set of skills. Firms specializing in law, engineering, architecture, accounting, tax preparation, and custom computer programming are examples of such firms. They are not identified with a particular raw material or a process; they may change their customers and the way in which they deliver services. The constant factor is the set of skills they offer.

A firm may be identified with the use of a specific raw material. That is, they utilize a specific raw material themselves. Wood products firms are an example of this identification with a raw material. They may switch their output from plywood to boards to chipboard as the market demand shifts. However, they are always concerned with finding something that can be made from wood. It is wood that is the identifying mark of the business, not the things that are made from the wood.

Only the managers of a firm can decide what business the firm is in. Outsiders can never be certain what decisions will be made in the face of change, e.g., to stick with a set of skills and seek different customers or to keep the same product but change distribution systems. However, the managers of a firm do not have unlimited freedom to choose the nature of their business. This is constrained by past history and present circumstances. Often it is a matter of discovering, rather than deciding, what business the firm is in. The managers must determine the aspect of the business that is the hardest to change, the one that would take the longest to change, or the one that is least profitable to change. Whether by decision or discovery, however, identifying the nature of the business in which a firm is engaged is one of the major responsibilities of its managers.

Once the managers of a company decide what business they are in, they are in a better position to evaluate the consequences of technological change. In some cases technological change can bring an end to the very business they are in. A frequently cited example is the destruction of the buggy-whip business by the automobile. No amount of improvement in buggy-whips could have saved the business, once the automobile replaced the horse on a large scale. A less drastic change in technology may leave the basic business in existence but change the way in which it is conducted. For instance, steel companies have adopted a succession of steel-making technologies, each of which reduced the cost of production. Petroleum refineries have adopted a succession of new processes for extracting gasoline from crude oil, each of which reduced cost, produced less pollutants, "cracked" more crude into gasoline, or provided some other advantage over the older process.

It is, therefore, extremely important for managers to be alert to technological changes that can alter the way they do business. Changes that destroy the business completely are fairly rare. Lesser changes that alter the way in which the fundamental business of the firm is conducted are more common. These changes, if not anticipated, may leave the firm unable to compete effectively with firms in the same business that did anticipate the changes properly. Hence, such changes can be just as fatal as a change that eliminates the business itself. Thus, the managers of a firm must also be alert to technological changes that leave the basic nature of the business unchanged but that alter some important aspect of the way in which it is conducted.

15.3 Business Planning for Technological Change

Once a firm's business has been determined, threats and opportunities from technological change can be more clearly identified. With regard to the business of the firm, a fundamental question is: Will this business continue to exist? If it will and the firm wishes to remain in that business, the next question is: How will this business be conducted?

Technological change can alter the fundamental nature of the firm's business; more frequently, it can alter the way in which that business is conducted. To illustrate these possible impacts, we will review some historical examples of technological change that altered important aspects of certain businesses.

Function. A firm may perform some function directly or it may sell a product that performs some function. In each case, the firm is vulnerable either to a technological change that makes the function unnecessary or to one that performs the function in some other manner. At the end of the nineteenth century, for instance, the most common means of lighting homes in large cities was the gas lamp. By 1910, however, the gas light was on its way out. The thing that destroyed the gas companies, however, was not a cheaper way of producing gas or a more efficient way of distributing it. Gas lights were replaced by a superior technology, electric lights. The function of lighting was performed in an entirely different manner. Gas, of course, is still used for heating. However, electricity is no longer used just for lighting but also for heating and for operating electrical and electronic appliances of all kinds.

Product. A product is sold to customers who utilize it for some purpose. A technological change may allow some different product to be utilized for the same purpose. Transistors, for instance, have replaced vacuum tubes except for very specialized applications. The consumer who buys a radio really does not care whether it contains vacuum tubes or transistors, just so it performs satisfactorily. The advantage of transistors was that they allowed radios to perform the same old functions, but more cheaply and in a smaller, lighter package. From the standpoint of the user, a transistor radio is a better radio, but it is still a radio. It is worth noting, however, that none of the firms that were leading manufacturers of vacuum tubes are now leading manufacturers of transistors. Moreover, most of the leading manufacturers of transistors never produced vacuum tubes. The advent of transistors gave new firms an opportunity to take the market away from those firms whose products served the same function.

Process. In many cases, a new process for manufacturing some product or providing some service can almost directly replace an old process. The advent of a sequence of new processes for making steel and for refining petroleum has already been mentioned. When these new processes were innovated, the firms that did not adopt them lost a share of the market or even went out of business because they could not compete with firms that did adopt the new processes.

These new steel-making and petroleum-refining processes replaced only one or a few steps in the total manufacturing process. The raw material remained the same, and most of the steps in the manufacture remained the same. In some cases, however, a technological change may affect the total process of manufacture. At one time in the electronics industry, for instance, one could distinguish among firms that made electronic materials, firms that made components out of these materials, and firms that assembled final products out of these components. A single firm might be involved in each of these steps, but the steps were still distinct and often carried out in different plants. With the advent of integrated circuits, however, the situation has changed. No longer are discrete components manufactured out of special materials and then assembled into products. The entire product may be formed as an integrated circuit by "depositing" layers of materials on a "substrate." By putting the layers down in a specific order and by masking and screening certain portions of the circuit between layers, thousands of individual components may be formed on a single circuit less than an inch in diameter. Producing a portable radio may mean nothing more than attaching a speaker, a battery, and a tuning circuit to an integrated circuit that contains everything else needed for a radio. The sequence of steps of preparing specialized materials, making components out of them, and assembling the components has been replaced by a single process. As with transistors and vacuum tubes, most of the firms that are leading producers of integrated circuits simply did not exist in the days of vacuum tubes, and few of the firms that were once leading manufacturers of components are now active in producing integrated circuits. A technological change in a process allowed new firms to displace those that had formerly dominated the market for electronic components.

Distribution. Technological change has altered the way in which firms distribute their products. The use of air freight has meant that high-value products can be delivered to a customer across the country on an overnight basis. While air freight costs more than other transportation on a ton-mile basis, its use may save enough to offset its cost. If regional warehouses near customers can be replaced by a central warehouse plus air freight, the size of the inventory needed can be reduced. This reduces storage costs, insurance costs, and the cost of capital tied up in inventory. Firms failing to make use of distribution innovations such as air freight may find themselves at a cost disadvantage relative to competitors who do make use of them. Thus, technological innovations that alter distribution may affect a firm's competitiveness.

Skills. Technological change may alter the skills required by the workers in a firm. There are at least two ways in which this can occur. If the process for making a given product changes, the workers may need new skills for the new process. If the nature of the product itself changes, the skills needed to make it may change as well. An example of the first situation is the replacement of manually controlled machine tools with computer-controlled ones. With the former tools, the skill of the operator was the critical element in making a part to specifications. The operator directly controlled the tool, measured the part during machining, and was responsible for the quality of the finished part. An automated machine tool generally performs these operations by itself. The operator, however, must be able to set it up, insert the proper program, and verify that the tool is working properly. The operator may actually need as much skill as the old-time machinist, but the specific skills needed are different. An example of the second situation is the replacement of mechanical cash registers by electronic ones. The firms that formerly made mechanical cash registers had large work forces of machinists, tool and die makers, and similar workers. When mechanical cash registers were replaced by electronic ones, a completely different set of skills was required by the work force. The skills that had provided the basis for the superiority of the leading mechanical cash register firms were no longer of any use. Electronic assembly skills were required. The firms involved had to replace not only their designs but their work forces as well.

Raw materials. Technological change can cause a change in the raw materials used by a firm to carry out some process or make some product. Frequently this comes about through the replacement of a natural material by a synthetic one that is cheaper or better. The replacement of silk by nylon and the replacement of cotton by a whole host of synthetic fibers are examples. In other cases the change may not involve the raw material directly but the product itself, causing an indirect change in raw materials. As aircraft technology improved, wood-and-fabric construction was replaced by aluminum construction. Later, aluminum was replaced by titanium and by stainless steel for applications requiring temperature resistance. These changes in aircraft structure meant that aircraft firms had to change the nature of the raw materials they used.

Historically, each of the major aspects of a business has been affected by technological change. However, a firm may be affected in two other aspects as well: management itself and support functions within the firm. Impacts on these are reviewed below.

Management. Technological changes may alter the way in which a firm is managed. The impact may arise directly from new management technology or it may arise from technology that allows more efficient management. New management technology has had a major impact on firms within the past several decades. The introduction of management science and operations research has reduced to routine calculations many of the decisions that formerly required management judgment. Examples include such things as setting inventory levels and reorder points in stockrooms and warehouses, selecting warehouse locations to minimize shipping costs, and selecting the number of stock clerks or service booths to balance the cost of providing service and the loss from time spent waiting for service. The introduction of a new technology, especially communications technology, may alter the ways in which management operates. For instance, the introduction of the telegraph in 1851 allowed the centralized dispatching of trains. Previously, when a train was delayed, all the other trains that were to meet it or use the same track were held up as well. Only local control was possible. Local dispatchers would not dispatch a train until they were certain the track would be clear through arrival of the delayed train. With the advent of telegraphy, the location of every train could be tracked by a central office, allowing central dispatching.

In general, improved communications allows a more central control of stocks of goods, of shipments from different locations, and of the movement of vehicles. This centralized control often (but not always) allows the total resources of a firm to be used more efficiently, since the need for, e.g., safety stocks at many locations or vehicles held in readiness at many points is reduced or eliminated.

Support. In a sense, this is a catchall term for the activities a firm must carry out simply to operate at all. There have been many technological changes that affect the support functions of a firm. Changes in behavioral technology, such as aptitude tests, have had an impact on hiring practices. Computers have affected support functions such as computing payrolls, writing paychecks, storing and maintaining records, and producing routine reports from stored data. The development of copying machines has altered how firms handle and file correspondence, how they maintain records, and how they distribute written communications within the firm. More recently, the development of the facsimile (fax) machine has affected not only communications within the firm but communications with customers and suppliers.

These examples have focused on one aspect of the firm at a time. However, a major technological change can bring about changes in almost every aspect of the firm's business. A current example will help illustrate this point. Photography has traditionally involved capturing an image in a photosensitive coating on a thin plastic film. The "latent" image remains on the film until the film is removed from the camera and developed. Firms that are in the photography business make cameras and film. This requires skills in optics, photochemistry, the mechanics of shutters, and electronics for controlling the camera. Manufacturing film requires that many of the operations take place in total darkness, almost necessitating that these operations be automated.

A combination of new technologies is now threatening to replace the film for photography, at least for a part of the market. This is the recording of still images on a magnetic medium, either disk or tape, which can then be played back through a computer or a TV set. With such a "filmless" camera, the picture does not need development. It is ready to view as soon as it is taken. Moreover, the magnetic image can be sent anywhere else via a modem and phone line, or incorporated into a document being prepared on a word processor.

Such a new technology would confront a film-and-camera firm with several challenges. Is it in the photochemical film business? Is it in the pictures business? Is it in the camera business? Or in some other business? How does the new technology fit in? If the firm attempts to adopt the new technology, new skills would be needed in its *research and development* (R&D) labs. Its film manufacturing facilities would be totally unsuitable for the new recording medium. New manufacturing facilities would be needed, and the work force either must be retrained or replaced. New raw materials would be required. Its sales staff would have to be retrained to sell the new products. Moreover, the change to the new technology would have to be managed without destroying its old business, which would have to provide the cash to finance the change. Clearly, the earlier a film-and-camera firm recognized the forthcoming technological change, the better position it would be in to respond.

These examples illustrate that technological change can have an impact on all the aspects of a firm's business, including its management and support activities. An aggregate look at the impact of technological change on firms is also of interest. Consider the 100 largest firms in the United States in 1917. Fewer than half of these are still among the top 100 firms, and over a quarter of them are no longer in business. Central Leather was in 1917 the twenty-fourth largest firm in the United States, with profits larger than those of Sears Roebuck, and a net worth greater than that of General Motors. It failed to respond quickly enough to changes in shoe-making technology that required different kinds of leather. It was undercut by other companies that responded more quickly and is no longer in business. American Woolen did not respond quickly enough to the consumer demand for synthetic fibers; it is no longer in business. In 1940, Curtiss-Wright was the twenty-eighth largest company in the United States. It was larger than either Boeing or Lockheed and was a major producer of high-performance aircraft during World War II. Its managers decided that the jet aircraft had no future as a civilian vehicle. They were wrong. The company is still in business, but not only is it not among the top 100 firms, it is not even among the top 500.

15.4 Market Share during Transition from Old to New Technology

One of the surprising results of research in technological forecasting deals with behavior of the sales of a "defender" technology under attack by a "challenger" technology when the total market is growing, i.e., firm(s) involved with the defender can unwittingly ignore signs of the impending replacement of their technology. We will examine how this can come about.

Assume that the total sales (or installations, capacity, etc.) of some technology is growing exponentially. This growth can be characterized by its doubling time, i.e., the time for the level of sales or capacity to double. The total sales can, thus, be expressed as:

$$\text{Total sales} = S_0 10^{\log(2)t/D}$$

where t is time, D is the doubling time for sales growth, and total sales are S_0 at $t = 0$.

Assume that the market share of the challenger is given by a Pearl curve. This can be characterized by the T-time, i.e., the time for the Fisher-Pry ratio to increase by a factor of 10. The market share of the defender is of course the complement of the market share of the challenger. If we let the Fisher-Pry ratio be 0.01 at time $t = 0$, then the fractional market share of the defender is given by:

$$\text{Market share} = 1 - \frac{1}{1 + 10^{2-t/T}}$$

The total sales for the defender technical approach are, of course, the product of total sales and the defender's market share. Let us assume the defender's sales at $t = 0$ are 100 units. Then by the assumption that the Fisher-Pry ratio is 0.01, the challenger's sales at $t = 0$ are 1 unit, and $S_0 = 101$ units.

We can parametrize the defender's sales in terms of the ratio of doubling time D to T-time T, or D/T. We can further simplify the expression by letting T-time $T = 1$.

Figure 15.1 shows the results, in parametric form. Defender's sales are plotted for three different ratios of D/T. The smaller this ratio, the faster the total

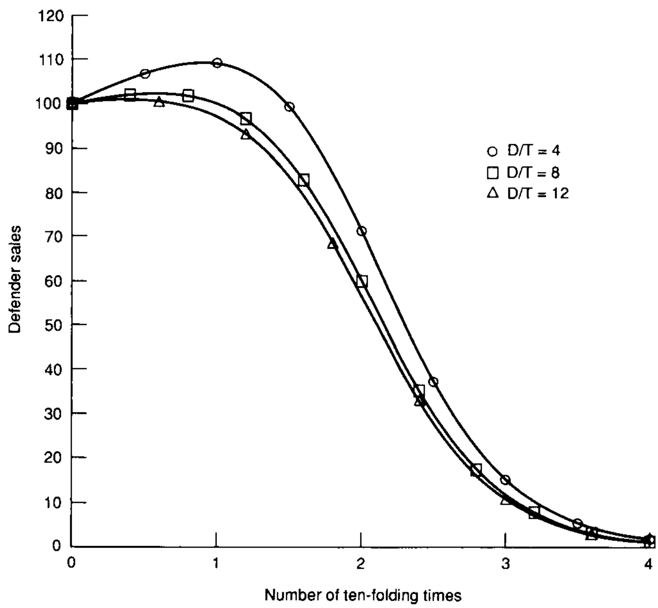

Figure 15.1 Defender sales for various D/T ratios.

sales are growing relative to the T-time of the challenger. If D/T is only 4, the defender's sales actually grow by about 10 percent from time zero to about one T-time. For D/T ratios of 8 and 12 (i.e., slower market growth), the defender's sales remain essentially flat for nearly one T-time. Regardless of the growth rate of the market, however, after about one T-time, the defender's sales go into a sharp decline.

The point is, it is futile for the defender to watch for some "triggering" event that brings about the sharp decline in the defender's sales. There is no such trigger event. The sharp decline is built into the mathematics of exponentially growing market size and Pearl curve takeover by the challenger. The sharp decline is already inevitable at time $t = 0$, unless the defender does something to derail the challenger's Pearl curve of growing market share.

The problem, however, is that the defender may not be aware of the nature of the interaction between growing market sizes and declining market share. The flat or slightly growing sales of the defender during the time when the challenger still has less than half the market (growth of the Fisher-Pry ratio ranging from 0.01 to 1) may mislead the defender into thinking there is no problem. It is essential for the defender to recognize that the sharp decline in sales shown in the figure is inevitable. Once the challenger has reached about 1 percent of the market, the defender has only about one T-time left to do something to meet the challenge. After that it will be too late. Sales will remain flat, then suddenly "drop off the edge of the table," apparently without warning. In two T-times, half the market will be gone.

This finding of technological forecasters about the relation between growing market share size and declining market share sheds considerable light on the failure of firms involved with a defender technology under attack by a challenger. Now that the finding is available, however, firms involved with a defender technical approach have little excuse for being caught by surprise.

15.5 Anticipating New Technologies

A critical task of business managers is to determine when the technology on which their business is based will change. A recent example is found in the residential air conditioning industry. Residential air conditioners traditionally used reciprocating compressors to liquefy the heat transfer fluid after its heat had been exhausted. Figure 15.2 shows the growth in *seasonal energy-efficiency rating* (SEER) of these compressors from 1981 through 1990. A Gompertz curve fitted to the data shows that SEER will grow only slowly through the 1990s. Clearly this technical approach was reaching its limit.

A further complication was the imposition of a federal minimum SEER requirement of at least 10 to take effect in 1992. It is clear that achieving this regulatory requirement under the normal course of development of reciprocating compressors would be costly. Moreover, even if it were achieved there would be little opportunity for improvement. The Copeland Corporation, a major supplier of these compressors, recognized the need to replace their technical approach to compressors with one not only capable of getting ahead of its

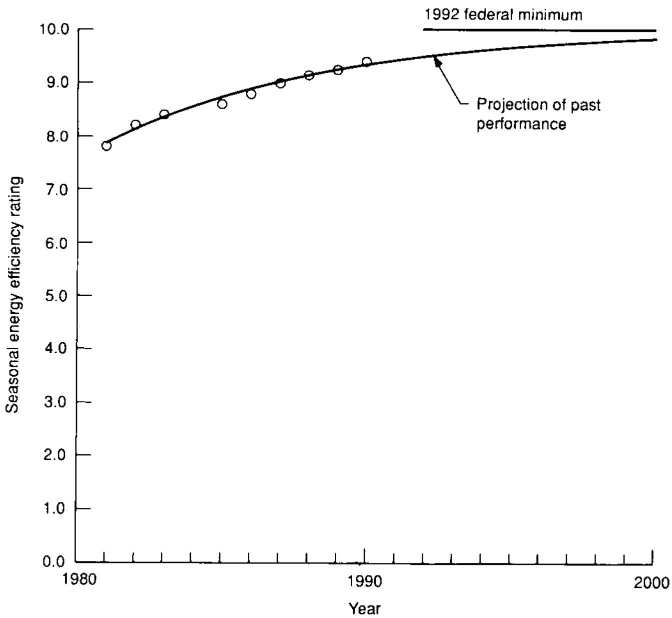

Figure 15.2 Efficiency of air conditioner reciprocating compressors.

competition but of meeting the regulatory requirements in a cost-effective way. Beginning in 1980, Copeland Corporation began experimenting with a completely new type of compressor, the compliant scroll compressor. In principle, it promised to achieve higher efficiency than the reciprocating compressor. Field tests began in 1984. By 1987, compliant scroll compressors had achieved efficiency and cost levels making them attractive in systems required to meet the federal minimum standards for 1992.

The new technical approach not only met the Federal standards, it had several other advantages: low noise and vibration, high liquid tolerance (liquefied heat transfer fluid), self-compensation for wear, high-speed capability, and greater durability than the reciprocating compressor. The result of this jump to a new technical approach was industry leadership for Copeland Corporation and a dominant position in the industry.

While the federal regulatory requirement placed immediate pressure on Copeland Corporation to switch to a new technical approach, competitive pressures would have made the switch inevitable eventually. Reciprocating compressors were reaching the limit of their inherent efficiency. Even without regulatory pressure, some firm in the industry would have seen a payoff from adopting a more efficient technical approach. By anticipating this payoff and recognizing that the existing technical approach was approaching a dead end, Copeland Corporation was able to get a jump on the rest of its industry.

This particular example illustrates what business firms should be doing on a regular basis. They should be constantly scrutinizing the technology on which their business is based. As this technology appears to be reaching a limit or is threatened by a replacement, they should be planning to adopt the new technology or otherwise make appropriate changes in their firm. Forecasts of their own technology, and possibly competing technologies, can tell them how much time they yet have to make a change. As the Copeland Corporation example shows, lead time of as much as 10 years may be required; hence, managers should be looking at least that far ahead.

15.6 R&D Strategy

As we observed in Chap. 9, the rank-time product for lag time from invention to innovation is approximately constant for any industry. This means that given any cluster of inventions available at a particular time, some will have a long gestation period before they can be marketed, a period which is characteristic of the industry, but most will have much shorter gestation periods.

From the standpoint of an individual business firm, this fact emphasizes the undesirability of placing all one's eggs in one basket. It may not be possible at the outset to identify which inventions will turn out to have the long gestation period. Therefore, the firm should back several "breakthrough" inventions, eliminating or reducing support for those that turn out to be either complete losers or winners only in the long term. By identifying those that can be expected to have a comparatively short gestation period for the industry, the firm can obtain a quicker payoff for its efforts.

Note that this advice flies in the face of the desire to avoid "duplication" in R&D. This advice actually encourages duplication, up to the point where enough information is obtained that one or more of the inventions under consideration can be dropped. Switching metaphors, the R&D program should be looked upon as a horse race, with horses eliminated only after they have clearly fallen behind the pack. An attempt to pick the winner beforehand, and back it only, will almost certainly lead to backing a loser, followed by a scramble to catch up.

The role of technological forecasting in this R&D strategy is to give advanced warning that a new technical approach will be required or will be available. When a new technical approach arrives, it will almost certainly come from outside the industry, as the electronic camera came from outside the conventional photochemical film industry. Hence, the forecasting methods employed as part of this R&D strategy should include environmental scanning, as described in Chap. 10.

15.7 Summary

Failure to adapt to technological change can destroy a firm or at best significantly reduce its market share. The demise of a firm that fails to adapt to technological change has some social value, of course. It transfers resources out of the control of those who have demonstrated incompetence into the control of those who have been proved to be competent. Nevertheless, the demise of a firm can bring hardship to a great many people. It would be better if firms adapted to new technology, instead of going out of business and letting other firms buy up their resources at fire-sale prices.

The role of technological forecasting in business planning is to prevent damage to firms by allowing them to anticipate technological change while they can still adapt easily. By forecasting technological threats and opportunities related to the major aspects of its business, management, and support activities, a firm can counter threats and capitalize on opportunities. Its managers will thereby demonstrate their competence and justify their continued control of resources. In addition, they will protect the interests of their stockholders, workers, and suppliers, all of whom might otherwise be hurt by technological change.

Problems

1 Using the data on surface and deep mining of coal in Table A.20, what forecast would you have made in 1950 for a company manufacturing deep-mine machinery? For a railroad whose major business was hauling coal?

2 Using the data on consumption of synthetic and natural fibers in Table A.48, what forecast would you have made in 1960 for a company producing cotton fiber? For a company producing knitting machinery? For a company producing men's dress shirts?

3 Using the data on tractive effort of locomotives in Table A.46, what forecast would you have made in 1940 for a company manufacturing steam locomotives? In 1945? In 1940 for a railroad company whose major business was hauling coal?

Chapter

16

Technological Forecasting in Government Planning

16.1 Introduction

Some activities of governments are very similar to the activities of other organizations. Planning for mission-oriented *research and development* (R&D) in a government organization is hardly, if at all, different from R&D planning in business firms. Similarly, when a government operates a business-type activity such as a post office, the planning required is very much like that in any other business. Hence, there is no need to discuss those activities of governments that are similar to the activities treated in the preceding chapters.

This chapter considers several types of activity that are unique to governments and, therefore, present planning problems different from those of other types of organizations. The role of technological forecasting in these activities is therefore unique and deserves specific treatment. The activities treated in this chapter are national defense, internal operations of government, the regulation of economic and social activity, and the provision of public goods.

16.2 National Defense

Much of the work in technological forecasting since 1945 was pioneered in the U.S. Department of Defense. It would seem almost redundant to say much about this topic. Nevertheless, the world situation after 1990 is so much different from the preceding four decades that a look at the role of technological forecasting for national defense is in order.

One of the most consequential applications of technological forecasting to U.S. national defense since 1945 involved the Polaris missile, intended to be carried by nuclear submarines and launched from below the surface of the sea. The initial design was intended to carry a physically large warhead of the type then in production. The warhead size drove the entire design of the missile, which would have been quite large, and in turn would have driven the design of the entire submarine. During the development program, the designers were provided with a forecast of lighter-weight warheads that would be available by the time the missile was deployed. The design was revised to take advantage of the lighter warhead. As a result, the missile that was finally developed was much smaller than originally planned. This size reduction rippled through the design of the submarine as well. A submarine of a given size could carry more missiles. By accepting the forecast and basing the design on it, the designers of the Polaris missile and the Polaris submarines were able to achieve a very capable weapon system at the earliest date at which it was technologically feasible.

This example illustrates one role of technological forecasting for national defense. If a new weapon system is to be developed, forecasts of the armament, the components, and the structural material to be available at the time of deployment can allow the design to take advantage of these developments. The weapon designers can, thus, escape to some extent the problem of a weapon that is already obsolete when it is built.

Moreover, forecasts of technology that will be available during the operational lifetime of the weapon system can be used to allow what are known as "preplanned product improvements," i.e., at the time the weapon system is deployed, provision has already been made to incorporate improved components or materials that are forecast to be available before the system is retired from service.

The whole point of national defense is to protect against foreign threats. Hence, an important part of national defense planning is to anticipate the technological capability of potential enemies. Here is another important role for technological forecasting.

During the so-called Cold War, technological forecasts of Soviet weapons played an important role in U.S. defense planning. What would be the performance of the next generation of Soviet aircraft? Missiles? Submarines? This role will continue to be important in the 1990s and beyond. The only major difference will be that the identity of the potential enemy may not be so clear. Defense planning in the United States and in other major industrial powers will have to anticipate technological advances in several countries, not just in a single one. Thus, forecasts of foreign-developed military technology will continue to be needed for planning one's own defense R&D and force structure.

One major difference between defense planning in the 1990s and that in the Cold War era is the much greater possibility of technology transfer to smaller nations. As a matter of policy, the Soviet Union never transferred its latest technology to other members of the Warsaw Pact nor did the Soviet Union sell it to its clients. This situation is clearly changing. Already by 1991, the latest Soviet military technology was available on the world arms market. The same was true of military technology from China. This is an extension of a trend that was already developing in the 1970s and after: Technology from France, Germany, England, Sweden, Austria, and other industrial powers was available on the world arms market. The result is that many nations will have military capabilities that they are not capable of producing for themselves. An important role for technological forecasting is to estimate the rate at which nonindustrial nations can absorb and use military technology from more advanced nations.

The pre-1990 Soviet Union exemplifies another situation that may become more common. From the time of the Bolshevik Revolution, the Soviet Union maintained a hypertrophied military sector in its economy, in parallel with a grossly underdeveloped civilian sector. Other nations may be capable of doing the same thing. Iraq, for instance, was by late 1990 within a year or two of producing nuclear weapons, this despite the fact that the civilian sector of its economy was well behind the standards of industrial nations. An important role for technological forecasting, then, is to estimate when specific nonindustrial nations might develop indigenous capabilities for particular kinds of weapons, either with or without technology transfer. The probabilistic distribution of time lags, discussed in Chap. 9, may be especially useful in this application.

In summary, the role of technological forecasting in national defense planning involves both determining what kinds of equipment one's own forces might be able to acquire and what kinds of equipment potential enemies might have at some time in the future. After 1990 the world situation requires that the range of technological capabilities of both industrial nations and still-developing nations must be forecast.

16.3 Internal Operations of Government

Most government decision makers find that the objectives of their organization are established by law and are, therefore, beyond their control. Their budgets are determined by the legislature, and even the sizes of their work forces are fixed by factors over which they have little control. In many ways, therefore, their problems are different from those of private decision makers.

However, government decision makers are still faced with technological changes that may alter the internal operations of their organization. Thus, even though their problems differ from those of private decision makers, there is still a need to utilize technological forecasts to anticipate these changes. We will consider two issues in which technological change may affect the internal operations of a government organization: centralization versus decentralization and the skills of the work force.

All government organizations are faced with the issue of the proper degree of centralization or decentralization. The advantages of decentralization are flexibility of operations, responsiveness to local conditions, and ease of access by the citizens who must deal with the organization. The advantages of centralization are reduced operating costs, more uniform policies, and ease of access to high-level decision makers and to central files. Any organizational structure represents a balance between centralization and decentralization. In a well-run organization, this balance reflects the relative costs of centralization and decentralization.

Technological change can alter the relative costs and, thereby, allow a shift in the balance between centralization and decentralization. Communications, for instance, can make it possible for remote offices to refer the "hard cases" to higher-level decision makers. This allows some of the benefits of decentralization while still retaining some of the benefits of centralization. Copying machines allow the replication of central files in remote locations, with frequent updates. This allows more decentralization, while retaining one of the benefits of centralization. A central computer can significantly reduce operating costs, thereby shifting the desired balance toward centralization. Remote access to that computer, however, can shift the balance back toward decentralization.

Planning for the future of an organization, in terms of location, equipment needed, etc., should be based on the technology expected during the period being planned for. To make commitments to specific locations and select personnel on the assumption of a static technology simply means that the organization will not have the best balance between decentralized and centralized operations. The agency will operate at costs higher than necessary and will not meet its legally required objectives as well as possible.

Decision makers in government organizations, thus, should use forecasts of the technology that will affect their operations, such as communications, record keeping, filing, document transmitting, and copying. These forecasts can help determine the proper balance between centralization and decentralization. If new technology permits a shift in the balance, planning for that shift ahead of time can allow the organization to carry on its legal responsibilities at a minimum cost.

Regardless of the effect on the balance between centralization and decentralization, a change in technology may require different skills on the part of the work force. Some instances are obvious: The adoption of new equipment requires that users or operators have the skills required to utilize the equipment. However, more subtle effects are also possible. Changes in equipment in one part of the organization may require changes in skills in other parts as well. The installation of a computer, for instance, requires the addition of programmers, computer operators, and other workers with skills directly related to making the computer function. However, people in other parts of the organization may also need to acquire new skills. Their previous decision-making procedures may have been shaped by the kind of information that was available in the agency's files and the ways in which that information could be extracted. The availability of a computer means that information can be extracted in ways that were not possible before. Both the depth and breadth of the information potentially available may increase. New skills will then be required if the decision maker is to take advantage of this information.

The government agency planner is faced with making some combination of three choices when technological change occurs: hire people with the newly required skills, train existing workers to have the newly required skills, and find other uses for workers whose skills are no longer required. The planner is often severely constrained in these choices by laws and regulations intended to protect the agency worker against arbitrary action. A forecast of technology that will require the availability of skills within the agency can allow the change to be made on an orderly basis, within the legal protection provided to workers, while maintaining both effectiveness and efficiency.

16.4 Regulatory Agencies

There are two different aspects to regulation by government agencies. One type of regulation is focused on a specific industry. It concerns itself with economics, i.e., the rate of return to companies in the industry, prices charged by the industry, entry into the industry, and operations of the industry as they affect services provided to the customers. The other type of regulation is concerned with the safety of products, of working conditions, and of third parties not involved in the industry. This second type may or may not be limited to a specific industry. The Food and Drug Administration is concerned with safety, rather than rate of return, but is restricted to particular industries. The Occupational Safety and Health Administration, on the other hand, is concerned with a single subject, workplace safety, but in all industries. Planning for each of these types of regulation requires technological forecasting. However, since the intentions of the two are different, they must be treated separately.

A major problem in the economic regulation of industry has been the failure of regulatory agencies to adapt to technological change. Examples are abundant. The Interstate Commerce Commission for years hindered the growth of "piggy-back" transportation of truck trailers on trains because of the potential threat to over-the-road trucking. Piggy-backing would have saved energy, reduced congestion and wear on the highways, and provided lower costs to customers. These advantages were not considered important, however, by comparison with the need perceived by the regulators to keep the trucking and railroad industries separate. The Federal Communications Commission restricted the growth of cable television for years because of a perceived threat to commercial television networks. Many similar examples could be cited.

The central feature in these examples is the attempt made by regulatory agencies to freeze industries into a pattern based on a particular set of technologies. This attempt is aggravated by the concept that competition must mean competition between firms using different technologies. For instance, transportation regulators assumed that competition meant that trucking companies, railroad companies, and barge companies must compete against one another. The concept of competition among integrated transportation companies, each providing a mixture of transportation modes, was never considered.

Because of this insistence on forcing regulated firms to restrict themselves to single technologies, regulatory agencies have a serious problem. They must take actions to prevent the extinction of firms that have been trapped into obsolete technologies by their own regulations.

Technological forecasts could be used to anticipate technological changes that will alter the competitive posture of firms required to utilize only a single technology. Currently, when a technological change threatens to alter the competitive situation, regulators tend to prohibit it or restrict its introduction and diffusion as in the examples cited above. It would be far more beneficial to society at large if these technological changes could be adopted rapidly. This would require that regulatory agencies anticipate the changes and their impacts, and tailor their regulations so the affected firms can adapt rapidly. The regulatory agency may even need to consider allowing a restructuring of the regulated industry instead of trying to force the new technology into a pattern based on earlier technologies. Such a change would benefit the public as well as the firms involved. Here again, technological forecasts could be of significant use in early planning for restructuring. Forecasts identifying the technological changes to be expected can provide the basis for an orderly change instead of a crisis.

The advent of cellular radios shows one way in which this can be done. The Federal Communications Commission decided to license two cellular radio companies in each major metropolitan area, to provide competition. However, one license would be reserved for the local telephone company, allowing it to enter the new business instead of freezing it into the old wired telephony while allowing its competitors to use the new technology. One might argue that a government-mandated monopoly on wired telephones is a bad idea; there are cities in the United States where two telephone companies successfully compete with each other, providing good service at competitive rates without going bankrupt. Nevertheless, so long as we are going to have a government-mandated monopoly in telephone service, the government's regulatory agencies should not freeze the regulated firms into obsolescing technology.

Health and safety regulations involve two types of problems. The first is anticipating threats to health and safety. The second is measuring them. Technological changes may alter the nature of the threats as well as improving people's ability to measure and identify them.

Despite much of the concern voiced about threats to health and safety from "new" technology, the fact remains that most of the threats the public faces today are from the massive use of comparatively old technologies, such as coal-burning, petroleum-refining, and manufacture of commodity industrial chemicals such as vinyl chloride and phosgene. Nevertheless, it makes sense to anticipate threats to health and safety when possible. In particular, given the limited resources available to agencies responsible for health and safety regulations, it may make more sense for them to get a head start on threats posed by new technologies than to expend their resources on the threats from old or obsolescing technologies. For instance, suppose a new industrial process is introduced that is expected to replace an existing process through normal substitution. This new process may expose workers to chemicals not currently present in the workplace. A regulatory agency that has a forecast of the substitution might begin investigating the health hazards posed by the new process. This would have two advantages. First, it might make it possible to alter the process, at low cost, to eliminate or reduce the hazard. Second, if the process cannot be altered, precautions can be taken against worker exposure before any harm is done. It may well make more sense to get a head start on the new problem and eliminate or control it before it becomes serious than to spend resources protecting workers against the problems of the declining technology while ignoring the new technology until serious damage is done. By failing to take advantage of forecasts of new technology and investigating potential hazards, the regulatory agency is always spending its efforts dealing with large but declining problems, allowing small problems to grow into large ones owing to lack of advanced planning.

Forecasts of the technology for measuring and identifying threats can also be of value to agencies regulating health and safety. As an extreme example, consider a regulation or law that prohibits the presence of a dangerous substance at any level whatsoever. (The Delaney Amendment, which bans food or drugs containing any detectable contamination by a known carcinogen, is a real-life example.) Below some level of contamination, the presence of the contaminant may be irrelevant, in the sense that the additional risk it poses is insignificant, or is offset by the benefits of legitimate use of the chemical. However, as the technology of detection improves, even these inconsequential levels of contamination will eventually be detected, and a "zero-tolerance" law or regulation will require the banning of a product that is actually harmless or even beneficial. Regulations on the levels of allowable contaminants should be based on known hazard levels and the capabilities of both contemporary and future detection technology.

Aside from the problem of zero-tolerance regulations, there is still the issue of tailoring the regulation of exposure and contamination levels to detection capabilities. Forecasts of detection technology can be of great use in determining what is worth regulating and how. For instance, if a contaminant cannot be measured accurately even at a level that is unsafe, elaborate and expensive precautions against contamination may be the only possible measure of protection. Once detection technology is adequate, however, emphasis can be shifted from prevention, which might approach "overkill," to actual measurement, with the screening out of those items actually contaminated. Thus, while the initial regulatory approach may emphasize manufacturing procedures to guarantee against contamination, technological forecasts can be used to determine when the emphasis should start shifting to actual measurement.

Whether regulatory agencies are concerned with economics or with health and safety, the nature of the problems they face will be altered by technological change. Forecasts of technological change can make it easier for these agencies to be on top of upcoming problems before they become serious.

16.5 Public Goods

One function of government is to provide so-called public goods. These are goods (or services) that by their very nature cannot be made available only to specific individuals but, if made available at all, must be available to everyone. There are three general categories of such public goods: protection such as law enforcement and national defense; infrastructure such as highways; and safety facilities such as lighthouses and air traffic control. We will discuss the use of technological forecasting in each of these categories.

The application of technological forecasting to national defense is widespread but has already been discussed. However, law enforcement is an important area of public protection in which technological forecasting is not as widely used as would be desirable. There are two ways in which technological change affects law enforcement. The first is the creation of new kinds of crime. The second is the creation of new ways of preventing or detecting crime.

Many kinds of technological change increase the criminal's scope of activity. The mobility provided by the automobile certainly aided the bank robber during the 1920s and 1930s. The adoption of telecommunications by the police eventually offset this mobility to some extent. However, the computer has created new opportunities for the bank robber, some involving the direct alteration of bank records, others involving the penetration of data links between banks and the insertion of false information. The importance of computer crime is revealed by the fact that the average nonelectronic embezzlement amounts to about $23,000, while the average computer fraud amounts to over $400,000.

It may be difficult to forecast just how criminals will exploit a new technology. If the manner of exploitation were obvious, it would often be designed out of the technology in the first place. The vulnerability of computers in banks is largely due to the failure to consider the possibility of computer theft. However, forecasts of technological change can often be the first step in identifying the potential for new crimes.

A forecast of the use of computers in banks, for instance, might have been the first step in predicting the scope of computer crime and in planning measures against it. Thus, while a technological forecast may not provide much information about the exact nature of future crime, it can provide information about what new technology might be vulnerable to criminal exploitation or available to aid criminals.

Much of the thought given to technological measures against crime focuses on so-called scientific crime detection. This generally involves technical means for connecting pieces of physical evidence with specific individuals. Fingerprints, blood typing, automobile paint matching, and so forth were long widely used for such purposes. Technological change can make new techniques available, such as the DNA typing that has come into use in the 1980s. In addition to making new techniques available, technological change can improve the capabilities of old techniques. The increased sensitivity and accuracy of measuring instruments will continue to contribute to this important aspect of law enforcement. Forecasts of detection and measurement technology can help police officials plan to keep their crime laboratories up to date.

Other technological changes also aid law enforcement. The use of telecommunications has already been mentioned. Computer files of criminal characteristics and behavior are becoming widespread. Burglar alarms of varying degrees of sophistication are widely available. The use of radar (and now laser radar) to detect speeders is now commonplace. Forecasts of these other technologies that aid law enforcement can be used to plan the acquisition of necessary equipment, identify new skills in which law enforcement personnel must be trained, and identify necessary alterations of law to allow new types of evidence or to protect the public from the unwarranted use of a new technology.

Governments often provide social infrastructure such as highways, flood-control dams, bridges, and tunnels. They also take responsibility for dredging rivers and harbors, digging canals, and maintaining waterways. This infrastructure is presumed to provide benefits to the public at large, as well as to the direct users. Land near highways, for instance, often increases in value because of the availability of better transportation.

As with any investment, the planners and builders of infrastructure may make either of two mistakes. They may overbuild or they may underbuild. Either error results in unnecessary costs to the public as well as to direct users. In a competitive business, errors committed by a single company are not particularly serious. If a company underinvests, some other company may take up the slack; if a company overinvests, the owners lose their own money, not anyone else's. With publicly funded infrastructure, however, the public at large pays for the mistakes of the planners. Thus, it is extremely important that the planners responsible for selecting the level of investment in infrastructure take technological change into account. Such change may alter the level or type of investment needed.

Lighthouses are a classic example of safety facilities that allegedly must be built by governments. It is allegedly impossible for private lighthouse operators to collect fees from the ships that use them. We need not examine the issue of whether this argument is correct. The fact is that governments do build lighthouses, air traffic control systems, harbor traffic control systems, etc. The planners for these facilities are faced with two issues: What functions must these facilities serve, and what technology will be available to serve them? A look at the history of aerial navigation will serve to illustrate these two issues.

Initially, commercial aircraft flew only in the daytime because of the difficulty of navigating at night. The first nighttime navigation aids were beacon lights, similar to the lighthouses long used at sea. Chains of these beacons were established along routes between major cities; however, they were ultimately unsatisfactory because they could not be seen by aircraft flying above clouds in bad weather, i.e., when they were most needed, they were not usable. The next major step was the "radio beam," an arrangement under which an aircraft could follow a specific track in the sky by properly matching signals from several transmitters. These beams in effect laid out aerial routes that could be followed despite bad weather or clouds. The problem with these routes was that they were narrow, leaving most of the sky unused. Eventually the level of traffic grew to the point that the system became saturated. Two later developments, radar and directional radio beacons, eliminated the need to have all aircraft follow the same narrow tracks.

In this example, we see that the function to be served changed over the years. The air traffic control system first had to meet the needs of individual aircraft flying on clear nights (they could not fly at night in bad weather anyway). Later it had to meet the needs of aircraft flying in bad weather and, ultimately, the needs of large numbers of aircraft flying in clouds or bad weather. The system was originally intended to allow aircraft to find their destinations. Later its function was expanded to that of keeping aircraft from colliding with each other. Thus, the function to be served became more difficult to carry out and broader in scope.

Similarly, the technology changed over the years. Visual means of navigation were abandoned for electronic means, and the electronic means, in turn, were refined and extended in performance. This refinement is still going on, with computer tracking replacing manual tracking of aircraft on their assigned routes.

The planners of a network of safety equipment such as an air traffic control system need forecasts of the technology available to the users of the system. For instance, forecasts of aircraft speeds, altitudes, and numbers define the performance that will be required of an air traffic control system. Forecasts of the technology that will be available for the system, on the other hand, define the system's possible performance. Technological forecasts can, thus, be used to avoid building systems that will not meet the needs of the users (e.g., chains of lighted beacons that cannot be seen from above the clouds) and to start the installation of systems that will meet the users' needs (e.g., a system that is usable above a certain altitude). Similarly, forecasts of capability will allow planners to avoid excessive investment in obsolete equipment when something else will soon be available.

16.6 Summary

Changes in technology can lead to changes in the operations and organization of government. The types of decision undertaken by government planners in response to changes in technology are somewhat different from those facing decision makers in nongovernmental organizations. Despite this difference, the use of forecasts of technological change can provide the same benefits as in other types of organizations. When changes are identified in advance, it is possible to adapt to them smoothly and routinely. Thus, planners and decision makers in government should make use of the forecasts of those technologies that will affect either their activities or the manner in which they are to carry them out.

Problems

1 Using one of the measures of performance for fighter aircraft developed in Chap. 6, prepare a forecast of U.S. fighter aircraft performance for the year 2005. How might the implicit trade-offs be taken (e.g., shorter takeoff roll, longer range, greater payload, and higher speed)? What implications does this forecast have for military R&D? For design of air bases? For location of air bases? For base-rights treaties with foreign countries?

2 Assume the year is 1975. Using data from Table A.79, prepare a forecast of the use of computers by municipal police departments. What are the implications of your forecast for personnel requirements for police departments? For police officer training? For changes in police procedure?

3 Assume the year is 1945. Prior to World War II, ammonia had been produced through a process using coke. The coke production itself generated air pollution, and the coke-based ammonia process generated additional pollution. A single plant using natural gas instead of coke had been built in 1931, but the process had not caught on

in the fertilizer industry. However, after 1941, more ammonia plants using natural gas were built, and by 1945 it appeared as though the natural gas process would displace the coke process. Using data from Table A.78 for the years 1931 through 1945, prepare a forecast of total ammonia production, ammonia production using coke, and ammonia production using natural gas. What implications would your forecasts have for an agency concerned with regulating air pollution?

4 Using data from Table A.23, prepare a forecast of the number of automobiles in the United States. What are the implications of your forecast for highway construction? For traffic safety? (Note: you have two alternative approaches to preparing this forecast. You might try to project the number of automobiles directly, or you might project adoption of automobiles (automobiles per capita) and make a separate projection of population. Which do you think is the better approach?)

Chapter 17

Technology Assessment

17.1 Introduction

There have been numerous definitions of the term *technology assessment* (TA). One of the best is the following definition by Joseph F. Coates (1974):

> [TA is] a class of policy studies which systematically examine the effects on society that may occur when a technology is introduced, extended, or modified. It emphasizes those consequences that are unintended, indirect, or delayed.

The idea behind TA is that technology does affect society, and its effects should be examined before it is deployed. Indeed, the very purpose of technology is change—to make something cheaper, easier, or simpler. To the extent that society is built around an older technology by which things are costlier, harder, or more complex, technology is bound to bring social change.

However, we must guard against the Marxian fallacy that there is a deterministic relationship between technology and society, or as Marx expressed it, "The hand mill gives you society with the feudal lord; the steam mill gives you society with the capitalist." This statement was specifically wrong. The English census of 1086 A.D. counted a total of over 5000 water-driven mills in England, at a time when England was a feudal rather than a capitalistic society. Conversely, many mills owned by capitalists at the beginning of the industrial revolution were driven by hand or by wind power before steam engines were adopted. Moreover, the hand mill itself can be traced back thousands of years to ancient Sumer, a society that can best be described as socialist (the government owned all land and tools, including farm implements, and issued regular food rations to everyone including farmers) (Shafarevich, 1980). Thus, there is no rigorous deterministic link between the nature of a technology and that of the society using it.

Nevertheless, technology does bring social change. The impetus for TA arises, then, not from the fact that technology has social consequences, but because some of these consequences are unforeseen and unintentional.

17.2 Some Historical Cases

Even though there is no deterministic link between technology and social change, new technology has often brought about social changes that were unanticipated and unintended by the sponsors of the technology. Here we will review the consequences of the automobile, the railroad, and the mechanization of agriculture. To provide a framework for analysis, we will utilize the dimensions of the environment introduced in Chap. 3. Not all the dimensions will be relevant in each case, but using them in a checklist provides a means of assuring completeness in the search for consequences, as well as an analytical framework that allows for comparison between different technologies.

The automobile. The history of the modern automobile can be traced to the German inventors Carl Benz and Gottlieb Daimler, who independently devised vehicles powered by gasoline engines. The basic design of the modern automobile was essentially fixed in 1891. In that year, E. C. Levassor, of the French firm of Panhard-Levassor, designed an automobile using Daimler's patents.

Early inventors of the automobile, such as Benz, visualized it as completely replacing the horse and revolutionizing transportation. However, for the first decade and a half of their existence, automobiles were primarily toys for the wealthy. In 1908, Henry Ford brought out the Model T, which was the first successful mass-produced automobile. While Ford is often praised for introducing mass production to the automobile industry, this was not an end in itself. Ford's basic goal was an automobile that would be cheap enough to be owned by the average person, rugged enough to operate successfully on the primitive roads of the day, simple enough for anyone to operate, and inexpensive to maintain. So Ford set out to realize the original dreams of Daimler and Benz for the automobile, i.e., he deliberately set out to produce social change.

The automobile did in fact achieve the social change Ford and the earlier inventors hoped for. It replaced the horse on a worldwide basis. However, it had many consequences beyond this first-order, intentional one. The secondary consequences, particularly the unintentional and unanticipated ones, are the concern of TA; hence, a few are worth examining.

In the technological dimension, it might be expected that widespread use of the automobile would lead to inventions directly related to its manufacture and use. However, as shown in Table 17.1, an enormous increase in the rate of inventions in the petroleum industry can be attributed directly to the need for new technology in producing gasoline from crude oil. Another aspect of the technological dimension is that from about 1960 on, the automobile represented about 25 percent of all American energy consumption (including the energy contained in asphalt paving). Some aspects of the economic dimension are shown in Tables 17.2 and 17.3. Table 17.2 summarizes the size of the au-

TABLE 17.1 Petroleum-Refining Patents Issued

Year	Number of patents	Year	Number of patents
1860	10	1915	91
1870	24	1920	248
1880	13	1925	831
1890	20	1930	646
1900	18	1940	461
1910	38	1950	237

TABLE 17.2 Economic Impact of Motor Vehicles in 1987

	Establishments, thousands	Sales, $ million	Payroll, $ million	Employees, thousands
Auto dealers	103	333,420	28,688	1373
Gas stations	115	101,997	6414	702
Automobile repair shops	114.6	28,664	7727	485.6
Automobile rental	11.4	16,441	2366	134.6
Automobile parking	9.3	2639	491	45.7
Automobiles and parts manufacturing	4422	204,678.5	22,864.1	748.2

SOURCE: U.S. Bureau of the Census, *Statistical Abstract of the United States: 1990* (110th edition) Washington, D.C., 1990.

TABLE 17.3 Automobile Insurance

Year	Premiums*	Losses Paid*
1950	2625	1069
1955	4644	2122
1960	6448	3645
1965	8358	5221
1970	14,612	11,198
1975	23,860	19,657
1980	37,621	27,887
1985	57,601	49,156

*In millions of dollars.

TABLE 17.4 Automobile Accidents

Year	Accidents	Injuries	Deaths
1930	9,859,000	1,373,000	32,900
1940	10,339,000	1,440,000	34,500
1950	10,418,000	1,799,000	34,800
1960	11,429,000	3,078,000	38,100
1970	22,116,000	4,983,000	54,800
1980	18,100,000	2,000,000	51,700

tomobile industry in terms of establishments, sales, payrolls, and employees. Table 17.3 shows insurance premiums collected and losses paid for automobile accidents. Some aspects of the social dimension of the impact are shown in three additional tables. Table A.23 presents the number of automobile registrations per year. By 1987, there was not quite one passenger car for every two persons in the United States. Table 17.4 shows the numbers of automobile accidents, injuries, and deaths per year. Table 17.5 shows the extent of automobile travel in the United States. The political dimension of the automobile's impact can be seen in part from Table 17.6, which shows state and local highway debt. This debt averaged between one-tenth and one-fifth of the total state and local debt for the period shown. The ecological dimension of the automobile's impact is shown by Table 17.7, which shows the sources of air pol-

TABLE 17.5 Automobile Travel

Year	Distance, billion vehicle miles
1960	588
1965	712
1970	901
1975	1051
1979	1163
1985	1353

TABLE 17.6 State and Local Highway Debt

Year	Debt, $ million
1960	13,166
1965	15,316
1970	19,107
1975	23,801
1980	25,804
1985	31,790

TABLE 17.7 Air Pollution Sources

Source	Percent*
Transportation	42.3
Stationary	21.4
Industrial	13.7
Solid Waste	5.2
Miscellaneous	17.4

*Data are for the year 1970.

lution in the late 1960s, prior to efforts to reduce pollution from automobiles. These dimensions could be examined in more detail and information from other dimensions could be added, but these examples are sufficient to illustrate that the automobile had impacts on society far beyond those intended or even imagined by its inventors and promoters. These unintended impacts are the concern of TA.

Railroads. The intent of the inventors of the railroad and of the people who built them was to provide cheap and reliable transportation. This they certainly achieved. However, the railroads also had consequences that were unforeseen and unintended by their inventors and sponsors. We will briefly examine some of these.

The technological dimension of the impact of the railroad is indicated by Table 17.8, which shows the number of railroad patents issued from 1837 to 1950. Railroads drew the attention of a significant fraction of the nation's in-

TABLE 17.8 Railroad Patents Issued

Year	Number of patents: Track	Nontrack	Total
1837	4	11	15
1840	4	21	25
1850	3	27	30
1870	48	278	326
1880	45	531	576
1890	242	1526	1768
1900	133	954	1087
1910	453	1438	1891
1920	284	1147	1431
1930	192	1038	1230
1940	53	602	655
1950	39	363	402

ventive talent. The economic dimension is shown in Table 17.9, which shows the gross capital formation for railroads and for all U.S. industry. During the era of railroad construction, the railroads drew on a significant fraction of the total U.S. investment capital. The social dimension of the impact of the railroad may have been its most important impact. In a very real sense, railroads provided the threads that held the United States together. Even in George Washington's time, there were separatist movements in the trans-Allegheny West. For the nation to survive, it had to be possible for trade and travel to be conducted over long distances and at tolerable cost. When agricultural or mineral commodities were hauled by wagon, the transportation costs ate up the profits for distances over about 150 miles. First the canals and then the rail-

TABLE 17.9 Average Gross Capital Formation

Year	Total U.S. industry*	U.S. railroads*	Railroads as percent of total
1873	1.47	0.26	17.7
1878	1.72	0.15	8.7
1883	2.18	0.59	27.1
1888	2.56	0.24	9.4
1893	3.06	0.27	8.8
1898	3.48	0.11	3.2
1903	4.91	0.24	4.9
1908	6.42	0.59	9.2
1913	6.94	0.54	7.8
1918	16.5	0.50	3.0
1923	16.2	0.80	4.9
1928	19.2	0.82	4.3
1933	6.68	0.20	3.0
1938	15.9	0.38	2.4
1943	28.7	0.58	2.0
1948	47.9	1.05	2.2

*In billions of dollars.

roads made it possible for the United States to exist as an economic and political entity. Without railroads, population and commerce would have been constricted to narrow ribbons along suitable waterways; separatist movements, centered around major river basins, would have been inevitable. But the railroad could go anywhere and haul goods more cheaply than competing means. Thus, the existence of a transcontinental United States may be one of the most important social impacts of the railroad. The cultural dimension of the railroad's impact was also important. During the era of railroad construction, newspapers and magazines were filled with accounts of railroad activities: projects, speed records of trains, profits of the companies, and accidents. Railroad lore appeared in poems, speeches, and songs (for some examples of the latter see Lyle, 1983). The "spirit of the times" was embodied in the image of the railroad—the Age of Steam.

The mechanization of agriculture. The intent of the inventors and manufacturers of agricultural machinery was to reduce the amount of human labor involved in producing farm products. They succeeded in their objectives, as shown in Table 17.10. The amount of labor required to produce certain crops has been reduced by a factor of 10 or more. However, there were other consequences to the mechanization of agriculture, some unintentional.

The economic dimension is illustrated in Table 17.11. As mechanization gained momentum in the late 1800s, the investment per farm increased dramatically; from 1850 to 1950 it grew by a factor of 20. But the social dimension of the impact is perhaps the strongest of the unintended consequences. Table 17.12 shows the migration to and from farms between 1921 and 1957. The mechanization of agriculture resulted in enormous numbers of people moving off farms and into towns and cities. The census data show that these migrants improved their condition as compared with their condition before migration; nevertheless, the migration represented a major social upheaval.

Conclusion. The brief review of these three technological changes show that technology does affect society. Moreover, many of the effects are unanticipated and unintentional. This is not to say that all unanticipated changes are bad. If

TABLE 17.10 Productivity of Farm Labor

Year	Wheat*	Corn*	Cotton†
1800	373	344	601
1840	233	276	439
1880	152	180	318
1900	108	147	280
1920	87	113	269
1940	47	83	191
1950	28	39	126

*Person-hours per 100 bushels.
†Person-hours per bale.

TABLE 17.11 Value of Farm Implements

Year	Value, $ millions	Farms, thousands	Average price per farm, $
1850	152	1449	104
1860	246	2044	120
1870	271	2660	102
1880	406	4009	95
1890	494	4565	108
1900	750	5737	131
1910	1265	6406	197
1920	3595	6518	550
1930	3302	6546	506
1940	3060	6350	482
1950	11,216	5648	2000

TABLE 17.12 Migration of Farm Population

Year	To farms*	From farms*	Net loss*
1921	560	896	336
1930	1604	2081	477
1940	819	1402	483
1950	995	2309	13,314
1957	459	2695	2236

*Thousand of migrants per year.

they are not bad, however, that is simply good luck, not good planning. These examples make the point that careful prior assessment of the changes brought about by new technology may help alleviate the undesirable impacts and gain more benefit from the desirable ones. Minimizing the unfavorable consequences of technological change is the ultimate objective of TA.

17.3 The Role of Technological Forecasting

There are many good books available on TA (Porter et al., 1991, is a recent example). This chapter cannot cover the subject in as much detail as these books. Instead, its purpose is to present the role of technological forecasting in TA. The other aspects of TA, important though they are, will not be treated here.

The examples in the previous section described the consequences of several past technological changes. However, TA is intended to identify in advance the effects of future technological change. Thus, technological forecasts play a key role in the conduct of a TA. The would-be technology assessor needs estimates of the technology's performance, the rapidity with which it will be adopted, and its ultimate scope of deployment. Without these estimates there is no way to estimate the magnitude of the consequences of the technology and even the nature of some of the consequences.

Forecasts of the performance of a technology are required to estimate its direct technical effects, such as its efficiency, the nature of its energy requirements and other inputs, the nature of its waste products, and other characteristics of the technology. Information of this type is precisely what technological forecasts can provide. Hence, before a TA can be completed, the assessor must make or obtain a forecast of the technical characteristics of the new technology.

The rapidity with which the new technology will be adopted determines the nature of some of its consequences. If the technology is likely to be adopted slowly, for instance, work-force adjustments can be made through normal retirements; however, users will not get the benefits quickly. If adoption is rapid, users and customers benefit quickly, but work-force adjustment may be more difficult. Forecasts of technological adoption or substitution, using the appropriate methods described in earlier chapters, are important inputs to the TA process.

The ultimate scope of deployment of the new technology determines the scope of its consequences. Even a technology that requires significant shifts in the relevant work force or that substitutes one type of raw material for another may have little overall effect if the total scope of deployment is small. Technological forecasts can be useful in estimating the scope of deployment by identifying those applications in which the new technology is likely to be superior to both improved versions of the current technology and other new technologies.

Thus, while technological forecasts do not address, for example, the social and economic consequences of a new technology, they provide the basis from which these consequences can be estimated. The role of technological forecasting in TA is to provide estimates of the extent and nature of the technological changes that will have social consequences.

17.4 An Example

Let us briefly examine some of the consequences of a new technology. In Chap. 8, a simulation model was presented for the growth of the use of solar energy in homes for space heating and hot water. The model focused only on the growth of the technology itself. Let us now expand that model to include some variables related to broader consequences of the technology. Let us add the following variables to the original model:

1. *Level of atmospheric pollution:* The current value is taken as 0.5 of its maximum and is given the plotting symbol "A."

2. *Fossil fuel use:* The current value is 0.7 of its maximum and is given the plotting symbol "U."

3. *Number of coal miners:* The current value is 0.8 of its maximum and is given the plotting symbol "M."

4. *Number of solar energy technicians:* This is taken as being 0.1 of its maximum value and is given the plotting symbol "T."

Next we consider the impacts on each of these variables. Because some variables have been added, some of the earlier impacts given in Chap. 8 have changed. Since the use of fossil fuel is now explicitly included, its price should no longer directly affect itself. Price should affect use, which in turn should affect price.

The proportion of homes using solar energy is affected positively by the price of fossil fuel, positively by the performance of solar devices, and positively by the availability of solar energy technicians.

The price of fossil fuel is affected negatively by the proportion of homes using solar energy, positively by the level of atmospheric pollution (pollution-control devices raise the cost of using fossil fuel), positively by fossil fuel use, and positively by the outside world (price will rise with increasing scarcity, even without the demand for home heating).

The performance of solar energy devices is affected positively by increased use and by solar energy R&D. R&D on solar energy devices is affected positively by the fraction of homes with solar energy and by the level of atmospheric pollution. Atmospheric pollution is affected negatively by the fraction of homes with solar energy and positively by fossil fuel use. Fossil fuel use is affected negatively by the fraction of homes with solar energy, negatively by the price of fossil fuel, negatively by atmospheric pollution, and positively by the outside world (it is used for other things besides home heating).

The number of coal miners is affected positively by fossil fuel use. The number of solar energy technicians is affected positively by the fraction of homes with solar energy and by R&D on solar energy.

The impacts among the variables are shown in Table 17.13, which should be compared with Table 8.4 to see the changes that have been made with the addition of the four extra variables. When the program is run, the results are as shown in Fig. 17.1, which should be compared with Fig. 8.5 to see the differences of the simpler model and the behavior of the additional variables.

The model here, of course, is not prescriptive in any sense. It does not tell the decision maker what should be done. However, it shows some of the consequences of the deployment of solar energy technology in terms of atmo-

TABLE 17.13 KSIM Input for Solar Energy Example

	1	2	3	4	5	6	7	8	9*	Initial value	Symbol
1	0	5	4	0	0	0	0	3	0	0.1	S
2	- 2	0	0	0	3	2	0	0	4	0.3	F
3	3	0	0	2	0	0	0	0	0	0.3	P
4	8	0	0	0	5	0	0	0	0	0.2	R
5	- 1	0	0	0	0	3	0	0	0	0.5	A
6	- 1	- 2	0	0	- 5	0	0	0	5	0.7	U
7	0	0	0	0	0	1	0	0	0	0.5	M
8	3	0	0	1	0	0	0	0	0	0.1	T

*Outside world.

```
SOLAR_USE               =      S          .1
FOSSIL_PRICE            =      F          .3
SOLAR_PERFORM           =      P          .3
SOLAR_R&D               =      R          .2
ATMOS_PULLUT            =      A          .5
FOSSIL_USE              =      U          .7
COAL_MINERS             =      M          .8
SOLAR_TECHNICIANS       =      T          .1

      0   .1   .2   .3   .4   .5   .6   .7   .8   .9   1.0
      +....+....+....+....+....+....+....+....+....+....+
    0|     *    R    *         A         U    M          |
   25|     TS    R   P  F       A        U    M          |
   50|     T S     R  P    F     A        U   M          |
   75|      T  S     RP       F   A       U   M          |
  100|      T   S      *         FA       U    M         |
  125|      T     S    P R         AF     U    M         |
  150|       T      S   P  R        A F   U    M         |
  175|       T         S P    R      A  F U    M         |
  200|        T          SP     R    A    *    M         |
  225|        T            PS      R  A   U F  M         |
  250|         T            P S      RA   U   FM         |
  275|          T            P   S     ARU     FM        |
  300|           T             P    S   AUR     *        |
  325|            T             P     S *    R  MF       |
  350|             T              P    UAS     RM F      |
  375|              T              P  U  A S    *  F     |
  400|                T              *   A   S  M R F    |
  425|                 T            U P   A    SM  RF    |
  450|                   T         U    P A     *   RF   |
  475|                     T      U      PA      MS  *   |
  500|                      T    U        AP     M S FR  |
  525|                        TU          A P    M  S *  |
  550|                        UT           A P   M   SFR |
  575|                      U    T         A  P  M   SFR |
  600|                     U       T       A   P M    *R |
  625|                   U          T      A    PM    S* |
  650|                  U             T    A     *     *R|
  675|                U                T   A     MP    *R|
  700|              U                   T  A     MP    *R|
  725|             U                      TA     M P   *R|
  750|           U                         *     M  P  F*|
  775|          U                          AT    M  P  F*|
  800|        U                            A T   M  P  F*|
  825|       U                            A   T  M   P F*|
  850|      U                             A    T M   P  *|
  875|     U                              A    T M    P *|
  900|    U                               A     TM    P *|
  925|   U                                A      *    P *|
  950|  U                                A       *    P *|
  975|  U                                A       MT   P *|
 1000| U                                 A       M T   P*|
      +....+....+....+....+....+....+....+....+....+....+
      0   .1   .2   .3   .4   .5   .6   .7   .8   .9   1.0
```

Figure 17.1 KSIM output for solar energy example.

spheric pollution, the employment of coal miners, and the employment of solar energy technicians.

Not all TAs are centered on computer models; in fact, most are not. Assessments do, however, attempt to take into account the full range of consequences of the introduction of a new technology. As shown in the model, these consequences include not only the physical effects on the environment, but impacts on people, such as changes in employment. Further impacts, such as illness and death from pollution, might also be included.

Whether a TA is centered on a computer model or utilizes some other means of examining the consequences of technological change, the emphasis is on what happens to society and not simply on the technology itself. However, the basis for forecasting the effects on society is a forecast of the technology itself.

17.5 Summary

Technology produces changes in society; indeed, the primary purpose of technology is social change. It is introduced to perform some function deemed useful by at least some people in society. However, technology often has consequences beyond those intended by its inventors and sponsors, and it is these unintended and unanticipated consequences that are the concern of TA. If these consequences are to be anticipated, it is necessary to have estimates of the performance, the rate of adoption, and the scope of deployment of the new technology. The technological forecaster, then, plays an important role on the team carrying out TA. The forecaster's task is to provide a description of the technology the team is trying to assess.

References

Coates, J. F., "Some Methods and Techniques for Comprehensive Impact Assessment," *Technological Forecasting and Social Change*, 6(4):341–358.
Lyle, K. L., *Scalded to Death by the Steam*, Algonquin Books, Chapel Hill, NC, 1983.
Porter, A. L., A. T. Roper, T. W. Mason, F. A. Rossini, and J. Banks, *Forecasting and Management of Technology*, John Wiley and Sons, Inc., New York, 1991.
Shafarevich, I., *The Socialist Phenomenon*, Harper & Row, New York, 1980.

Problems

1 In the model presented in the chapter, atmospheric pollution is affected only slightly by the use of solar energy. Increase the size of this effect from -1 to -5 and run the revised model. What difference does this make?

2 In the model presented in the chapter, the number of solar energy technicians lags behind the growth of the use of solar energy. Is this reasonable? Increase the impact of solar energy use on the number of solar energy technicians from $+3$ to $+8$ and run the revised model. Does this decrease the lag? Are there any factors in the "outside world" that might affect the number of solar energy technicians (either positively or negatively) in addition to the use of solar energy itself? If so, make the appropriate change to the model and run it.

3 Add an "illness and death" variable to the model in the chapter. How is this variable affected by atmospheric pollution? How is it affected by the outside world? Is it affected by any other variables in the model? Make the necessary changes in the cross impacts and run the revised model.

Chapter

18

Some Common Forecasting Mistakes

18.1 Introduction

It has been frequently suggested that the art of forecasting can be improved by studying past forecasts and determining why they went wrong. As we indicated in several chapters in this book, whether or not a forecast comes true is hardly the proper criterion by which it should be judged. Nevertheless, there is certainly some appeal inherent in this criterion. First, it appears unambiguous. Second, if the forecast itself is not of a self-defeating nature, then its utility for decision making is probably highly correlated with its eventually coming true. Third, in retrospect, there is probably no other measure of the utility of a past forecast, since in general we do not have access to the decisions that were based on the forecast. Hence, despite the admitted shortcomings of "coming true" as a measure of the utility of past forecasts, until better records of decision making are kept and made available to scholars, it may be the best we can manage.

It is interesting to observe that, in some cases, it is hardly possible to tell whether a past forecast came true or not. For instance, the forecasts contained in the original Gordon-Helmer Delphi study of 1964 by Gordon and Helmer were examined 6 years later in an attempt to assess their validity. It was found that despite all the attempts during the original study to make statements of events as unambiguous as possible, in many cases it was impossible to tell whether an event in question had occurred yet or not. From the hindsight vantage point of 6 years later, it was clear that many of the statements were a great deal more ambiguous than they were originally thought to be. Hence, it may not always be possible to tell whether or not a forecast has come true.

Despite this problem, and that of self-defeating forecasts, however, there is much to be learned by examining past forecasts, particularly those that "went wrong" in some spectacular fashion. To the extent that the sources of error can be specifically identified, the present-day forecaster can learn to avoid them. We will look at some typical examples of forecasts that turned out to be dramatically wrong and examine some of their identifiable sources of error.

One example of an erroneous forecast is given by Lord Bowden (1970) in a description of some of his early experiences with Ferranti, the British computer firm:

> I went to see Professor Douglas Hartree, who had built the first differential analyzers in England and had more experience in using these very specialized com-

> puters than anyone else. He told me that, in his opinion, all the calculations that would ever be needed in this country (i.e., Great Britain) could be done on the three digital computers which were then being built—one in Cambridge, one in Teddington, and one in Manchester. No one else, he said, would ever need machines of their own, or would be able to afford to buy them. He added that the machines were exceedingly difficult to use, and could not be trusted to anyone who was not a professional mathematician, and he advised Ferranti to get out of the business and abandon the idea of selling any more of them.

George Chatham (1970) quotes a forecast regarding the future of the turbine engine for aircraft:

> In its present state, and even considering the improvements possible when adopting the higher temperatures proposed for the immediate future, the gas turbine could hardly be considered a feasible application to airplanes mainly because of the difficulty in complying with the stringent weight requirements imposed by aeronautics. The present internal-combustion-engine equipment used in airplanes weighs about 1.1 pounds per horsepower, and to approach such a figure with a gas turbine seems *beyond the realm of possibility with existing materials.* The minimum weight for gas turbines even when taking advantage of higher temperatures appears to be approximately 13 to 15 pounds per horsepower.

(Note: This forecast was prepared by The Committee on Gas Turbines, of the U. S. National Academy of Sciences, and was published on June 10, 1940, nearly a year after the Heinkel 178 had actually flown in Germany using von Ohain's jet engine.)

Several examples of erroneous forecasts are given in Jantsch (1967). He quotes an IBM estimate made in 1955 that by 1956 there would be 4000 computers in use in the United States, whereas the actual figure was 20,000. He also quotes the chairman of the board of United Airlines as stating that he had "never visualized the speed and comfort of today's jet travel, the volume of air traffic, or the role of air transportation in American national life."

In 1935, C. C. Furnas wrote a book entitled *The Next Hundred Years* in which he made a sizable number of forecasts of the technological advances to be expected by the year 2035. In the mid-1960s he reviewed some of his own forecasts as being spectacularly off the mark (1967). Some of the forecasts he selected as being in serious error were the following:

> It seems obvious that one task confronting aviation is to improve design of planes so that they can improve performance at the higher altitudes. It certainly can be done...Three hundred fifty miles per hour for transport planes at the higher altitudes is not an unreasonable figure to expect. That brings San Francisco within 8 hours of New York and London within 10 hours...
>
> If man could only induce matter, any kind of matter, to explode into energy then he would have a power house that really is something. Half a pound of matter, says Professor Aston, would contain enough energy to drive a 30,000 ton vessel from New York to Southampton and back at 25 knots. Can the atomic physicists show us how to tap that reservoir profitably? The last word has not been said yet but do not buy any stock in an Atomic Energy Development Company. You will certainly lose....

> It is about time that some of the glory of taking lunar trips be robbed from the pages of Jules Verne. Silly chatter? Perhaps. Perhaps not. There is nothing fundamentally erroneous in the idea. I do not doubt that some day a great crowd will gather at some aviation field to watch a man be locked into a peculiar bullet-like machine on the first bonafide trip to the moon. I hardly expect it to be in my day....

Some of the most extensive studies of past forecasts have been carried out by James R. Bright. One forecast on which he has reported (1970) is a study sponsored by the U.S. Department of the Interior in 1937. According to Bright, about 60 percent of the individual forecasts in this study have actually come to pass. Nevertheless, the authors of the study missed radar, the jet engine, nuclear power, antibiotics, nylon, and a number of other economically significant technological advances of the decade or two following their study.

These examples of the errors of past forecasts have all been instances of underestimating possible progress. It should not be thought, however, that this is the only possible direction in which to err. Errors of overestimation are made as well. For instance, the authors of the 1937 forecast mentioned above anticipated a rapid growth in private flying in the 1940s and 1950s, which simply did not take place. They also forecast for the immediate future the facsimile printing of newspapers in the home.

Thomas J. O'Connor (1969) made a study of the so-called von Karman report, prepared in 1945 by a team of some of the most outstanding scientists in the United States under the leadership of Theodore von Karman. This report was prepared at the request of General H. H. Arnold, Commander of the U.S. Army Air Forces and was intended as a guide for Air Force research and development in the post-World War II era. The authors of this report forecast the following: the use of Flying Wing aircraft, which in fact turned out to be inferior to other types; the use of nuclear power for the propulsion of aircraft, which is still far from being technically feasible; the supersonic propeller, which was bypassed by the jet engine; and the widespread use of "supersonic pilotless aircraft" (later to be known as "cruise missiles") of intercontinental range, which turned out to be bypassed by the ballistic missile. In all of these cases, the authors of the report either overestimated the rate of progress of the technology needed to overcome the difficulties facing these devices or missed out on alternatives that turned out to be superior.

These instances of errors contained in forecasts made in the past are not intended in any way to denigrate the competence of their authors. In general, the forecasts quoted above were made by persons of outstanding stature. Especially in the cases of reports requested or commissioned by government organizations, the forecasters were among the most competent scientists in the country. Our objective in studying incorrect forecasts is to recognize that even competent people can make errors. Our main concern is to identify the causes of errors and avoid repeating them.

In the remainder of this chapter, three major categories of causes of error will be discussed: environmental factors, personal factors, and core assumptions. Procedures that can help reduce the likelihood of errors from these sources will also be discussed. Even if these errors of the past are avoided, however, this will not guarantee perfect forecasts. As Winston Churchill is reputed to have once remarked, "These mistakes will not be repeated. We will make enough of our own." Nevertheless, eliminating the known causes of error to the greatest extent possible cannot help but improve the utility of forecasts to decision makers.

18.2 Environmental Factors that Affect Forecasts

Many of the studies of past forecasts have concluded that a significant cause of error is the underestimation or omission of certain factors in the environment that could have an impact on the technology being forecast, i.e., the forecaster takes a narrow look at only the technology to be forecast. As a result, the forecast may be rendered invalid by some factor that was either completely ignored or whose impact was grossly underestimated. The possibility of this type of mistake can be reduced by systematically examining environmental factors that can have an impact on the technology being forecast. We will look at some of these factors, utilizing as a framework the dimensions of the environment first introduced in Chap. 3. Each dimension will be examined separately, and some factors whose impact had led to the failure of past forecasts will be identified.

Technological. A common mistake of forecasters is to fail to look at all of the stages of innovation, particularly the early ones. Most of the items missed in the 1937 forecast discussed in the previous section, such as penicillin, were actually in existence in one or another of the early stages of innovation. Had the forecasters looked closely at items that were then perhaps at the stage of scientific findings or laboratory feasibility, they might not have missed taking them into account. Hence, the forecaster should deliberately search for items in the early stages of innovation that might turn out to have an impact on the forecast.

Another common failing of forecasters is to ignore developments in other fields, or other countries, that may supersede the technology to be forecast. For instance, someone trying to forecast the future of vacuum tubes in 1945 would have missed the work on semiconductors that led to the transistor had only the field of vacuum tubes been investigated. In an actual case, a 1950 forecast of materials for the United States failed to identify the impact that the basic oxygen process would have on steel making within the subsequent decade, even though this process was already being adopted on a large scale in Austria. Thus, the forecaster should look at other technologies, other industries, and other countries for developments that might supplant or supersede the technology to be forecast. Not only the later stages of innovation but also the earlier ones should be examined.

Another common failing of forecasters is to ignore the impingement of one technology on another. The widespread growth of the computer, for instance, might not have been possible without the transistor or something that shared its properties of low cost, high reliability, and low power consumption. Hence, forecasters should examine not only the technology with which they are concerned but also its supporting and complementary technologies. The interactions between them may be such that a predictable improvement in a supporting or complementary technology may greatly alter the growth rate of the technology being forecast.

Finally, a common failing of forecasters can be classified as "scientific error" or "incorrect calculation." This may take the form of an assumption that the present state of scientific knowledge represents some final or ultimate level of knowledge. It may also take the form of a calculation that is formally correct but uses unduly pessimistic or optimistic values for certain factors that have not yet been evaluated. Finally, it may take the form of an assumption that something will be done "the hard way" or "by brute force." Such an error may be particularly significant when it occurs in the calculation of an upper limit to a growth curve. The forecaster must then make sure that the assumptions and calculations of the forecast do not treat the present level of scientific knowledge as the last word, that the values assumed for some unknown factors are not all biased in the same direction (a good practice is to carry out the calculations with both an optimistic and a pessimistic bias and see what the differences are), and that the forecast takes into account the cleverness people have shown in the past in overcoming supposedly "ultimate" barriers.

Economic. Typical failures here arise from an overoptimistic estimate of the acceptance of some innovation or an overpessimistic estimate of the problems of introducing it. For instance, the forecast of facsimile newspapers in the home, cited earlier, failed to take into account the costs to the consumer. This failure often amounts to mistaking the "operational prototype" stage for the "commercial introduction" stage. Technical feasibility does not necessarily imply commercial success. On the other hand, innovations in financing a new device may overcome the cost barrier. For instance, had IBM insisted on selling its computers, or Xerox on selling its copying machines, probably neither would have been a commercial success. By leasing their products to the user, both IBM and Xerox overcame the barrier represented by the high cost of manufacturing the devices. Hence, if a forecaster is required to predict the "commercial introduction" stage of innovation or some later stage, both overoptimism and overpessimism in the evaluation of economic factors must be avoided.

Managerial. The primary problem here seems to have been the inability to see the impact of advances in managerial technology. New managerial techniques may make a much more rapid advancement of technology possible. For instance, the managerial technique known as PERT (*project evaluation and review technique*), which was devised for the Polaris program, had a major impact on making that program manageable. It is now widely used in the construction industry and has had a significant impact on management practices in that industry. Managerial technology has not reached its limit any more than have many other technologies. Hence, the forecaster should be alert to the possibility that the advancement of a technology may be speeded up significantly as a result of better management techniques.

Political. Changes in the political environment can have a significant impact on technological progress. This may arise from the creation of new agencies charged with achieving specific goals (e.g., the National Aeronautics and Space Administration, Atomic Energy Commission, and Federal Water Pollution Control Agency). It may arise from the replacement of one official having one set of views by another having different views. Such a replacement may take place through death, retirement, transfer, or the loss of an election. Changes may also arise from efforts to foster a specific technology such as agriculture; to control the detrimental effects of another technology such as automobiles; or to restrict the growth or spread of yet another technology, as the U.S. Atomic Energy Commission and its successor agencies have attempted to restrict the development of centrifuge technology for separating uranium isotopes on the grounds that such technology would foster a proliferation of nuclear weapons. The end of the "Cold War" is almost certain to slow the development of U.S. military technologies. If some event of this nature has already taken place (i.e., the death of an individual or the creation of a new agency), the forecaster should take into account its impact on the technology being forecast. If some change is possible (e.g., an officeholder is reaching retirement or the creation of a new agency has been proposed), the possible impact should be considered. If necessary, the forecaster may attempt to obtain a political forecast from an appropriate source. In any case, the impact of possible political changes on the technology being forecast must be taken into account.

Social. A major source of error in past forecasts has arisen from the failure to take into account changes within society, particularly population growth, changes in the age distribution, and the growth of affluence. In addition, the impact of special interest groups within society, such as labor unions, conservation groups, and consumer organizations, must be taken into account. Many forecasts made during the 1930s were seriously in error because they grossly underestimated the population growth in the United States, known as the "baby boom" following 1945, and thereby underestimated the demand for certain technologies and devices such as the automobile. The effect of the "baby bust" or "birth dearth" in the late 1960s and 1970s is being incorporated into current forecasts of population for the year 2000. However, increasing birth rates during the 1990s may make those forecasts equally wide of the mark. Thus, the forecaster must realize that population can have an impact on a forecast and should therefore be consciously aware of the population growth assumptions (or forecasts) being built into it. The same holds true with assumptions about per capita income. The forecaster must examine the possible impact of actions by various groups within society. Even similar kinds of groups may have different attitudes. The United Mine Workers Union, for instance, traditionally encouraged the use of new technology in coal mines. The International Typographers Union, by contrast, was very slow in permitting the utilization of new technology in the printing industry. Thus, the forecaster must identify the groups that may have an impact, determine what their attitudes are, and take these into account. Resistance from some group may succeed in slowing the rate of progress well below what would be feasible from the technical standpoint.

Cultural. Changes in the values subscribed to by society can have an impact on a forecast. One example of a radical shift in the values of American society is the change in attitudes toward education in science and mathematics that resulted from the launch of the first successful artificial satellite, the sputnik, by the Soviet Union. Conversely, in the 1960s and 1970s, some commentators on the American scene claimed to see a shift toward a "postindustrial society," that would downgrade the values of the preceding industrial society, values that were responsible for the rapid economic and technological growth during the years 1945 to 1970. Such a possibility would certainly have had an impact on any technological forecasts intended to cover the subsequent two or three decades. Beginning in the 1980s, we saw a shift back toward "reindustrialization" of the United States and claims that a "service economy" was the road to universal poverty.

These shifts in values may be set off by some external event such as the sputnik or the growing Japanese share of the U.S. automobile market. They may also be set off by the actions of some individual or by some book, article, or play that brings some problem to sudden public attention. Ralph Nader's book *Unsafe at Any Speed* brought the problem of automobile safety to public attention and triggered a shift in values regarding this subject. On the other hand, values may shift slowly and cumulatively, as a result of factors either internal or external to society, with little notice being taken of the shift until it has reached significant proportions. For instance, the decade 1960 to 1970 saw a slow but steady shift in the values of youth in the United States until a significant proportion of the young people no longer subscribed to the value systems of their parents. In retrospect, this shift seems obvious, but the rate at which it crept up on society was so slow that it was hardly noticed until it had become quite large. Many forecasts made during the late 1960s and 1970s, based on this shift, also turned out to be in error, as a contrary shift began in the 1980s. The technological forecaster must take into account any recent shifts or current trends in the cultural values subscribed to by society and should also be aware that a forecast may be invalidated by a subsequent shift in such values. A technology may have its rate of growth slowed down simply because people no longer care.

Intellectual. Here again we are concerned with values, those of the intellectual leadership of society. These values are of concern to the technological forecaster from two standpoints. First, the values of the leadership may differ from those of society at large. In the short term, the values of the intellectual leadership may be more important than those of society at large, since the former are likely to have more influence on the course of the development of technology. Conversely, over the long term the attitudes of society at large may have more influence than those of the intellectual leaders, who cannot remain leaders indefinitely if they have no followers. If the values of society at large remain unchanged, new intellectual leaders will arise who reflect those values. Second, the values of the intellectual leaders may be precursors of the values of society at large. The technological forecaster may be able to identify forthcoming shifts in the values of society at large by determining what values are currently subscribed to by the intellectual leaders. In either case, whether the values of the intellectual leadership shift to match the values of society or vice versa, the forecaster should not ignore the impact of this dimension on technology. Instead, an attempt must be made to determine what the impact will be, and it should be included in the forecast.

Religious-Ethical. At the present time, it seems unlikely that any religious group could have a large-scale impact on technological advance, at least on doctrinal grounds. However, it is quite possible that at some time in the future, religious, ethical, and professional groups, together or separately, may come to have an impact on technological advance by raising questions about the proper goals and objectives for humanity. The "we can, therefore we should" attitude regarding technology has been under attack for a decade or more, and the strength of this attack has not diminished. In addition, these groups may take stands regarding the personal responsibility of the technologist for any dangers or unwanted side effects arising from a new design. There have already been suggestions to the effect that a proper code of ethics for engineers would include such personal responsibility. The first of these attitudes may slow the rate of growth of technology generally; the second may give rise to increased empiricism and incrementalism in design, making engineers hesitant to depart too far from something that is known to work safely and reliably, although at a comparatively low level of performance. If either of these changes took place, they would have a significant impact on technology, and, therefore, they are important to the forecaster. In addition, it is possible that other similar changes will have an impact on technology and will, therefore, also be of concern to the forecaster. Hence, the religious-ethical dimension must be considered in preparing forecasts.

Ecological. This dimension is likely to have a growing impact on the course of development of technology and is therefore of considerable importance to the technological forecaster. For instance, the 1937 forecast mentioned above presented the following forecast regarding waste disposal (Bright, 1969): "There is no cheaper or better way of disposing of sewage than by dilution with water...Rivers...must carry at least all of the soluble wastes of life and industry." The authors of this forecast clearly did not anticipate the impact on the environment of a large and affluent population following such practices. By contrast today, there are laws in place limiting emissions of pollutants, and more stringent limits are continually being demanded. In addition, there are demands that future product designs take disposability and recyclability into account. Clearly, then, the forecaster must consider this dimension of the environment. In particular, it must be recognized that if a certain scale of use of a particular technology produces an intolerable level of damage to the environment, that technology may not be allowed to reach so large a scale or may be supplanted by a less damaging technology.

Each of these dimensions of the environment represents a possible source of impact on technology. Therefore, in order to prepare a useful forecast of a technology, it is essential to look at each dimension to attempt to identify the impact. The impact may be the acceleration or deceleration of the rate of progress of the technology because of a change in the level of interest, demand, or support, or because of a consequence arising from an interaction of several technologies. The impact may also take the form of an acceleration or deceleration of the rate of progress of an alternative technology. Whatever the impact, it must be identified and incorporated into the forecast if it is to be of any great utility for decision-making purposes. A forecast that ignores some major change in any dimension of the environment may even mislead the decision maker into neglecting the impact of some important developments.

18.3 Personal Factors that Affect Forecasts

In the previous section, we have examined some environmental factors that could have an impact on technology and that, if ignored, could render a forecast worthless. These were derived from studies of the failures and errors of past forecasters. In this section, we will examine some factors internal to the forecaster that have likewise been found to invalidate forecasts. These may be harder to identify in specific cases, because the forecaster may not want to admit to any shortcomings. However, forecasters who are aware of the possibility of their existence will have a better opportunity to search for and recognize them in themselves. To the extent that a forecaster can identify and make allowances for these personal factors, the result will be that the forecast will be more useful, and if not self-defeating, more likely to come true.

Vested interest. This problem arises when forecasters have a personal interest in an organization or a particular way of doing things that might be threatened by a change in technology. They may have spent their professional lifetimes with a particular organization or industry; they may have many favorable or pleasant personal associations or experiences with an organization or a group of people; or they may have a feeling of personal commitment to an organization or a way of life. In any of these cases they may tend to suppress a forecast that appears to threaten any of these interests. Bright (1964) cites two statements made by executives of locomotive-manufacturing companies that illustrate this point. Robert S. Binkerd, vice-president of the Baldwin Locomotive Works, said in a 1935 speech to the New York Railroad Club:

> Today, we are having quite a ballyhoo about streamlined, light-weight trains and diesel locomotives, and it is no wonder if the public feels that the steam locomotive is about to lay down and play dead. Yet over the years certain simple fundamental principles continue to operate. Sometime in the future, when all this is reviewed, we will not find our railroads any more dieselized than they are electrified.

W. C. Dickerman, president of the American Locomotive Company, said in a 1938 speech to the Western Railway Club in Chicago:

> For a century, as you know, steam has been the principal railroad motive power. It still is, and, in my view, will continue to be.

However, as we saw in Chap. 8 from Mansfield's work on the rate of diffusion of innovations in various industries, the diesel locomotive followed the typical growth curve quite closely. In fact, Mansfield's data showed that by 1935, 50 percent of the railroads in the United States had already bought at least one diesel locomotive. Diesels accounted for far less than 50 percent of all locomotives at that point, but the trend was clearly there (at least it is clear now in hindsight; obviously it was not clear to everyone in 1935). At any rate, here we have a clear-cut example of two forecasters who had a commitment to a particular technical approach, so much so that their companies failed to anticipate the transition to a new technical approach despite the available evidence. Thus, forecasters must be aware that a vested interest in some organization or some way of doing things may obscure the true implications of the evidence they are examining. They may consciously or unconsciously refuse to forecast a technological development that would threaten their vested interest. Note, however, that this would be a disfavor to the forecaster or to the decision maker who utilizes the forecast. The impact, when it comes, will be all the more severe for being unexpected. Thus, forecasters must do everything possible to make sure they are not being biased by vested interests.

Narrow focus on a single technology or technical approach. This error is committed when the forecaster looks at only one specific technology or technical approach. It differs from the previous one in that the narrow focus arises from the forecaster's experience or training, rather than from a vested interest. Because of experience or training, the forecaster is conditioned to think only in terms of one type of solution. This may lead to the forecaster making a "window-blind forecast" of the type described in Chap. 1 and to missing or overlooking a technical approach that could replace or supplant the one being forecast. A major example of this error is the behavior of firms making mechanical calculators and cash registers. The major innovations in their industries involved electronics and computer technology, which were introduced by firms outside or peripheral to the calculator or cash register industries. The companies that led these industries in the 1960s either dropped out by the 1980s or had to scramble to catch up to the newcomers to the industry. Their major problem was a focus on a particular mechanical technology, which led them to overlook the electronic technology that took over the industry.

Commitment to a previous position. This situation arises when a forecaster has prepared a forecast on a particular topic at some previous time and is asked to take another look at the same topic or a related one. As a result of the earlier forecast, the forecaster may be unwilling to make any alterations, even in the face of new information. This may result from the forecaster's strong personal commitment to a previous forecast, leading to an unwillingness to recognize that recent information may require the forecast to be changed. Alternatively, it may result from the forecaster's feeling that if the previous forecast is changed, the forecaster will lose credibility or stature with the client or the public. Thus, even when the forecaster realizes that more recent information has significant implications for the field in question, this personal commitment may make the forecaster unwilling to make changes in the previous forecast. It should be apparent that no matter what the reason for this reluctance to change a previous stand, the situation can be dangerous. The forecaster's reputation is likely to suffer much more if clients recognize an unwillingness to take new information into account than it is if the forecaster admits that an earlier position was based on incomplete information and no longer appears as soundly based as it did at the time of the previous forecast. Forecasters must examine their own attitudes carefully to be sure they are not inadvertently committing this mistake.

Overcompensation. This situation arises when the forecaster bends over backward to prevent a personal commitment to a particular view, a vested interest in some technology, or a strong desire to see a particular outcome, from distorting the forecast. As a result, the forecast is distorted in the opposite direction. This distortion is just as bad as the one arising from a commitment to a particular technology, organization, or viewpoint. It is harder to detect, however, by both the forecaster and the clients, because it looks so much like an attempt to be "fair." Hence forecasters must be just as aware of this possibility as they are of the more obvious forms of bias.

Giving excessive weight to recent evidence. It is only natural that recent events will loom large in the mind of the forecaster, larger than those of the more distant past, even when the latter may contradict the recent events. If some activity has been following a long-term rising trend, for instance, but has recently had a downward fluctuation similar to many minor downward fluctuations in the past, the forecaster may give more weight to the recent downward fluctuation than to the long-term trend. While this may be natural, it is certainly not conducive to effective and useful forecasting. The forecaster must look for long-term trends and make a forecast that is contrary to these trends only when there is good reason to believe that the phenomenon is departing from the trend. To ignore the trend and base a forecast solely on the latest fluctuation, either up or down, is asking for trouble. If the long-term trend is still continuing, there will inevitably be a counterbalancing fluctuation to offset the one on which the forecast is based. It is essential, then, that the forecaster give the proper weight to all relevant evidence, including that from the bulk of the past. If one of the rational and explicit means of forecasting described earlier in this book is used to prepare the forecast, the danger from this problem is somewhat reduced. In such a case, it is obvious to both the forecaster and the forecast user which data have been used and which have been omitted. The reasons given by the forecaster for omitting specific items of data are more readily analyzed. However, even in this case there is still the possibility that the forecaster will fail to use earlier data, not because of any rational conviction that it is no longer relevant, but simply because of a subjective feeling that "it doesn't fit with more recent events." Hence, there is no sure cure for this problem, and the forecaster must take pains to avoid making such a mistake.

Excessive emphasis on the troubles of the recent past. This is a more vicious form of the preceding deficiency. Serious troubles are likely to make a deep impression, predisposing the forecaster toward a particular outlook. Troubles may come to be viewed as permanent instead of as a temporary phase that may, but need not, recur. Schoeffler (1955) gives an extensive description of this problem, in terms of forecasts of the U.S. economy prepared in the early 1940s, with regard to the reconversion of the economy from wartime to peacetime production. Virtually all the forecasts made at that time foresaw extensive unemployment, slow growth in the level of production, and general economic stagnation. Schoeffler identified some specific errors made in these forecasts, a few of which are itemized below:

1. Industrial reconversion tempo underestimated
2. Construction revival misjudged
3. Quick retirement of emergency workers after V-J day
4. Underestimation of effective foreign demand
5. Underestimation of rush to build up inventories

There were a considerable number of these specific errors, which Schoeffler has summarized thus:

> Continued reflection upon the etiology of the forecasting errors leads the writer to the conclusion that the underlying cause of the errors was *psychological*—not theoretical, statistical or methodological.
>
> *The originating cause of most of the reconversion forecasting errors was the prevailing psychological 'mind-set toward depression' among the forecasters.* Most of the forecasters had an unmistakable predilection to look at the economy through the dark glasses of the 1930s. This predilection had profound consequences for

> economic forecasting of postwar conditions. Almost inevitably, it led to an acceptance of the stagnation of the 1930s as the economic norm for the American economy of the twentieth century. Given this bias, it was natural to regard *re*-conversion as a return to basic prewar conditions and to minimize the continuing economic consequences of the war.

Nor were the effects of the depression limited to the years immediately following it. Ikle (1967), writing in 1966, stated:

> We make the mistake of focusing our predictions where our shoe hurts. If we have been hurt particularly badly, our predictions will look backwards to our old pain for a long time. The Great Depression of the 1930s had a depressing effect on all social predictions; this cast a longer shadow into the future than the real dislocations of the Depression itself. On a number of issues, our predictions are *only now* recovering from the Great Depression. Thus, the volume *Recent Social Trends,* written just before the worst years of the depression, takes a view of the future in many ways closer to our current agenda for predictions than the planning and forecasting done for some twenty years following the Depression.

Since this problem is closely related to the one described in the preceding section, the procedure for handling it is much the same. The forecaster must make a deliberate effort to look for long-term trends and behavior, instead of concentrating on recent events. This will be especially difficult, since as both Schoeffler and Ikle indicate, serious troubles may easily be interpreted as representing basic structural changes in the situation that render older data irrelevant. The awareness that this can happen, however, can be helpful in counteracting the tendency to assume a structural change, since rendering this assumption explicit makes it easier to examine and either verify or invalidate.

Unpleasant course of action. Sometimes the course of action that seems to be made necessary by a forecast will appear so unpleasant that the forecaster shrinks from making the forecast itself in the hope of thereby avoiding the necessity for the action. Ikle (1967) made the following comments on this problem:

> Some predictions imply such a horrible course of action that we experience a failure of nerve and find ourselves unable to go through with the indicated choice. This might be called the 'non-Freudian Oedipus effect.'

The reference here is to the parents of Oedipus, who lacked the courage to make sure their son was killed, despite the serious threat implied by his continued existence. Our concern here is with the forecaster who lacks the courage to take the necessary action and therefore refuses to make the forecast that would require the action. This attitude can only be described as "head in the sand." The forecaster whose nerve fails in an adverse situation is helpful to no one. It is admittedly not very helpful to give advice such as: "One must have the courage to face the consequences of one's own forecasts, or else stay out of the forecasting business." While the advice is sound, the forecaster who needs it is probably least likely to heed it. However, if one is aware of the possibility of stumbling into this pitfall, one has a better chance of avoiding it.

Dislike of the source of an innovation. There are cases where an innovation that already exists should be taken into account as having an impact on the technology to be forecast. However, the forecaster may for some reason have acquired a dislike for the source of the innovation. This source may be an organization or an individual. In either case, such a dislike of the source may cause the forecaster to ignore the innovation and its potential consequences. This attitude is not rational nor is it conducive to making an effective and useful forecast. It is, however, human. If the forecaster is aware that the possibility of this attitude exists and is honest about whether he or she suffers from it, the forecaster has a much better chance of avoiding this mistake.

Systematic optimism-pessimism. It has been observed, through studies of many forecasts, that there is a systematic shift from optimism to pessimism as the time length of the forecast increases. In the short run, many forecasters tend to be optimistic, especially about work they are responsible for. They feel confident that the immediate obstacles can be overcome and often underestimate the effort required to overcome them. In the long run, however, they tend to be pessimistic. They see many difficulties and barriers that they do not at the moment see how to overcome. Thus, these difficulties loom large and make the forecaster forget that difficulties of similar magnitude have been overcome in the past. In effect, the forecaster makes an implicit forecast that the historical rate of innovation will slow down or come to a complete halt. A useful rule of thumb is that the shift from optimism to pessimism comes at a point about 5 years in the future. While this may not be absolutely accurate, it can be helpful in warning the forecaster of where to expect one psychological attitude to be replaced by another. In any case, the forecaster should know that this tendency to shift from short-term optimism to long-term pessimism is a natural one and that by being aware of it, he or she can minimize its effect on forecasts.

It should not be thought that every forecaster is inevitably afflicted with each one of these deficiencies. Many will be afflicted with at most one or two, and some fortunate souls will escape them entirely. However, forecasters should take into account the fact that they are human beings and are, therefore, afflicted with human frailties. Some of the frailties that have in the past contributed to reducing the usefulness of a forecast have been mentioned above. If forecasters take into account the possibility that they may be afflicted with these frailties, they have a better chance of avoiding them, thereby increasing the usefulness of their forecasts.

18.4 Core Assumptions

William Ascher (1978) has shown that one of the most serious sources of error in forecasts is what he has called "core assumptions." These are the underlying assumptions made by the forecaster, regarding the subject area to be forecast. As Ascher puts it, once the core assumptions are made, selection of the forecasting method is usually obvious or trivial. An example drawn from the beverage can industry will illustrate the importance of choosing core assumptions correctly (Machnic, 1980).

During the 1960s and 1970s, there were three technical approaches to the function of providing a metal can to contain beverages. They were the following, in their order of development:

1. *Three-piece steel:* The can is a cylinder of tinned steel, with a tinned steel top and bottom soldered to the cylinder.
2. *Two-piece aluminum:* The can is a drawn, cup-shaped aluminum piece to which an aluminum top is crimped.
3. *Two-piece steel:* The can is the same as the two-piece aluminum, except that it is made of steel.

The two newer technical approaches supplanted the older three-piece steel can. Decision makers in the can industry and in the steel and aluminum industries needed to know which of the two new technical approaches would dominate. There were (at least) three possible interpretations of the situation:

1. A multilevel substitution was taking place. The two-piece aluminum can was substituting for the three-piece steel can, while the two-piece steel can was substituting for the two-piece aluminum can. Ultimately the two-piece steel can would dominate the market. This multilevel substitution is plotted in Fig. 18.1.

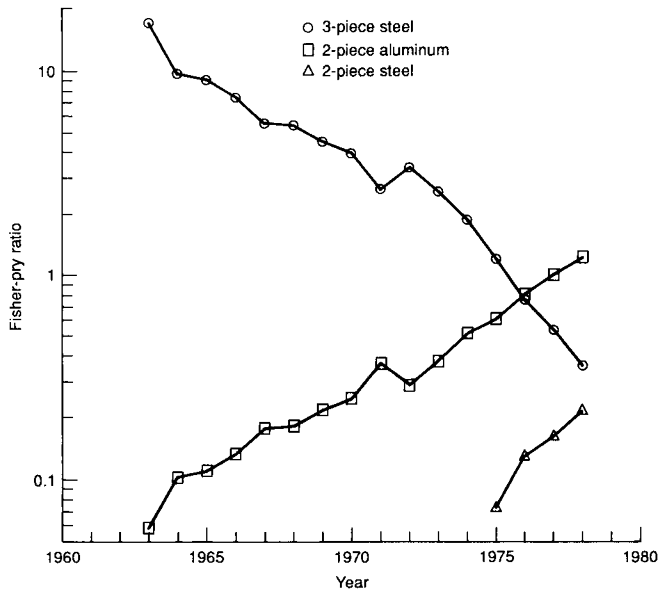

Figure 18.1 Multilevel substitution in beverage cans.

2. The three-piece can is being replaced by the two-piece can. Ultimately the market would be dominated by two-piece cans, with price determining whether steel or aluminum was used as the raw material. (By the 1970s, can-making machines were capable of using either steel or aluminum as the raw material.) Substitution of the two-piece can for the three-piece can is shown in Fig. 18.2.

3. The aluminum can is substituting for the steel can, with the two-piece steel can simply being a tactical action on the part of the steel industry to reverse their decline in market share. The substitution of aluminum for steel is shown in Fig. 18.3.

Each of the substitutions, shown as Fisher-Pry plots, appears plausible, yet they lead to dramatically different outcomes. One leads to the dominance of steel, another to the dominance of aluminum, and the third to price competition between aluminum and steel. How do we select among these predictions? We can make a statistical analysis of each plot.

If the first interpretation (multilevel substitution) were correct, we would expect the rate of decline of the three-piece steel can to remain unchanged after introduction of the two-piece steel can, but the rate of growth of the aluminum can should slow down after introduction of the two-piece steel can. Regressing the logarithms of the Fisher-Pry transformed market shares on time, we obtain the following results: The rate of decline for the three-piece steel can increases from -0.223 (0.0141) in 1963 to 1974 to -0.397 (0.015) in 1975 (the numbers in parentheses are standard errors). This difference is about 14

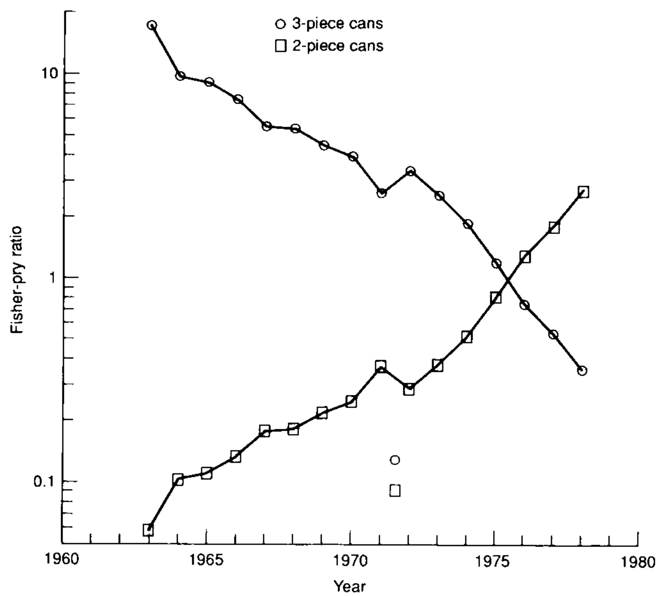

Figure 18.2 Substitution of two-piece cans for three-piece cans.

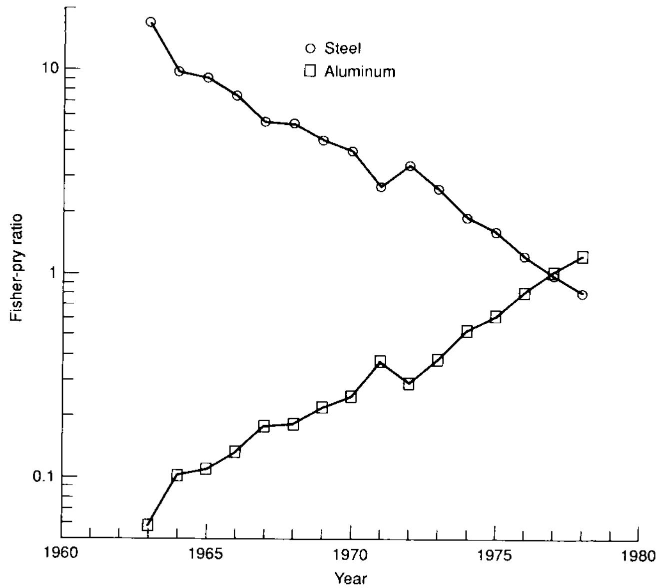

Figure 18.3 Substitution of aluminum for steel in beverage cans.

standard errors and is highly significant. The growth rate of the two-piece aluminum can *increases* from 0.172 (0.0119) to 0.231 (0.0114). This difference is about 4.9 standard errors and is also highly significant. Thus, we can reject the explanation that a multilevel substitution is taking place.

If the second interpretation (the two-piece can substituting for the three-piece can) were correct, we would expect the rate of decline of the three-piece steel can to remain unchanged after introduction of the two-piece steel can. We would also expect the combined rate of growth of the two-piece steel can and the two-piece aluminum can to be the same as the growth rate of the latter by itself prior to introduction of the former. But we have already seen that the rate of decline of the three-piece steel can increased after 1975, and the rate of growth of the two-piece aluminum can alone was greater in 1975 to 1978 than it had been in 1963 to 1974. Thus, the growth rate of all two-piece cans has to be greater after 1975 than it was previously.

If the third interpretation (aluminum substitution for steel) were correct, we would expect the combined rate of decline of the two-piece and three-piece steel cans to be the same after 1975 as the rate of decline of the three-piece steel can prior to 1975. We would also expect the growth rate of aluminum cans to be the same before and after 1975. The rate of decline of all steel cans for 1975 to 1978 is −0.228809. The difference between this and the rate of decline of the three-piece steel can in 1963 to 1974 is not statistically significant, being much less than one standard error. However, we have already seen that the growth rate of aluminum cans did not remain constant after 1975. On the other hand, the actual Fisher-Pry values for aluminum cans for the period 1975 to 1978 are above the 50 percent confidence limit but inside the 90 percent confidence limit for a projection based on 1963 to 1974. Similarly, the values for all steel cans are within the 90 percent confidence limits for a projection based on the steel market share for 1963 to 1974.

What can we conclude from this? The data are not consistent with either the first or second interpretation. The data for steel cans are completely consistent with the third interpretation, whereas the data for aluminum cans are consistent with it on one criterion but not on another.

The important point here is that no matter how precisely we fit curves to the data, the forecast is going to be dominated by our core assumptions about what is actually taking place. If we assume that aluminum is being substituted for steel, then our forecast will show an ultimate dominance af aluminum, and the fitted curve will only tell us how rapidly this will occur. However, if we assume that a multilevel substitution is taking place, with a succession of technical approaches, then our forecast will show the dominance of the two-piece steel can, and the fitted curve will only tell us how rapidly this will occur. No amount of precision in curve fitting can save the forecast if the forecaster has *chosen the wrong interpretation* for what is taking place.

Although the data are most consistent with the third assumption, this does not constitute sufficient grounds for assuming that this interpretation is correct. The forecaster should attempt to verify this core assumption by consulting with experts in the technology. The technological forecaster should not make the mistake of assuming that a knowledge of forecasting techniques can be substituted for an understanding of the technology itself, any more than an understanding of the technology can be substituted for knowledge of forecasting techniques.

More generally, forecasters should recognize that their core assumptions will dominate the outcome of a forecast. They should therefore attempt to identify the assumptions being made and verify them from appropriate sources such as experts in the technology. Unfortunately, there are no clear-cut rules or procedures for identifying core assumptions. All too often these are assumptions of which forecasters are unaware. Nevertheless, if forecasters recognize the problem, they are more likely to avoid it.

18.5 Summary

This chapter has covered some common mistakes made in forecasting. These have been identified by studies and reviews of past forecasts and can be grouped into three categories: environmental factors, personal factors, and core assumptions.

Environmental factors represent influences originating in the environment of the technology being forecast. If the forecaster concentrates solely on the technology concerned, these factors are likely to be missed, thereby producing a forecast that is less useful than it might have been. By systematically examining the environment, it is possible for the forecaster to reduce the likelihood of omitting or overlooking some significant factor and thus invalidating the forecast.

Personal factors represent psychological and other influences on the forecaster that may cause the inadvertent or unconscious distortion of a forecast.

This chapter has provided a checklist of such factors to help the forecaster avoid them.

Core assumptions represent those assumptions that underlie the forecast and that are often made implicitly by the forecaster. These assumptions dominate the forecast and may have more influence on the outcome than does the actual methodology used. The forecaster should attempt to identify the core assumptions behind a forecast and verify their validity.

There can be no absolute assurance that the forecaster is avoiding the mistakes described in this chapter. However, by being aware of them and making a conscious effort to avoid them, the forecaster has a much better chance of preparing an error-free forecast. The forecaster should always keep in mind that mistakes have been made by some of the most eminent scientists and engineers of the past and that no one is immune to them, but by a conscious and systematic effort, they can certainly be minimized.

References

Ascher, W., *Forecasting—An Appraisal for Policy-Makers and Planners,* Johns Hopkins University Press, Baltimore, MD, 1978.

Bowden, V., "The Language of Computers," *American Scientist,* 58 (1):43–53 (1970).

Bright, J. R., *Research, Development and Technological Innovation,* Richard D. Irwin, Homewood, IL, 1964.

Bright, J. R., "Some Insights from the Analysis of Past Forecasts," in J. R. Bright and M. E. F. Schoeman (eds.), *Technological Forecasting,* XYZYX Information Corp., Canoga Park, CA, 1970.

Calder, N. (ed.), *The World in 1984,* Penguin Books, Baltimore, MD, 1965.

Chatham, G. N., "Open Season on Aeronautics," *Aeronautics & Astronautics,* 22, February 1970.

Furnas, C. C., "The Next 100 Years—Three Decades Later." (Paper given to the Thursday Club, April 13, 1967).

Gamarra, N. T., "Erroneous Predictions and Negative Comments Concerning Exploration, Territorial Expansion, Scientific and Technological Development," Washington, D.C., Library of Congress, Legislative Reference Service, 1967.

Gordon, T. J. and O. Helmer, "Report on a Long Range Forecasting Study," RAND Paper P-2982, RAND Corporation, Santa Monica, CA 1964.

Ikle, F. C., "Can Social Predictions be Evaluated?" *Daedalus,* 733–758 (Summer 1967).

Jantsch, E., *Technological Forecasting in Perspective,* Organization for Economic Cooperation & Development, Paris, 1967.

Machnic, J. A., "Multilevel versus Single-Level Substitution: The Case of the Beverage Can Market," *Technological Forecasting & Social Change,* 18:141–149 (1980).

O'Connor, T. J., "A Study of Some Effects of the Scientific Advisory Group's Recommendations on Air Force Research and Development for the Period 1945–1950," unpublished Master's thesis, University of Denver, Denver, CO, 1969.

Paterson, R. W., *Forecasting Techniques for Determining the Potential Demand for Highways,* University of Missouri Press, Columbia, MO, 1966.

Schoeffler, S., *The Failures of Economics: A Diagnostic Study,* Harvard University Press, Cambridge, MA, 1955.

Whelpton, P. K., "An Empirical Method of Calculating Future Population," *Journal of the American Statistical Association,* 31:195 (1936).

Problems

What were the errors made by the authors of the following forecasts?

1 Roger Revelle, in Calder's *The World of 1984,* written in 1964, forecast that by 1970 a worldwide network of meteorological satellites would be established, which would be

coordinated with buoys in the open ocean, carrying meteorological instruments and reporting surface conditions by radio to the satellites for retransmission to central ground stations.

2 In 1936, P. K. Whelpton prepared a forecast of the U.S. population through 1980. Most of his work involved some very elaborate statistical methods for estimating unregistered births, so as to obtain an accurate estimate of the population in each age bracket in 1936. He then assumed three different fertility rates and two different mortality rate, as follows.

High fertility: The historic decline in fertility would halt and fertility would continue at the 1930 to 1934 rate, 2177 births per 1000 women.

Medium fertility: The historic decline in fertility would slow down, reaching a level of 1900 births for 1000 women living through the childbearing period in 1980.

Low fertility: The historic decline in fertility would slow down only slowly, dropping to 1500 births per 1000 women.

Low mortality: Life expectancy at birth would rise from 62.4 years in 1934 to 73 in 1980.

Medium mortality: Life expectancy at birth would rise to 70 in 1980.

High mortality: Life expectancy at birth would rise to 67 in 1980.

Whelpton combined these possible birth and death rates into six different population series. The highest and lowest series are shown in the following table. The entries show Whelpton's forecast as a percentage of the actual population. Whelpton's highest series was still a significant underestimate.

Year	Low series, percent	High series, percent
1940	99.2	100.3
1945	95.8	99.4
1950	89.7	96.3
1955	89.2	92.4
1960	76.3	88.7
1965	70.1	85.6

(Whelpton's forecast can be found in his 1936 article. The material given here is taken from the discussion in Paterson, 1966).

3 After the Wright brothers' successful flights in December 1903, there was considerable popular interest in aviation. Octave Chanute, the aviation pioneer, whose work with gliders antedated the Wrights' experiments, made this forecast of the future utility of aircraft in March 1904 (quoted in Gamarra, 1967):

> The machines will eventually be fast, they will be used in sport, but they are not to be thought of as commercial carriers. To say nothing of the danger, the sizes must remain small and the passengers few, because the weight will, for the same design, increase as the cube of the dimensions, while the supporting surfaces will only increase as the square. It is true that when higher speeds become safe it will

require fewer square feet of surface to carry a man, and that dimensions will actually decrease, but this will not be enough to carry much greater extraneous loads, such as a store of explosives or big guns to shoot them. The power required will always be great, say, something like one horse power to every hundred pounds of weight, and hence fuel can not be carried for long single journeys.

Chapter

19

Evaluating Forecasts as Decision Information

"You acted unwisely," I cried, "as you see
By the outcome." He calmly eyed me;
"When choosing the course of my action,"
said he,
"I had not the outcome to guide me."

AMBROSE BIERCE

19.1 Introduction

How does a decision maker know whether a forecast is a good one? The obvious measure of the goodness of a forecast is whether or not it comes true. However, this measure cannot be used by the decision maker, since at the time a forecast is needed, the outcome is not known. A forecast can be compared against the outcome only after the fact. Our concern here is with evaluating forecasts before the fact, when they are needed.

Another consideration about using outcome as a measure of forecast goodness is self-altering forecasts. Suppose the forecaster presents a forecast of some catastrophe. As a result of the forecast, the decision maker takes action that successfully averts the catastrophe. Conversely, suppose a forecaster presents a forecast of some good outcome. As a result of the forecast, a decision maker takes action that achieves the good outcome. Clearly it would be wrong to categorize the first forecast as "bad" because it didn't come true while categorizing the second as good because it did come true. Both were good forecasts because they triggered actions that led to better outcomes than would have been achieved without them.

The point is that the purpose of a forecast is to help the decision maker reach a better decision than would have been possible without the forecast. The criterion for judging a forecast, then, must be based on its utility for decision-making purposes. Even here, we must be careful to distinguish evaluation before the fact from evaluation after the fact. A decision that looked entirely reasonable in the light of information that was available when it was made may turn out to be worse than some other alternative that looked unreasonable before the outcome was known. However, the most we can demand of a decision is that it look reasonable when it is made. To demand that it appear right after the outcome is known creates the same problem as demanding that a forecast come true. It leaves us without any means of judging the quality of a decision at the time we need to judge it. Thus, we must demand that a decision take into account the best information available when it is made. We can demand no more than this. To judge a forecast's utility for decision-making purposes, then, we must determine whether the estimate of the future it provides is based on the best information available.

A forecast is a statement about the future, based on data about the past. But past data, in and of itself, never implies anything about the future. The forecaster must use some law, pattern, or logical relationship to link data about the past with statements about the future. Therefore, any judgment about the quality of a forecast must consider not only the data utilized, but the pattern or logic linking that data with the future.

This chapter presents a forecast-evaluation procedure that focuses on how well a forecast makes use of information about the past and about the nature of change in the subject area of interest. It is used to determine whether the forecast provides the best possible estimate of the future that can be obtained on the basis of available information.

19.2 The Interrogation Model

The evaluation procedure is called an interrogation model because it consists of a series of questions to be asked about the forecast. The order in which the questions are asked is chosen deliberately to narrow down the evaluation as quickly as possible. Early in the procedure, items that cannot be evaluated are identified and eliminated from further consideration; only those that can be evaluated further are allowed past each screening step. The interrogation model, thus, allows an efficient search for items that are important in each specific case.

The interrogation model involves four steps:

1. Interrogation for need
2. Interrogation for underlying cause
3. Interrogation for relevance
4. Interrogation for reliability

The first two steps address the problem area; the third and fourth steps address the specific forecast to be evaluated.

The sequence of steps in the interrogation model is shown in Fig. 19.1. The steps are described in detail below.

Interrogation for need

This step focuses on the decision maker and addresses the issue of why the forecast is needed. It deals with the issue raised in several earlier chapters: What

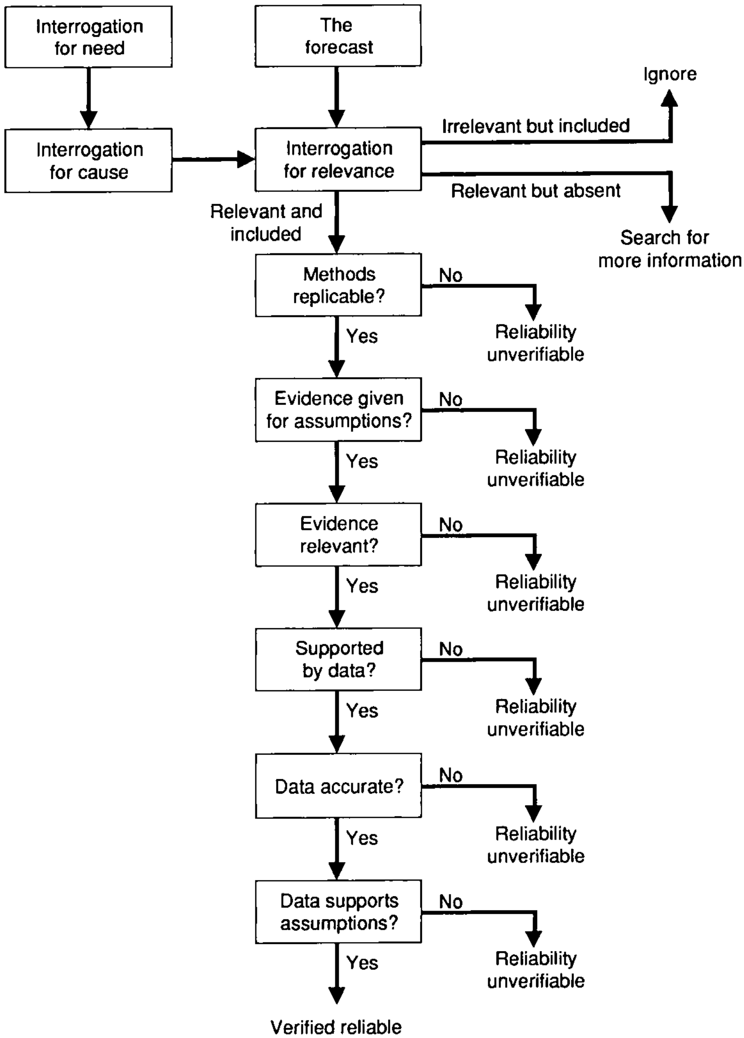

Figure 19.1 Interrogation model.

question are you trying to answer? A forecast cannot be evaluated for its utility in decision making until it is known what information the decision maker needs.

The following specific questions help elicit the specific need for the forecast. Their order is important, since the answer to each question determines the subject of the next question.

1. Who needs this forecast? What specific person, agency, business firm, or other actor is to make the decision for which the forecast is needed? (In what follows, this specific actor will be referred to as "we.")

2. What do we need to know about the future? This depends on both the identity of the actor, and the decision to be made. The initial statement of

need should be as broad as possible to avoid the risk of excluding elements that might later turn out to be important.

3. How badly do we need to know the future? This depends upon the decision to be made. What is at stake? What is the cost of not knowing the future? How much is knowing it worth?

4. Through what future time range must the forecast apply? What are the lags, delays, or rates of change involved in our decision? How much time is involved between making the decision and obtaining the payoff?

Interrogation for underlying cause

This step focuses on the subject area of the forecast. It attempts to identify the nature of changes that can occur in the subject area during the time range found to be of interest in the preceding step. The interrogation is carried out iteratively, i.e., each time a possible source of change is identified, that source is addressed in turn to determine what sources of change affect it. The result of this iteration is a relevance tree of the causes of change. The following specific questions can be used at each step of the iteration:

Step one: In what ways can the subject area change? Can there be changes in magnitude (size, numbers, performance, or capability)? Can there be changes in composition (increase or decrease in the number of components, or addition or deletion of components)? Can there be changes in its character (organization or interrelation of components)? Once the modes of change have been identified, the next question is asked.

Step two: What can cause these changes? At this step each mode of change is examined to determine possible causes of change. Note that there may be more than one possible cause of change at a given step. A possible cause may not be sufficient or even necessary for change; it may be only contributory. Nevertheless, all these possible sources of change must be identified at this step. The sectors of the environment that have been used before can serve as a checklist for searching out possible causes of change. Appropriate questions in each sector are given below.

Technological. Is there a competing technology that can cause change? Is a change in supporting or complementary technology possible? What effect would that have?

Economic. Will the costs of production, use, or operation change? Will the general economic climate change? Will financing be available on the scale required? Will there be a change in the potential market?

Managerial. Would a possible technological change require a scale of enterprise beyond the capabilities of present managerial techniques? Will there be a change in managerial techniques that would make a larger-scale enterprise feasible or make current enterprises more efficient?

Political. Who benefits by change? Who suffers from it? What are these parties' relative political strengths? Will there be changes in the missions or responsibilities of existing institutions? Will new institutions be established? Will there be changes in political leadership? Are vested interests involved? Would certain political objectives unrelated to the technology in question be hurt or helped by change? Would the acceptance of change from a particular source help or hurt some political objective?

Social. Will there be a change in population size? In age distribution? In the geographic distribution of the population? Will there be changes in major institutions such as the family, schools, and business? Will there be changes in society's "ideal image" of itself?

Cultural. Will there be large-scale changes in the values held by society? Will there be changes in the values held by major groups within society? Will demographic or geographic changes shift the relative importance of groups with different values?

Intellectual. Will there be changes in the values held by society's intellectual leaders? Will there be changes in the makeup of society's intellectual leadership through death, retirement, and so forth? Will the values of the intellectual leaders be subject to fads or fashions?

Religious-ethical. Is a change possible in the religious or ethical attitudes that have supported the current situation? What is the likely response of each major religious or ethical group to a possible change?

Ecological. Is the ecological impact of current practice likely to cause change? What would be the ecological impact of a specific change? Are there alternatives with less ecological impact?

At this point we have identified all the possible sources of change in the subject area we are concerned with. In effect, this gives us level one of a relevance tree (the original subject area is level zero). We continue then with the following steps:

> *Step three:* Iterate the process for each of the factors identified as a source of change in step two, identifying the causes of change and again using the sectors of the environment as a checklist. This iteration continues until a set of "fundamental" causes of change is obtained. This set consists of causes of change that are not worth further investigating (e.g., it is probably not worth investigating whether the sun will continue to shine during the period of interest). Once these fundamental causes have been identified, the relevance tree of causation is complete.

Figure 19.2 shows an example of such a relevance tree of causation. Changes in energy use can arise from changes in total population or from

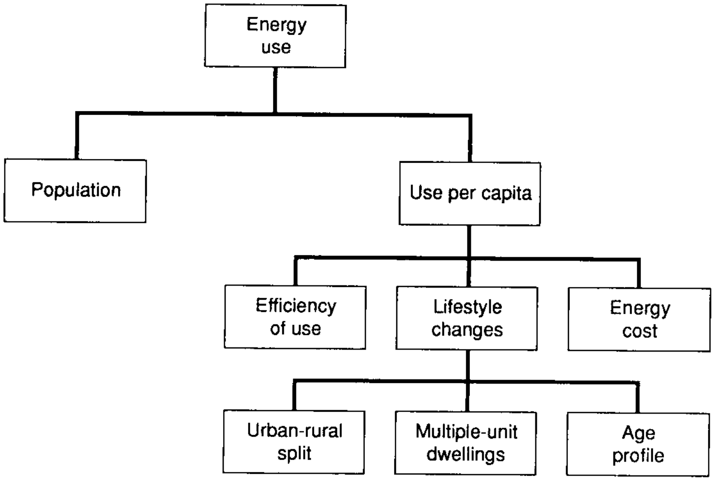

Figure 19.2 Example of relevance tree of causation.

changes in use per capita. The former can be considered a fundamental cause and need not be pursued further. The latter is affected by efficiency of use, by changes in life-style, and by energy cost. For the purposes of the example, efficiency and cost are taken as fundamental. Lifestyle changes are affected by changes in the rural to urban split, in greater or lesser use of multiple-unit dwellings, and in the age profile of the population. These can be considered fundamental causes, not needing further explanation.

Step four: What are the key elements of the subject area? Once the iteration of searching for causes of change has terminated in a set of fundamental causes, it is important to identify those that are the key elements. If the same causes appear several times at the bottom of different branches of the relevance tree, these can be considered key causes. Their effects are transmitted through multiple branches to the top of the tree. In addition, there may be causes that appear only once but have such a strong influence that they become key elements. The influence of each fundamental cause at the bottom of the tree as well as those items identified as important causes of change in the topic to be forecast must be determined.

Interrogation for relevance

This step focuses on the forecast to be evaluated. The information generated in the two preceding steps is utilized to determine the relevance of the forecast as a whole and of its parts. This is done by comparing the forecast with the key elements identified in the preceding step.

There will be three categories of key elements: (1) relevant and included, i.e., found to be key elements and included in the forecast; (2) relevant but omitted, i.e., found to be key elements but absent from the forecast; and (3) irrelevant but included, i.e., not relevant to our needs but present in the forecast.

The presence of irrelevant material does not necessarily indicate a fault in the forecast. The forecast may have originally been prepared for some other purpose or for a broader topic area. The elements that are irrelevant can simply be ignored. They do no harm other than to clutter up the forecast.

Elements that are relevant but not included must still be sought out. These are items of information that are still needed. Without them the forecast does not provide complete information and is to this extent inadequate as a basis for decision.

Elements that are both present and relevant are then subjected to the next step in the interrogation model.

Interrogation for reliability

The information that was both relevant and included in the forecast must be evaluated for reliability. This process involves examining past data, the assumptions, and the methods connecting past data with estimates of the future. These three features are inevitably intertwined, but we will consider them separately.

1. *The methods.* The methods used to produce the forecast must be identified. In some cases, they will be described clearly; this is usually the case when models or explicit procedures are used. In other cases, the methods will not be described at all and only the results will be given. For each result, the method must either be identified or categorized as not given. When the method is not given, the forecast cannot be evaluated for reliability. Hence, some results may have to be classified as having unverifiable reliability. When the methods are given, two further questions are possible.
   a. *Are the methods replicable?* A replicable method is one that can be used by someone other than the original forecaster to obtain the same result. For instance, fitting a Pearl curve to a set of data is replicable. Assuming the work is done correctly, every forecaster should get the same results. If the methods are not replicable (e.g., a Delphi forecast), the forecast is not necessarily wrong. It simply means that the reliability of the methods cannot be verified (however, see below about supporting the credibility of experts).
   b. *Are the methods formally consistent?* Once a method has been determined to be replicable, it must be examined to determine whether it is logically consistent. Does the conclusion follow from the premises? Have any logical errors or fallacies been committed? A logical fallacy will make any further conclusions unreliable. These questions group conclusions into three categories: (1) verified as arising from reliable methods; (2) verified as unreliable; and (3) of unverifiable reliability. This last category indicates an area where further information must be

sought. The results in the second category must be rejected. Results in the first category can be examined further.

2. *The assumptions.* Assumptions are used in place of facts when the facts are unknown or are inherently unknowable. The assumptions, then, are critical to the forecast. The following questions should be asked about the assumptions, in the order given.
   a. *What are the assumptions?* In many cases, the assumptions are not clearly listed. Some may be hidden. In some cases, the forecaster might not have been consciously aware of all the assumptions; hence, ferreting them out may require a careful examination of the forecast. Some of the kinds of assumptions found in forecasts are as follows: that a particular method is appropriate in a specific case; that a given group will behave in a particular way (like another group or like they did in the past); and that data from a sample are representative of the entire population.
   b. *Are the assumptions adequately defined?* Before an assumption can be tested for validity, it must be adequately defined. The definition must specify the conditions to which the assumption does and does not apply and the evidence that is necessary for affirmation or denial. The task of defining assumptions can be made easier by the following questions.
      (1) *Is this a static or a dynamic assumption?* Static assumptions refer to states or conditions. They must specify the time at which they are valid. Dynamic assumptions involve a rate, an acceleration, a sequence, a delay, or a causal linkage. They must specify a dimension of change and an interval of time during which the change takes place. A static assumption, valid for one point in time, may not be reliable if applied to some other point. A dynamic assumption, valid for one time interval, may not be reliable if applied to some other interval.
      (2) *Is this a realistic or a humanistic assumption?* Realistic assumptions concern phenomena essentially independent of human opinion. These may include facts about the physical universe or the past actions of humans. A realistic assumption must specify the time and place to which it applies. Humanistic assumptions refer to values, perceptions, future actions, intentions, goals, etc., of human beings. A humanistic assumption must specify the group or individual to which it applies, since it may not be true of other groups or individuals. Assumptions of worth, cost, or risk must specify to whom they apply.
   c. *How are the assumptions supported?* An assumption for which no support is offered is of unverifiable reliability. If support is offered, reliability may be determined by the following steps.
      (1) *Is the assumption a necessary consequence of some law, principle, or axiom?* If so, how is this law, principle, or axiom supported? Is the law so well accepted that no support is needed? If not, is evidence or data offered in support?

(2) *Is evidence offered in direct support of the assumption?* This evidence may be examples of other cases in which the assumption was valid or other reasons why the assumption should be accepted.
(3) *Is the assumption supported by the opinion of an expert in the field?* If so, is evidence offered in support of the expert's credibility?

d. *If evidence is offered, is that evidence relevant to the law, expert, or assumption it is to support?* Is the evidence of the type demanded by the nature of the assumption? Is it an adequate test of the assumption? The following criteria can be used to evaluate the relevance of the evidence:
(1) *Evidence to support or deny a static assumption.* The evidence must be obtained from the particular state or point in time specified by the assumption.
(2) *Evidence to support or deny a dynamic assumption.* The evidence must be gathered sequentially in time or process. The frequency of observation and the time span over which the observations are taken must be consistent with the nature of the assumption.
(3) *Evidence to support or deny a realistic assumption.* The evidence must be gathered, processed, and presented in a manner as free from human evaluation and prejudice as possible. This is important since what people do or do not believe is irrelevant to matters of fact. The inclusion of human evaluations only serves to reduce the reliability of the purported facts.
(4) *Evidence to support or deny a humanistic assumption (e.g., concerning the values or perceptions of some person or group).* The evidence must be obtained from the physical, written, or verbal behavior of the person or group in question. Different groups of people may interpret the same reality in different ways and have different ideas about what is desirable or undesirable. The only valid means for determining what people believe or perceive is through the examination of their behavior. In particular, it is necessary to avoid the error of attributing one's own values to others or assuming that they perceive reality in the same way.
(5) *Evidence to support or deny a law or principle.* The evidence must be gathered under the circumstances in which the law or principle is assumed to apply. If the evidence is gathered under some other set of circumstances, it cannot be used either to prove or disprove the validity of the law in the case considered in the forecast.
(6) *Evidence to support or deny the credibility of an expert.* The evidence must deal with his or her degree of knowledge regarding the case under consideration in the forecast. The expert's knowledge or lack of knowledge about some other cases is not relevant.

If the evidence offered turns out not to be relevant to the assumption, the reliability of the assumptions cannot be determined. If the evidence is relevant, it must then be checked for validity.

3. *The data.* Data may be used to support an assumption or as raw material from which a pattern is to be extracted (e.g., by some mathematical fitting procedure). By this point in the interrogation process, only data that are relevant either to an assumption or to a forecasting method should have survived the screening. For both kinds of evidence, accuracy is an important issue.
   a. *How accurate is the data?* There are two possible sources of error in data: uncertainties of observation or measurement, and conscious or unconscious distortion.
      (1) *Are there uncertainties of observation?* These are inherent in any observation technique. In observations of physical phenomena, the uncertainties are usually known. In the social sciences, they may be harder to evaluate. However, in either case it is necessary to estimate the uncertainties to determine whether the data are sufficiently accurate.
      (2) *Are there uncertainties of measurement?* These uncertainties arise from aggregation, rounding, etc., and from sampling from a large population. Again, it is necessary to estimate these uncertainties to determine whether the data are sufficiently accurate.
      (3) *Is there conscious distortion?* Have the observations been distorted or selected to fit preexisting human values? The agency that gathered the data must be identified, and its past actions examined to identify possible biases. If these are found, the data would be suspect.
      (4) *Is there unconscious distortion?* This may arise from the philosophy, culture, or ideology of the observer. It may originate in factors such as "self-evident truths" of a culture, the unchallenged axioms of a scientific discipline, or the implicit assumptions of a common philosophy. Unconscious distortion can show up in both the selection and the classification of data, and it is often difficult to identify, especially when the forecast evaluator shares the unconscious assumptions of the data gatherers. However, it is important that this distortion be identified, especially when the data concern people who might not share the unconscious assumptions.

      At this point, evaluation is complete for data to be used for fitting models, extracting patterns, and so on. However, one additional question must be asked about data intended to support assumptions.
   b. *Does the evidence presented tend to confirm or deny the assumptions?* Assumptions about future conditions can never be denied or confirmed absolutely on the basis of available evidence. The evaluator must use judgment to determine whether to accept or reject an assumption on the evidence available. Even assumptions about the past cannot always be confirmed absolutely, since some of the evidence may have been lost or destroyed or it may cost more than it is worth to gather "all" the data. Thus, given the data available, the evaluator must de-

termine whether it is sufficient to support or deny an assumption or whether the assumption must be evaluated as having indeterminable reliability.

This completes the description of the interrogation model. In summary, the model leads the evaluator through four steps:

1. *Interrogation for need:* The decision to be made is analyzed to determine what kind of forecast is required, including what should be forecast, how accurately it must be forecast, and over what time range.

2. *Interrogation for underlying cause:* Elements that have been identified in the previous step as required parts of the forecast are analyzed to identify causes of change.

3. *Interrogation for relevance:* The forecast itself is analyzed to determine how much information it contains about the factors previously identified as affecting the things to be forecast.

4. *Interrogation for reliability:* Those items found to be included and relevant are evaluated for reliability.

By asking this structured sequence of questions on the subject and the given forecast, the evaluator is led through an evaluation of the relevance and reliability of the latter. A forecast that is highly relevant to the subject in question and highly reliable (i.e., draws sound conclusions from the best data available) is very useful as an input to a decision. This is not an assertion that the forecast will come true. Instead, it is an assertion that the decision maker is likely to do better by taking the forecast into account than by ignoring it.

If the forecast turns out to be irrelevant or unreliable or both, then the decision maker is well advised to reject it. If the matter is sufficiently important and the available forecast is not sufficiently relevant or reliable, the decision maker is well advised to initiate the preparation of another, more useful forecast.

19.3 Summary

The interrogation model has been presented in terms of its use by a decision maker to evaluate a forecast; however, it can also be used by forecasters. Forecasters should anticipate that their work will be evaluated by some means similar to the interrogation model. They should therefore plan their work from the beginning so that it will successfully pass the evaluation. A forecast should be focused on the needs of the user, incorporating all the relevant sources of change. The methods, assumptions, and data should be clearly spelled out in such a way that their reliability is verifiable. By applying the interrogation model to their own work while preparing a forecast, forecasters can increase the utility of their work and the likelihood that it will be accepted and used.

References

Bierce, A., *The Sardonic Humor of Ambrose Bierce,* Dover Publications, Inc., New York, 1963.

Problems

1 Why is it desirable to be able to evaluate a forecast before it is used, rather than waiting until the actual outcome is known?

2 Carry out an interrogation for underlying cause on each of the following situations and develop a relevance tree of causation for each:

a Number of automobiles in use in the United States
b Number of first-class letters handled by the post office
c Total demand for passenger air travel in the United States

3 The year is 1895. You are the U.S. Secretary of Agricultures. You have asked your staff to prepare a 30-year forecast of American farm production and acreage. The portion of this hypothetical forecast dealing with the acreage planted to oats is given below. Using the methods of this chapter, evaluate this forecast, carrying out all steps of the interrogation model. To what extent does your analysis exhibit 1895-style foresight and to what extent does it exhibit hindsight?

> Our forecast of the demand for oats is based on the following considerations: The primary market for oats is as horse feed. We assume this will continue to be the case. Most farm workhorses receive a diet containing a mixture of feeds, primarily hay. However, the average farm horse consumes 31 bushels of oats annually. We assume this will continue to be the case. Since 1867, the average yield of farmland in oats has been about 26 bushels per acre. We assume this will continue to be the case. The number of horses has grown as shown in the following table, which represents an excellent fit to an exponential growth law. We assume the exponential growth will continue. Projecting this growth until 1925, there will be 53 million horses in the U.S. This will require 63 million acres planted to oats.

Year	Number of horses on U.S. farms, thousands
1867	6280
1868	7051
1870	7633
1875	9333
1880	10,356
1885	12,700
1890	15,732
1895	17,849

SOURCE: *Historical Statistics of the United States,* Department of Commerce, Washington, D.C., 1976.

Chapter

20

Presenting the Forecast

20.1 Introduction

After preparing a forecast based on one or more of the methods described earlier in this book, after having taken into account the application of the forecast as described in one of the chapters on applications, and after having made sure to avoid the pitfalls and mistakes described in the immediately preceding chapters, there is still one more hurdle the forecaster must get over: The forecast must be accepted by a decision maker and used in the decision-making process.

This is the final step and the reason for all the work that came before. If the forecast is not accepted and used by a decision maker, all the work and resources that went into preparing it were wasted. No matter how interesting the work was to the forecaster, how elegant the procedures that were used, or how extensive the compilation of data that went into the forecast, it was all wasted effort if the forecast does not contribute to the making of an appropriate and timely decision.

The forecaster must remember that this criterion reflects the viewpoint of the decision maker, who is not a professional forecaster. All the professional niceties that mean so much to forecasters and their colleagues are of no consequence to the decision maker. It does the forecaster no good to complain that the decision maker "just didn't understand what I was trying to get across." This may well be true, but the fault is more likely that of the forecaster than that of the decision maker.

It may be that in some cases the forecaster will encounter a decision maker who rejects the concept of using rational and explicit forecasts as inputs to decisions. There are some decision makers of this type still in responsible positions. Their number will grow fewer, however, as death and retirement take their toll, and it is unlikely that such a decision maker would have called on a forecaster for help in the first place. The progressive decision maker accepts all the help available in clarifying a tough decision problem. If a technological forecaster is asked to contribute to the decision-making process, it is because the decision maker appreciates the benefits that can be derived from a forecast of technological change. If one of these decision makers rejects a forecast after having asked for it, the fault is usually that of the forecaster. No matter how well the forecast was prepared, in some way it was not presented properly.

There are two basic reasons why a forecast will not be used. The decision maker does not believe it to be useful or does not believe it to be credible. In this chapter, we will look at some of the problems of presenting a forecast to a decision maker so that its utility and credibility are apparent.

20.2 Making the Forecast Useful

A forecast prepared by a technological forecaster is stated in terms of some functional capability, i.e., that a specific functional capability will (or could) have achieved a specified level at a definite time in the future. A graph, such as that shown in Fig. 20.1, with historical data points and a best-fit curve projected to some future date must never be interpreted as though the projected line somehow lies in the future that has not yet arrived. The curve shown in the figure was implicit in the historical data and was extracted by a mathematical fitting technique. It is connected with the future growth of technology only by an assumption about the relationship between past and future behaviors, an assumption that forms the basis for the forecast. The point is that the forecaster could simply have presented the decision maker with a tabulation of the historical data without a trend extraction or projection. The sole justification for the activity of extracting and projecting the trend implicit in the data is that it is more useful to the decision maker than a tabulation of the historical data would be. The forecast is justified only to the extent that it is useful to the recipient, providing information that could not have been obtained as readily (or at all) from the historical data alone, and that it is recognized as being of help in making a decision.

The decision maker will see the forecast as being useful to the extent that is meaningful and understandable. If the forecast is incomprehensible, it is, of course, of no help. Even if it is understandable, it must still be meaningful in terms of a specific decision to be made. If the forecast does not illuminate the

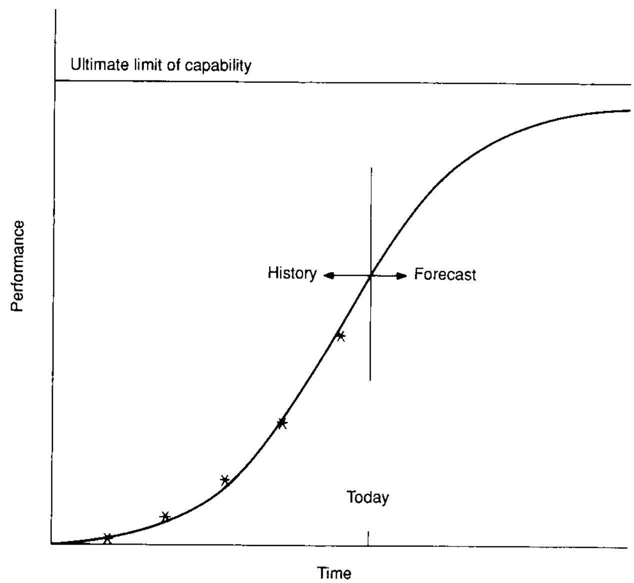

Figure 20.1 Forecast showing trend extracted from historical data.

alternatives in a specific decision, leading to a choice of actions, it is not meaningful. These two properties of a good forecast will be discussed separately below.

Meaningful forecasts. Former President Lyndon Johnson is reputed to have responded with the query, "Therefore, what?" following a briefing on the Middle Eastern situation (Hare, 1970). To the extent that a decision maker responds to a forecast with such a query, the forecast has not been meaningful. It has failed to get at whatever the decision maker considers to be the important issues. It may have addressed the important issues, at least as seen by the forecaster, but this fact is not conveyed to the decision maker.

This points out the importance of tailoring the forecast to the needs of the decision maker in a specific situation. In other words, the forecast must be pertinent to the type of decision being made. How much control does the decision maker have over the technology being forecast? Will the decision affect the progress of that technology to some degree (great or small)? Or is the decision maker trying to choose the appropriate response to a technological change resulting from the decisions of others?

Depending on the type of decision situation, the forecaster prepares either a conditional or an unconditional forecast. The forecaster states either that a certain event is going to come to pass regardless of what the decision maker chooses, i.e., "Other people are developing this; you have to adapt to it," or conversely that the forecast is conditional upon the choice made by the decision maker, i.e., "You can achieve this outcome if you choose this particular course of action."

This issue can be especially important in forecasting the future growth of some technology. If the decision maker's organization has played a major role in establishing the trends that led to the achievement of the current level of functional capability in the area concerned, then the organization cannot drop out and assume that the trends will continue. On the other hand, if the organization has not contributed to the progress of that technology in the past, then the forecast will be based solely on assumptions about the behavior of others who have been contributing to its progress.

Consider, for instance, the following basic (and hypothetical) forecast: By the year 2000, half of all business travel will be replaced by long-distance video teleconferencing. If this forecast were prepared for a decision maker in a telephone company, it would have to be a conditional forecast, i.e., it would have to indicate that it was conditional on the decision maker taking the proper actions. In that case, the forecast might be phrased, "If we continue installation of new equipment at the present rate, by the year 2000, half of all business travel will be replaced by long-distance video conferencing."

If, however, the forecast were being prepared for a decision maker in an airline company, it might be phrased, "At the present rate of improvement, by the year 2000, long-distance video conferencing will replace half of the business travel that is an important part of our market." From the standpoint of the recipient, this forecast is unconditional; no action has to be taken to see it fulfilled. It is important that the forecaster state explicitly whether or not the forecast is conditional on certain actions undertaken by the decision maker receiving the forecast. In particular, it must be clear whether the decision maker's organization can simply go along for the ride or whether they must take some positive action to fulfill the forecast.

Another situation where even a well-prepared forecast can turn out to be meaningless from the decision maker's standpoint is that of the implied superiority of a technique or technical approach. Consider, for instance, the (hypothetical) forecast that low-cost composite materials (plastic and fiber) will be used in (1) secondary and primary commercial building structures by 1998 and (2) secondary and primary structures for highway bridges, airfield construction, etc., by 2000.

It is not clear from the statement of the forecast whether the forecaster is saying that this capability will simply be available, or whether he or she is saying not only that it will be available but will be superior to all competing techniques. Unless the decision maker knows which is the case, he or she does not know what actions to take, if any at all. For instance, suppose this forecast were given to decision makers in the cement industry. They would naturally be concerned with the possibility that plastic and fiber composites might replace concrete in some structural applications. However, they need to know whether composites will be superior to concrete, and for what applications, before they can determine whether any action at all is needed, and if so, what actions would be appropriate. Hence, it is important that the forecaster make this point clear in the presentation of the forecast.

From time to time, the forecaster will find it appropriate to forecast diminishing returns for a given technology, i.e., the rate of progress will decline significantly during the period covered by the forecast. It is not enough, however, simply to state a forecast of diminishing returns. The cause must be identified, and the implications of the fact must be made clear. Some common causes of diminishing returns in the progress of a technology are the following:

1. A physical upper limit is imminent.
2. Progress is slowed by standardization, which puts significant constraints on possible new approaches.
3. The demand for the technology decreases; hence, less effort is expended on advancing it.
4. A breakdown in communications occurs because the field has become very large, with many people working in it.
5. Continued growth at the previous rate requires excessive resources.

There are other possible reasons in specific cases, but the foregoing cover a large proportion of the instances of diminishing returns. The forecast can be useful to the decision maker only if it presents the specific cause for diminishing returns.

Not only is the cause of the diminishing returns important but so are the implications. What must be done if the previous rate of progress is to be restored? If the previous rate of progress cannot or should not be restored, what other actions should be considered? Some possible decision implications are the following:

1. Organizational changes to permit more rapid progress
2. More resources to allow continued rapid progress
3. Shift to another (specific) technology capable of more rapid progress (i.e., move over to a new growth curve that has already been identified)
4. Search for a new technical approach, not yet identified, that is not faced with an imminent limit
5. Abandonment of the existing technology, since the decline in the rate of progress is fundamental and cannot be reversed

Note that the statement of these implications does not force a choice by the decision maker. It only indicates what must be done if the previous rate of progress is to be restored. The decision maker is still free to decide whether or not to restore the previous rate and to choose the most appropriate action under the circumstances, i.e., the forecast does not force a particular decision. However, an intelligent choice cannot be made unless the decision maker knows why diminishing returns have set in and what can be done to restore the situation, or adapt to the changed situation. Hence, it is important that the forecaster not stop with a forecast of diminishing returns but explain why the situation has arisen and what its implications are.

The points made above regarding meaningful forecasts have all tacitly assumed that the functional capability being forecast would be meaningful to the decision maker and that the forecaster's only concern is to present the forecast in such a way that the possible decisions are clearly illuminated. However, this is not always the case. Consider, for instance, a decision maker in a bank who is concerned with planning the bank's future purchases of computers. The decision maker wants to maintain adequate computer capability over the planning period but not be stuck with too many obsolescent computers at some future date. Therefore, near-term purchases, which provide immediate capability, have to be balanced against waiting for better computers. Suppose a forecast of computer capability is presented in terms of technical parameters such as the ratio of memory size to access time. Unless the decision maker is a specialist in computers, this forecast is not going to be helpful at all. The forecast should be presented in terms of functional parameters that are meaningful to the user. In the case of a bank official, the forecast should be presented in terms of bank operations such as the number of transactions (deposits, withdrawals, etc.) to be handled each day. This situation is in fact fairly general. There are many technologies where the functional capability to be forecast can best be expressed in some technical measure that is not likely to be meaningful to a decision maker who is not a specialist in the technology. In such cases, it is necessary to carry out the forecast in the appropriate technical parameters, then present the findings in the functional parameters meaningful to the decision maker.

One of the best ways to render a forecast meaningless is to make it qualitative instead of quantitative. A forecast that states, "The performance of this class of devices will increase significantly," "This technique will be used more extensively," or "This device will be used in a wide variety of applications," is not very helpful. Even if the forecast comes true, it is useless because it is so vague. Besides, the decision maker is probably already aware that performance will be improved and new applications will be found. The decision maker expects the professional forecaster to provide information that is not already known. In particular, this should include a quantitative estimate of the improvement in performance or the extent of use or application, as well as an estimate of the uncertainty in the forecast. Quantitative estimates of the nature and degree of change help a decision maker make a more appropriate and timely decision.

One situation that sharply outlines the problems of making a forecast meaningful is that of making several successive forecasts for the same date. For instance, suppose an organization has established the practice of maintaining a 10-year plan based on a 10-year forecast. Each year the forecast is redone completely, covering a 10-year period. Year 19*XX*, which was the tenth year the first time the forecast was prepared, becomes the ninth year after the first annual revision, the eighth year after the second annual revision, etc. No matter what forecasting method is used, each version of the forecast will have a somewhat different description of conditions in the year 19*XX*. Suppose, for instance, that a growth curve is fitted to the performance of a technical approach to show how it will move toward its ultimate upper limit. The calculated upper limit should remain the same year after year (unless the theory on which it was calculated is found to be wrong). As additional data points are collected, however, and included in the regression fit computations, a slightly different curve will be produced each year. (The only exception will be the unlikely circumstance where the new data point falls precisely on the curve computed the previous year.) What meaning should be attached to these changes in the forecast? Is new action or a reversal of previous action required of the decision maker as a result of the "revised" forecast? Clearly if the underlying process has not changed and the original decision was the appropriate one for that underlying process, then no changes are required. Hence, when this situation occurs, the presentation should make clear whether the difference is simply the result of statistical fluctuation from the additional data point or whether it represents a departure from the previous forecast, and if it is the latter, what is the reason for the change. The use of confidence limits can be helpful here by showing whether the new forecast is within the X percent confidence bounds on the previous forecast. (Note that this should be the confidence bounds on the fitted curve itself, not on the scatter about the fitted curve.) The percentage of confidence should be chosen to display clearly whether or not the change is significant. If there is some reason to believe that the underlying process has in fact changed, the forecaster is obliged to stress this fact. In this case, there has been a significant change in the forecast, and the decision maker may be called upon to alter a previous decision. The same is true if there is a significant change in the calculated upper limit. The forecaster is obliged to make this fact clear.

The important point is that when a forecaster is required to produce a sequence of forecasts for the same point in time, a forecasting method or a means of presentation or both should be chosen that plays down unimportant variations or statistical fluctuations in the forecast and emphasizes significant differences that call for a change in the course of action being pursued. If the forecast is to be meaningful to the decision maker, the forecaster must show clearly whether the differences between the succeeding forecasts are significant or not.

Understandable forecasts. The decision maker should not be in any doubt about what future the forecaster is forecasting. The events in that future and their uncertainties and the qualifications (the ifs, maybes, possiblys, etc.) must be laid out clearly. The decision maker, after hearing or reading the forecast, should not be left wondering, "Now, what did the forecaster say?"

One of the biggest blocks to understanding is ambiguity. The forecaster inevitably deals with uncertainty and is, therefore, frequently tempted to make ambiguous forecasts. If this is done with the intent of hedging against being caught with an incorrect forecast, it is reprehensible. The forecaster's obligation is to be useful, even if this means risking being wrong. However, producing an ambiguous forecast is often done innocently. The forecaster is not certain what is going to happen, does not want to leave out any reasonable possibilities, finds it difficult to assign relative likelihoods to various possibilities, and, therefore, ends up with a forecast that is ambiguous. This forecast will then not be understood by the decision maker, who may then fail to take appropriate action.

In her study of the Japanese attack on Pearl Harbor in December of 1941, Roberta Wohlstetter (1962) describes a series of messages sent to General Short and Admiral Kimmel, the Army and Navy commanders, respectively, at Pearl Harbor, during the year prior to the attack. These messages were based on information obtained by decoding Japanese messages to their own military forces and to their diplomatic officials in various parts of the world. Wohlstetter summarizes these messages as follows:

> Except for the alert order of June 17, 1940, none of the messages used the word "alert." They said, in effect, "Something is brewing," or "A month may see literally anything," or "Get ready for a surprise move in any direction—maybe."

> Phrases characterizing contingencies as "possible" or "barely possible" or "strongly possible" or saying that they "cannot be entirely ruled out" or are "unpredictable" do not guide decision-making with a very sure hand. To say, for example, in the Army warning of November 27, 1941 that Japanese future action was unpredictable added nothing at all to the knowledge or judgment of General Short or Admiral Kimmel. Local pressures being what they were, General Short read a danger of sabotage into his final messages from Washington, while Admiral Kimmel read danger of attack by Japan on an ally, or even on an American possession somewhere near the Hawaiian Islands.

This example illustrates clearly the problems that can arise if a forecast is worded ambiguously. Even though the United States was in possession of decoded Japanese messages that provided reasonably clear signals that the Japanese were planning some offensive action, the value of this information was lost because the forecasts based on it, and supplied to the decision makers on the spot, were not worded unequivocally.

The forecaster is engaged in decoding signals not *from* the future but *about* the future. This process is not done for its own sake but to provide information for decisions. The results of this decoding process should be presented as unambiguously as possible. When uncertainty is inevitable, the forecaster must estimate the degree of uncertainty as well as possible and present the forecast in such a way that the decision maker is aware of this uncertainty. Ambiguity should not be used to hide uncertainty.

One of the best means of presenting a forecast is the use of graphs. Quantitative forecasts lend themselves well to presentation in this fashion. Growth curves and trend curves, when presented graphically, can be grasped and understood readily by the decision maker whose time is limited; they are easy to explain. Even intuitively derived forecasts, such as those obtained from a Delphi panel, can be presented in graphical form if the forecaster has obtained estimates of several levels of capability or estimates of the dates by which a technology will achieve different stages of innovation. Figure 20.2 illustrates a graphical display of functional capability. Figure 20.3 illustrates a graphical display of Delphi estimates of achievement of various stages of innovation. Graphical display also lends itself well to the display of degrees of uncertainty, whether these be in the form of calculated confidence limits or simply intuitive estimates.

Some items lend themselves well to tabular display, especially forecasts obtained for a single point in time. Table 20.1 illustrates a hypothetical forecast of structural materials for aircraft in the year 19*XX*. Each of the properties of the material is forecast separately in terms of its degree of improvement over the present level of functional capability. This format can be grasped readily by the decision maker who is pressed for time. It focuses on the essentials, presents these in quantitative terms, and describes the implications of the forecast. As shown, the table is an unconditional forecast for a decision maker whose actions will not influence the outcome. In a conditional forecast, it might be desirable to replace the final column with a description of the actions required by the decision maker if the forecast is to be fulfilled.

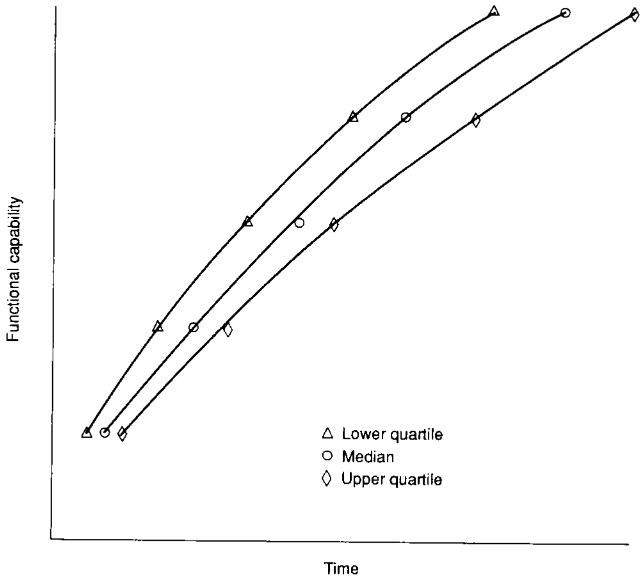

Figure 20.2 Graphical display of Delphi estimates.

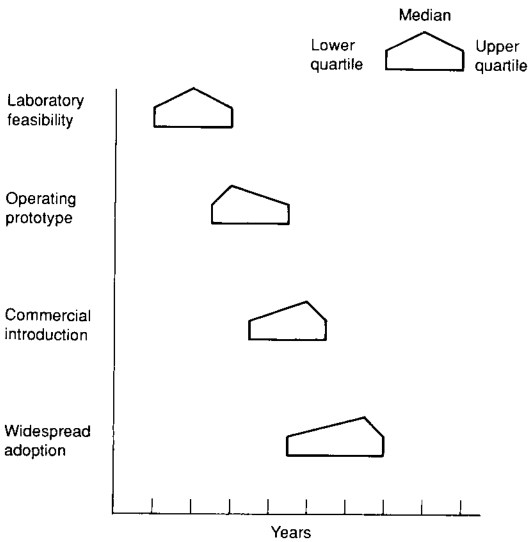

Figure 20.3 Graphical display of Delphi estimates of discrete events.

Another very useful method for making a forecast understandable is the scenario. This method is particularly applicable when several related but independently produced forecasts are to be combined into a composite one. In such cases, the decision maker is concerned not with discrete events or indi-

TABLE 20.1 Hypothetical Forecast of Aircraft Structural Materials for 19XX.

Property	Level of capability	Implication
Yield strength	75 percent increase	Lighter structure for the same load-carrying capability
Corrosion resistance	20 percent increase	Longer life in the same corrosive environment
Low-temperature performance	Same as current	No change
High-temperature performance	40 percent increase in allowable temperature for short exposure; 10 percent increase for long exposure	Higher speed; lighter structure at same speed
Fatigue life	50 percent increase	Longer service life for airframes; longer period between major overhauls

vidual technologies, but with a broad picture of the future. The scenario can provide a context in which the various options open to the decision maker can be placed so as to determine which one is the most satisfactory on an overall basis.

20.3 Making the Forecast Credible

A prerequisite that a forecast must meet if it is to have any impact on the decision-making process is that it be credible. The decision maker to whom it is presented must believe the forecast, but not, of course, in the sense that it is certain to come true. (Progressive decision makers are well aware of the problems of uncertainty and also of the fact that their own actions may have some impact on the outcome of the situation being forecast.) The forecast must be believed in the sense that the forecaster appears to have extracted all or most of the relevant information from the historical data and has connected this with the future through an appropriate logical structure. In short, the decision maker must believe that the forecast presents the best picture of the future that can be obtained from existing knowledge.

Unfortunately, many forecasts that represent the good use of existing knowledge and the sound application of forecasting methods, are rejected because the way they are presented robs them of credibility. The decision maker is always short of time. He or she knows that only part of all the desired decision information is available and that the time available to study the information is less than he or she would like. In addition, the decision maker knows that arguments for and against each of the available courses of action will be presented. These arguments are often presented by people who have an interest in a particular course of action, and therefore each argument is calculated to make their preferred course of action look good. Finally, these arguments will usually be couched in technical terms or will deal with technical matters. The decision maker does not always have the technical competence to judge these arguments in detail and is aware of this. He or she, therefore, looks for quick means of assessing the validity of arguments or the credibility of people presenting the arguments. For instance, the decision maker may take the attitude, "If I know something about the field in which this person claims to be an expert, something that the individual ought to know but doesn't, I'll disregard everything the expert says." Moreover, such an attitude cannot be faulted. It is simply self-defensive, a "quick and dirty" means of sorting out which alleged experts should be believed and which should be ignored. Nevertheless, the existence of such an attitude on the part of the decision maker means that the work that went into a forecast can go for naught if the forecaster destroys his or her own credibility or that of the forecast through a mistake in presentation. Many of these mistakes do not really reflect on the quality or utility of a forecast and are, therefore, avoidable if the forecaster has done a good job up to that point. The most common of these credibility-destroying mistakes will be discussed in this section.

One such mistake is to present the forecast in a narrative instead of a quantitative form. A page of prose that boils down to the statement that things are good and getting better is of no value. The decision maker needs to know in quantitative terms how good things are now and how good they are expected to be at some definite time in the future. Furthermore, there should be no need to hunt through paragraphs of high-sounding irrelevancies to find the few nuggets of quantitative information hidden there. A decision maker who is confronted with large quantities of prose narrative, even if it is based on a well-done forecast and contains some useful statistics, will most likely reject the whole thing on the basis of lack of credibility. If a good job has been done, the forecaster should have no difficulty presenting the forecast in a compact and quantitative form. To do otherwise would look like a cover-up for a shoddy job.

Even when the forecaster has presented the forecast in graphical form, it still may not be credible. There are three common failures in the graphic method of presentation that can cause a decision maker to doubt the credibility of the forecast, and perhaps even that of the forecaster. In fact, if the forecast is well done in the first place, these credibility-destroying mistakes need not appear. Hence, the appearance of one or more of them is a clue that the forecast was poorly done.

The first of these mistakes with graphics is shown in Figure 20.4. Here the forecaster has apparently extrapolated from a single point. If this is actually the case, the forecast is worthless. If an adequate time series of historical data has been used, this data should appear on the graph. Omitting it in the mistaken notion that doing so unclutters the graphs and allows more room for the forecast (i.e., the "future" portion of the curve, for which there is no data as yet) is to invite the suspicion that there is no historical data backing up the forecast. A decision maker spotting a graphical display such as Fig. 20.4 cannot be blamed for suspecting the worst and rejecting the forecast. Therefore, a

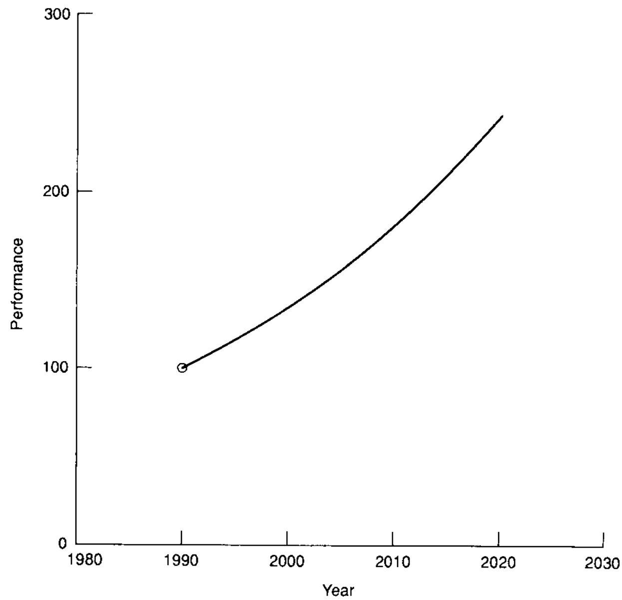

Figure 20.4 Forecast apparently extrapolated from a single point.

graphical presentation should include the historical data as well as the best-fit curve extrapolated into the future.

The second of the mistakes is the "northeast curve," illustrated in Fig. 20.5. Here the forecaster has included some data points in the lower left corner of the graph but has covered them with a fuzzy band going generally northeast across the page. Sometimes a wavy line is used instead of a fuzzy band, but the effect is the same. There is no indication of how the band was calculated,

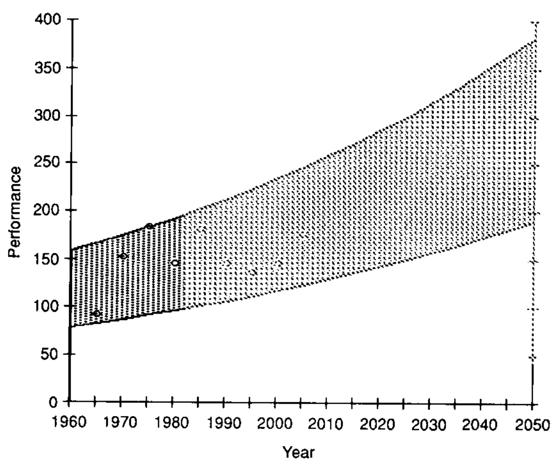

Figure 20.5 Forecast showing northeast curve.

nor what level of confidence is to be associated with its upper and lower limits. The decision maker is fully justified in assuming that the fuzzy band represents a freehand drawing that is absolutely unreliable beyond the region covered by the historical data. If a curve is fitted by some appropriate means, or if confidence limits are calculated, they should be shown and labeled clearly, and the method of curve fitting should be indicated.

The third of these mistakes is "loss of nerve," illustrated in Fig. 20.6. Here the forecaster has fitted a trend to a set of data. extended this trend into the future, and then indicated a halt to progress. If there is some fundamental limit to the possible level of functional capability that will bring progress to a halt, it should be indicated and explained. However, if the forecast is simply presented as shown in Fig. 20.6, the decision maker is very likely to assume that the forecaster suffered a loss of nerve and would not project a continuation of an obvious trend. This may well lead to the rejection of the entire forecast.

Any of these mistakes, if actually representative of the way the forecast was prepared, is the graphical equivalent of a narrative forecast. Despite their quantitative appearance, the graphs provide no information at all and will be of no value to the decision maker. If they are used in an attempt to cover up a shoddy job of forecasting, this fact will undoubtedly be detected. If they are used by mistake, they can easily destroy the credibility of a well-prepared forecast.

Even if the forecaster avoids the mistakes described here, there is one more way of running into trouble. A situation that frequently occurs is that of anomalous data. In the set of historical data used to prepare the forecast there may be one or more points that do not seem to belong. This situation is illus-

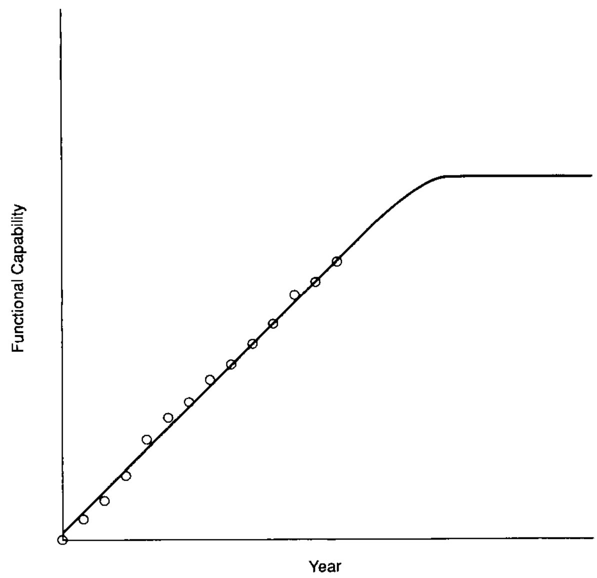

Figure 20.6 Forecast apparently showing loss of nerve.

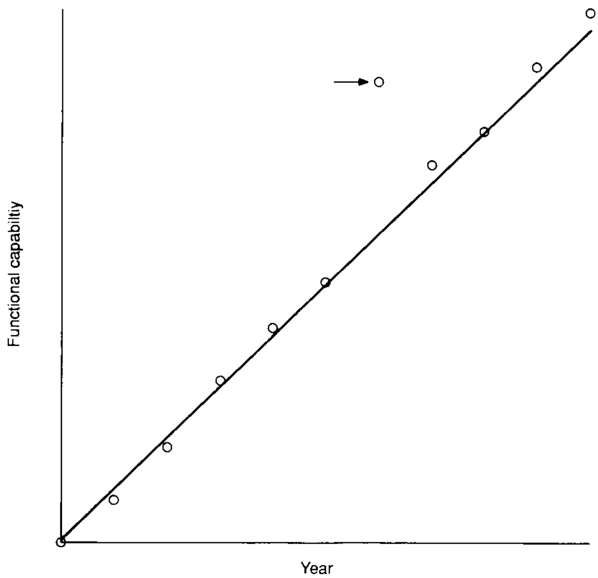

Figure 20.7 Forecast containing an apparently anomalous data point.

trated in Fig. 20.7, where the arrow indicates one point deviating markedly from the rest. The mistake, of course, is not in finding the anomalous data point or even in including it in the historical time series. The mistake is in not explaining it to the decision maker. A forecaster should never grind out a mathematical fit to the historical data without examining the anomalous data points. There is no magic in mathematical fitting techniques that can make up for bad or anomalous data; hence, the forecaster should determine whether these data points do belong with the rest. If so, they should be used, since they represent real deviations that might recur. If they do not belong, they should be eliminated from the calculations of the trend fit. If the forecaster fails to explain the anomalous data points, the decision maker may assume that a mathematical fitting procedure was applied blindly, without examining the data that went into the mathematical machinery. The following are some possible reasons for points appearing "too early":

1. They represent unsuccessful freaks.
2. They are suspect because of inadequate documentation of the conditions of measurement.
3. They represent unexploited breakthroughs.
4. They were "ahead of their time" for some reason not likely to be repeated.

Similarly, points may appear to be "too late." Some of the reasons for this are the following:

1. They represent development projects that were delayed for nontechnological reasons.
2. They represent developments that did not have to meet stringent performance requirements and therefore did not fully utilize the then-existing state of the art.
3. They represent the use of an older and well-tested technology in order to satisfy extreme reliability requirements (e.g., amplifiers for undersea telephone cables often use technology that is well behind the state of the art but has demonstrated that it will operate without failure for many years).

If one of these reasons applies, or if there is some other reason for the existence of anomalous data points, the explanation should be given to the decision maker.

Another category of mistakes that destroys the credibility of a forecast is that of the apparent avoidance of responsibility. Whereas the decision maker rightly does not want the forecaster to make decisions for him or her, the decision maker does want the forecaster to take the problem seriously. The decision maker wants to feel that the forecaster has a personal interest in the correctness of the decision being made and, therefore, would attempt to make the forecast as specific and useful as possible. If the forecaster attempts to avoid any responsibility for the forecast or for the correctness of the decision, the forecaster is not living up to his or her professional responsibilities. If the forecaster makes the mistake of appearing to try to avoid responsibility, the forecaster destroys the credibility of his or her forecast.

One of the ways of appearing to avoid responsibility is to take the attitude, "That's what the analysis showed." If the forecast appears unrealistic in some sense or seems to contradict common-sense notions of appropriateness, the decision maker may ask the forecaster to explain the peculiarities of the forecast. For the forecaster to respond to the effect that these are the results the analysis produced appears irresponsible. The forecaster cannot simply give the data, describe the forecasting technique used, and present the results. The historical data should have been selected on the basis that it is a true representation of what actually happened in the past. Similarly, the forecasting method should have been selected on the basis that it is the most appropriate one for the situation. The results should have been examined to see if they appear reasonable. If the forecaster conveys the impression that the forecast was simply a mechanical routine of applying a method to a set of data without any judgment involved, the decision maker cannot be blamed for refusing to accept the forecast.

Another way of appearing to avoid responsibility is to use multiple sets of assumptions. For instance, let us consider some process that is affected by both a growth rate and a decay rate. Population is one such process; the total number of people alive at some time in the future will depend on the birth and death rates in the intervening period. Similarly, the number of items of a technological device in use at some time in the future will depend on the production rate and the rate of failure or scrappage in the intervening period. When a process is governed by a birth rate and a death rate, the forecaster can assume different values for each rate. If two different growth rates and two different decay rates are assumed, a computer can be made to produce four separate and distinct forecasts. This in itself is not a mistake. However, the forecaster is obliged to choose the rates on a reasonable basis, i.e., they appear to be the most likely rates, they represent upper and lower bounds on likely rates, or some such rationale. For the forecaster to say to the decision maker, "Here are several combinations of rates; you choose the one you like," appears irresponsible. The forecaster is likely to be better informed than the decision maker on which rates or ranges are most likely. Therefore, the decision maker should not have to be responsible for selecting the appropriate rates on which to base a forecast.

The same problem appears in the more general form of multiple assumptions. If the forecaster produces a range of forecasts, each based on a different set of assumptions, but makes no attempt to indicate their respective likelihoods, the decision maker may suspect that the forecaster is trying to avoid responsibility for the forecast. Thus, the mistake of presenting several forecasts based on multiple assumptions without indicating their relative likelihoods can destroy the credibility of the forecast.

It should be noted that there are cases where it is appropriate to present the decision maker with forecasts based on multiple assumptions. One case is when the decision maker is better informed about the various rates of growth or decay, for instance, than the forecaster. This situation is highly unlikely to arise in practice, however, since in such a case the decision maker is unlikely to need a forecaster. A more common situation occurs when the decision maker has some control over the value that will be taken by one of the factors. In such a case, the decision maker's real interest is in the effect of exercising that control. Then it is entirely appropriate for the forecaster to say, in effect, "If you do such and such, the results will be so and so." If all the relevant factors are outside the control of the decision maker, however, the forecaster must judge the likelihood of the various assumptions or risk seeing the forecast rejected as not credible.

A final way of appearing to avoid responsibility is to present a forecast with an escape hatch. The most common way of doing this is to use progressive tenses in making statements, for example, "The rate of R&D expenditures in this field is rising," "The number of engineers in this field is increasing," "Progress in this field is continuing at a steady pace." These may all sound like forecasts, at least by implication, but they are not. They are all statements *only* about the present or the past. The implication is that things will continue as they have been or are. However, the forecaster does not actually say that they will. If they do not, the escape hatch is that the statement was true when it was made. It is not enough for the forecaster to observe that some trend is rising. He or she is also obliged to determine whether it will continue to rise and at what rate. If the forecaster makes a forecast on the basis of previous trends, the forecaster must state whether, in his or her professional opinion, these trends will continue. If the forecaster gives the appearance of leaving an escape hatch, he or she cannot complain if the decision maker rejects the forecast as lacking credibility. Hence, if the forecaster has done the job properly, the conclusions should be stated boldly so as not to make the mistake of appearing to leave a way out should the forecast turn out to be seriously wrong.

The final type of mistake that can diminish the credibility of a forecast is the omission of the data on which it is based. If a forecast is prepared in written form, the data should be included in an appendix or final section so that the continuity of the presentation is not broken. If the forecast is presented "live," the data should be readily available in case they are requested. In any case, it is good practice to include the data on which the forecast is based. From the standpoint of the forecaster, including the data saves the need to look it up the next time he or she makes a forecast in the same subject area. More importantly, however, from the standpoint of concern here, it increases the credibility of the forecast. It makes clear the degree of effort to which the forecaster has gone in order to make the forecast valid and useful. In practice, if the forecaster has done a good job, it would be a mistake to detract from the credibility of the forecast by omitting the data.

There is yet another consideration regarding the inclusion of data. If the forecaster has rejected a historical data point for legitimate reasons, it should nevertheless be included in the historical data tables with an explanation of why it was not used. If the data point is omitted, the forecaster runs the risk that the decision maker will happen to be aware of that very item and take its omission as an indication of sloppy historical research. Hence, it is a mistake to omit the historical data that was rejected for good reason. Including it, along with the reasons for not using it in the forecast, will increase the credibility of the forecast, since this serves as an indication of the thoroughness with which the historical research was carried out.

20.4 A Checklist for Forecasts

The two preceding sections have concentrated on mistakes that the forecaster is subject to and that may render good work useless. Avoiding these mistakes, which are common and serious, can be very helpful in making the presentation of a forecast effective. However, it is probably better to take a more positive approach to ensure that the forecast as a whole meets certain minimum criteria. Such a set of criteria is presented in this section.

These criteria are presented in the form of questions. They are based on the interrogation model presented in Chap. 19. They represent questions that the alert decision maker will (or at least should) ask about a forecast. The decision maker may not voice these questions explicitly, but they will at least implicitly be in his or her mind. If the way the forecast is presented answers these questions satisfactorily and the mistakes discussed in the previous section are avoided, then the forecaster may reasonably expect that the forecast will be accepted and used in the decision process.

1. *Why is this forecast needed?* In answering this question the forecaster must address the decision to be made and the alternatives open to the decision maker. The uncertainties that exist regarding each of the alternatives must also be identified. Then the relevance of the forecast to these alternatives and uncertainties must be demonstrated, which will mean showing that the forecast covers the appropriate subject area as well as the time range of interest and has the correct scope.
2. *What portion of the needed decision information is contained in the forecast?* First of all, the technological forecast never provides all the information needed for a decision. There are always considerations other than that of technological change that also bear on the decision. These may include, for instance, demographic or economic change. Hence, the technological forecaster is called upon to provide only certain portions of the needed information. Therefore the presentation of the technological forecast should identify the following:
   a. Which portions of the total set of decision information is provided?
   b. What portions must be obtained from some other source?
   c. How does the latter interact with the technological forecast?
   d. Which portions should be provided by a technological forecast but are not contained in the one being presented? There may well be legitimate reasons for failure to include all the needed information. For instance, the forecast may be only an interim report on work not yet completed, in which case, all the information may not yet be available. It may be a forecast that was originally prepared for some other purpose and that will be supplemented by a specially prepared forecast. In any case, answering this question places the forecast in proper context. It shows how the forecast fits in with the other information available to the decision maker and how it provides information not available from any other source.
3. *What method was used to prepare the forecast?* The decision maker is not usually concerned with the technical details of the forecasting method used other than to be aware of its strengths and weaknesses. The decision maker will want to know whether the method is really appropriate to the situation, whether a better method was available but not used, the kinds of failures the method is prone to, and the type of surprises that lie in store should the forecast based on the method fail. The decision maker is also interested in the replicability of the method. For instance, if the forecast was obtained from a panel of experts, the decision maker may be concerned that another panel of equally eminent experts might produce a totally different forecast. In the forecast presentation, the forecaster should attempt to satisfy the decision maker on these points by showing that the most appropriate method or combination of methods has been used or that the fore-

cast represents the consensus of the best experts available. In the latter case, it is also appropriate to indicate whether there was a strong minority view, and the consequences should the minority view turn out to be right. If the method is not an intuitive one, the forecaster should show that an adequate sample of historical data was obtained and that the method has worked satisfactorily for similar situations in the past.

4. *What are the assumptions on which the forecast is based?* One of the assumptions behind the forecast is the logical structure connecting the past with the future. This will already have been covered in point three above. Other assumptions may include estimates of numerical values for which adequate data are not available; assumptions about the behavior of other significant actors involved in the situation or about the continuation (or change) of certain governmental or organizational policies; other forecasts used as inputs to the technological forecast, such as economic or demographic forecasts; and assumptions about the success or failure of ongoing actions when these can have an impact on the events being forecast. If the assumptions behind the forecast are specifically identified, the decision maker will be able to evaluate them more readily. In particular, if a specific assumption seems incorrect, it can be changed and the forecast modified appropriately. When the forecaster does not specifically identify the assumptions and the decision maker has doubts, there may be no choice but to accept or reject the forecast as a whole.
5. *Why should the assumptions be accepted?* Whenever an assumption is used in place of a fact or in place of certain knowledge, presumably the opposite assumption, or at least a different one, could also be used. The forecaster must show that the assumptions used in the forecast are more reasonable, more plausible, more likely, or in some way better than alternative assumptions that might have been used. For instance, the forecaster might be able to place bounds, through physical considerations, on the possible value of some quantity. If an assumption is made about the future behavior of some person or group, it might be demonstrated that the assumption is consistent with their past behavior. If another forecast is used as input, it may be shown that forecasts from that source have a good batting average in the past. In short, the forecaster must show that the full range of possible assumptions has been examined and that the choice made from among them is based on the preponderance of the best evidence available.
6. *What are the sources of information and data?* The decision maker may be legitimately concerned that the forecast may be biased, deliberately or inadvertently, and that it may be invalid because of inappropriate data. By indicating the source of the forecast's data, the forecaster can allay some or all such apprehension. If (1) the data sources are authoritative, have no known bias, or have a known bias that can be discounted; if (2) the data are relevant in the sense that they have been gathered from the appropriate time periods, individuals, or groups; and (3) the data are complete, then the decision maker is more likely to accept the forecast. Thus, if the forecaster has done a good job of assembling the data for use in the forecasting

method or for use in validating the assumptions of the forecast, the decision maker's confidence in the forecast can be increased by identifying the sources of the data.

The questions in this checklist represent questions that any decision maker should want answered about a forecast before basing a decision on it. Likewise, they are questions that any forecaster should be prepared to answer with regard to a forecast. If the forecaster has done a good job, a presentation that answers these questions clearly should result in the acceptance and use of the forecast.

20.5 Summary

The ultimate test of a forecast is whether it is accepted and used in the decision-making process. If it is rejected by a decision maker, the work that went into it was wasted.

Even when a forecast is well done, it may fail to be used because of its presentation. Hence, it is important that the forecaster present the forecast in such a way that the quality of the work and the utility of the results will be readily apparent. In particular, the forecast must be useful and credible to the decision maker. If the latter does not see its utility or does not believe in it, it will not be accepted.

From the standpoint of utility, the forecast must be meaningful and understandable. The decision maker must see it as having meaning with respect to the decision to be made and must be able to understand it. The forecast must be more meaningful and understandable than the raw data that went into it. Graphical presentations can be very helpful in making the forecast understandable. However, there are a number of mistakes the forecaster can make that will render the forecast either meaningless or incomprehensible. If the forecaster is aware of these mistakes, they can be avoided more easily.

From the standpoint of credibility, the decision maker must believe that the forecast presents the best picture of the future that can be obtained on the basis of available data about the past and knowledge of how the past is related to the future. Again, there are a number of common mistakes that can be made in presenting a forecast that destroy its credibility. Again, if the forecaster is aware of these mistakes, they can be avoided more readily.

From an overall standpoint, there are several questions that the decision maker will have about the forecast. These may be explicit or implicit. In either case, the forecaster should make sure ahead of time that the presentation of the forecast will answer these questions. A checklist can be used to determine whether the presentation is satisfactory. If the presentation adequately answers all the questions in the checklist, it should be convincing to the decision maker.

Since the presentation of the forecast represents the last hurdle in the process of getting it accepted and used, the forecaster should give it adequate attention. Considerable time and effort have been expended in preparing the forecast. If the presentation is not convincing, all this work will go down the drain. The forecaster should, therefore, be willing to spend adequate time and effort in preparing a convincing presentation. By following the suggestions in this chapter, the chances of making a forecast convincing can be increased.

References

Hare, F., "How Should We Treat Environment," *Science,* **167**:352–355 (1970).
Wohlstetter, R., *Pearl Harbor: Warning and Decision,* Stanford Univ. Press, Stanford, CA, 1962.

For Further Reference

Botez, Mihai C., "Anticipating Decision Makers Attitudes: A Methodological Suggestion," *Technological Forecasting & Social Change,* **19(3)**:257–264 (1981).

Problems

1 Consider the following forecasts. Under what circumstances would their presentation as given here be a mistake, and what would be the nature of the mistake?

a By the year 2000, most of the skin panels of aircraft will be made of high-strength, high-modulus composite materials consisting of strong fibers embedded in a binder.

b By the year 1998, most takeoffs and landings of commercial aircraft will be computer controlled, with the pilot serving as a monitor only.

c By the year 2005, space satellites will be used for manufacturing plants in order to take advantage of microgravity, high vacuum, and abundant solar power.

2 You have obtained the following data in the table below for the growth in the functional capability of a specific technology. What is your forecast of this technology out to year 45 if there is no imminent limit to this growth and no radical changes seem likely in the industry that produces the technology? Prepare your forecast for presentation to a decision maker who has no influence on the progress of the technology.

Year	Capability	Year	Capability
2	105	15	223
3	122	20	259
6	142	21	300
11	165	24	349
14	192	29	406

3 A prospective decision depends on the level of functional capability to be achieved by five related technologies. The historical data on each of them are shown in the table below. You have been asked to present forecasts for each for years 60 and 70. You know of no reason for any of them to deviate from their past patterns of change. Prepare a forecast for each of them. In what form can your forecasts be presented most effectively?

	Technology				
Year	A	B	C	D	E
2	14	62	210	340	3000
5	24	64	190	420	3500
10	30	86	250	430	2600
14	54	130	230	540	2900
20	70	180	355	550	2200
23	110	185	360	600	2100
29	140	300	360	860	2400
34	280	360	480	800	1700
39	345	475	520	1100	2000
42	540	640	580	1200	1500
48	730	910	630	1400	1600

Appendix A

Historical Data Tables

Table	Subject
A.1	U.S. Electric Power Production
A.2	Low Temperatures Achieved in the Laboratory
A.3	Speed Trend of U.S. Military Aircraft
A.4	Transport Aircraft Characteristics
A.5	U.S. Population Since 1790
A.6	Illumination Efficiency
A.7	Solid-Propellant Rocket Engine Performance
A.8	Digital Computer Performance
A.9	Gross Weight of U.S. Single-Place Fighter Aircraft
A.10	Characteristics of Integrated-Circuit Gates for Electronic Computers
A.11	Speeds Achieved by Experimental Rocket Aircraft
A.12	Number of Papers Published on Masers and Lasers
A.13	Steam Engine Efficiencies
A.14	Power-Generation Efficiency of Public Utilities
A.15	Telephones per U.S. Population
A.16	Conversion of U.S. Merchant Marine from Sail to Mechanical Power
A.17	Conversion of U.S. Merchant Marine from Wood to Metal
A.18	Percentage of U.S. Dwellings with Electric Power
A.19	Energy Consumption from Various Sources
A.20	Deep Versus Surface Coal Mining
A.21	Power Output of UHF Transistors
A.22	Maximum Thrust of Liquid-Propellant Rocket Engines
A.23	Automobiles in the United States
A.24	U.S. Space Payloads Injected into Orbit
A.25	Characteristics of STOL and VTOL Fixed-Wing Aircraft

TABLE A.1 U.S. Electric Power Production

Year	Production, $\times 10^6$ kW · h
1945	271,255
1946	269,609
1947	307,400
1948	336,808
1949	345,066
1950	388,674
1951	433,358
1952	463,055
1953	514,169
1954	544,645
1955	629,010
1956	684,804
1957	716,356
1958	724,752
1959	797,567
1960	844,188
1961	881,495
1962	946,526
1963	1,011,417
1964	1,083,741
1965	1,157,583
1966	1,249,444
1967	1,317,301
1968	1,436,029
1969	1,522,757
1970	1,639,771
1971	1,718,000
1972	1,853,000
1973	1,965,000
1974	1,968,000
1975	2,003,000
1976	2,125,000
1977	2,212,000
1978	2,286,000
1980	2,286,000
1981	2,295,000
1982	2,341,000
1983	2,310,000
1984	2,416,000
1985	2,470,000
1986	2,487,000
1987	2,572,000
1988	2,704,000
1989	2,779,000

SOURCE: *Statistical Abstract of the United States,* Department of Commerce, Washington, D.C.

TABLE A.2 Low Temperatures Achieved in the Laboratory

Year	Experimenter	Temperature, K
1721	Fahrenheit	273
1799	de Morveau	240
1834	Thilorier	216
1837 (approximately)	Faraday	163
1883	Wroblewski and Olszewski	137
1885	Wroblewski and Olszewski	121
1885	Wroblewski and Olszewski	48
1898	Dewar	20.4
1899	Dewar	14
1908	Kamerling-Ohnes	4
1909	Kamerling-Ohnes	0.8
1933	Giauque and de Haas (independently)	0.25
1933	de Haas, Wiersma, and Kramers	0.085
1934	de Haas and Wiersma	0.031
1935	de Haas and Wiersma	0.0044
1949	De Klerk	0.0029
1950	DeKlerk	0.001
1952	Kurti	0.00002

SOURCES: *Encyclopedia Britannica*; Asimov, *Biographical Encyclopedia of Science and Technology*; *National Bureau of Standards Circular* 519 (1951); Garret, *Magnetic Cooling*, 1954; Burton, Smith, and Wilhelm, *Phenomena at the Temperature of Liquid Helium*.

TABLE A.3 Speed Trend of U.S. Military Aircraft

Year of first delivery	Airplane	Maximum speed, mi/h
1909	Wright Brothers B	42
1916	Curtiss JN-4	80
1918	Nieuport 27 C.1	110
1918	Spad XIII C.1	135
1921	Boeing MB-3A	141
1924	Curtiss PW-8	161
1925	Curtiss P-1	163
1927	Boeing PW-9C	153
1929	Curtiss P-6	180
1929	Boeing P-12	171
1933	Boeing P-26A	234
1934	Martin B10-B	212
1937	Seversky P-35	381
1938	Boeing UB-17	256
1938	Curtiss P-36A	300
1939	Curtiss P-40	357
1940	North American B-25	322
1940	Bell P-39C	379
1941	Martin B-26	315
1941	Republic P-43	350
1942	Republic P-47D	420
1942	North American P-51A	390
1943	North American P-51B	436
1945	Lockheed P-80A	573
1946	Republic XP-84A	619*
1948	North American F-86A	671*
1950	Boeing B-47A	600
1953	Convair F-102A	860
1954	McDonnell F-101C	1200
1956	Convair B-58	1330
1958	Lockheed F-104A	1410*
1961	North American B-70	1800
1965	Lockheed SR-71	1800
1967	Convair F-111	1900

*World special record.

SOURCES: J. C. Fahey, *U.S. Army Aircraft 1908–1946*, Ships & Aircraft, New York, 1946: and *Aviation Week & Space Technology*, annual inventory issues.

TABLE A.4 **Transport Aircraft Characteristics**

Year	Speed, mi/h	Payload, t	Number of passengers	Aircraft
1925	95	NA*	NA	Fokker F-IV
1926	111	NA	11	Ford 4-AT-B
1927	116	NA	NA	Fokker Trimotor
1928	148	NA	13	Ford 5-AT-B
1933	161	NA	NA	Curtis T-32
1933	200	NA	10	Boeing 247D
1934	192	NA	10	Lockheed 10 Electra
1934	213	NA	14	Douglas DC-2
1934	225	NA	8	Vultee V-1A
1935	220	2.8	21	Douglas DC-3
1938	228	NA	33	Boeing 307B
1940	241	NA	36	Curtis C-46
1940	272	NA	14	Lockheed Lodestar
1940	275	8.6	42	Douglas DC-4
1941	264	NA	NA	Curtis CW-20
1946	329	NA	64	Lockheed 649
1946	NA	NA	NA	Martin 202
1947	315	15.5	63	Douglas DC-6
1947	347	4.7	40	Convair CV-240
1948	375	20.5	134	Boeing C-97A
1949	375	NA	86	Boeing 377
1950	312	NA	40	Martin 404
1950	370	20.0	92	Lockheed 1049
1951	314	NA	44	Convair CV-340
1952	265	87.0	200	Douglas C-124C
1954	409	NA	95	Douglas DC-7
1955	360	18.3	92	Lockheed C-130A
1955	400	12.3	95	Douglas DC-7C
1956	377	16.0	92	Lockheed 1649A
1956	300	25.0	NA	Douglas C-133A
1958	579	47.6	189	Douglas DC-8
1958	450	10.8	98	Lockheed 188 Electra
1959	300	45.4	NA	Douglas C-133B
1959	597	48.4	189	Boeing 707-320B
1959	615	11.3	110	Convair 880
1959	622	21.6	140	Boeing 720B
1961	600	44.5	126	C-135
1963	571	35.4	154	Lockheed C-141A
1963	632	14.7	114	Boeing 727
1968	550	110.0	NA	C-5A
1968	1550	16.5	140	Tupolev TU-144
1969	640	110.0	490	Boeing 747
1969	1354	14.0	100	Concorde
1970	575	53.2	270	Douglas C-10
1970	610	48.5	256	Lockheed 1011
1972	570	38.1	336	Airbus 300

*NA, not available.
SOURCE: *Aviation Week & Space Technology,* annual forecast and inventory issues.

TABLE A.5 U.S. Population Since 1790

Year	Population
1790	3,929,214
1800	5,308,483
1810	7,239,881
1820	9,638,453
1830	12,866,020
1840	17,069,453
1850	23,191,876
1860	31,443,321
1870	39,818,449
1880	50,155,783
1890	62,947,714
1900	75,994,575
1910	91,972,266
1920	105,710,620
1930	122,775,046
1940	131,669,275
1950	150,697,361
1960	178,464,236
1970	203,849,000
1971	206,076,000
1972	208,088,000
1973	209,711,000
1974	211,207,000
1975	212,748,000
1976	214,446,000
1977	216,058,000
1978	217,874,000
1979	219,699,000
1980	227,757,000
1981	230,138,000
1982	232,520,000
1983	234,799,000
1984	237,001,000
1985	239,279,000
1986	241,625,000
1987	243,934,000
1988	246,329,000
1989	248,239,000
1990	248,710,000

SOURCE: *Statistical Abstract of the United States,* Department of Commerce, Washington, D.C.

TABLE A.6 Illumination Efficiency

Year	Source	Illumination efficiency, lm/W
1850	Paraffin candle	0.1
1879	Edison's first lamp	1.6
1892	Acetylene lamp	0.7
1894	Cellulose filament	2.6
1901	Mercury arc	12.7
1907	Tungsten filament	10.0
1913	Inert gas filled	19.8
1928	Sodium lamp	20.0
1935	Mercury lamp	40.0
1942	Fluorescent lamp	55.0
1962	Gallium arsenide diode	180.0

SOURCE: *Encyclopedia Britannica,* 1964.

TABLE A.7 Solid-Propellant Rocket Engine Performance

Year	Engine or vehicle	Thrust, lb	Impulse, $\times 10^6$ lb · s
1956	Sergeant	50,000	1.19
1958	Algol	105,000	3.625
1959	Polaris	62,000	10.0
1961	Minuteman first stage	155,000	10.0
1964	AGC FW-1	NA*	35.0
1962	AGC FW-4	300,000	40.0
1964	TV-412	1,500,000	170.0
1965	Titan III	1,100,000	90.0
1965	AGC-260 (120 in)	3,500,000	380
1967	AGC-260	5,800,000	840
1969	7-Segment (120 in)	1,000,000	140.0

*NA, not available.

TABLE A.8 Digital Computer Performance

Year	Computer	Memory capacity, bits per access time (μsec)	Matrix inversion time, min
1946	Eniac	NA*	5.0
1951	Edvac	0.0011	7.5
1951	Univac I	0.0011	5.2
1951	MIT Whirlwind	0.0011	0.028
1952	IAS Johnniac	NA	1.1
1953	IBM 701	0.013	0.22
1953	Univac 1103a	0.040	0.71
1955	IBM NORC	0.014	0.10
1956	IBM AN/FSQ-7 (SAGE)	0.0390	0.047
1956	IBM 704	0.11	0.20
1958	Philco 2000/210	0.17	0.074
1959	IBM 7090	0.50	0.059
1960	Philco 2000/211	1.2	0.047
1960	CDC 1604	0.50	0.038
1960	Univac LARC	0.93	0.015
1961	IBM Stretch	3.4	NA
1963	Univac 1107	0.93	0.045
1963	Philco 2000/212	2.39	0.021
1964	CDC 6600	8.3	0.0013
1965	Univac 1108	12.5	0.0079
1966	IBM 360/67	17.7	0.012
1967	B-8500	24.5	NA
1967	CDC 6800	31.6	0.00030

*NA, not available.
SOURCE: Dr. Robert U. Ayres, Carnegie-Mellon University, personal communication.

TABLE A.9 Gross Weight of U.S. Single-Place Fighter Aircraft

Year	Aircraft	Gross weight, $\times 10^3$ lb
1918	Nieuport 27 C.1	1.5
1919	Spad XIII C.1	2.0
1921	Boeing MB-3A	2.5
1924	Curtiss PW-8	3.2
1925	Curtis P-1	2.8
1927	Boeing PW-9C	3.2
1929	Curtis P-6	3.2
1930	Boeing P-12B	2.6
1933	Boeing P-26A	3.0
1937	Seversky P-35	5.6
1938	Curtiss P-36A	6.0
1939	Curtiss P-40	7.2
1940	Bell P-39C	7.2
1941	Lockheed P-38	15.3
1941	Republic P-43	7.8
1942	Republic P-47D	14.5
1942	North American P-51	9.0
1943	North American P-51B	11.8
1943	Bell P-63A	10.0
1945	Lockheed P-80A	11.7
1946	Republic YP-84	16.5
1948	North American F-86A	13.8
1952	Republic F-84F	25.0
1953	North American F-86F	17.0
1953	North American F-100	28.0
1954	McDonnell F-101	49.0
1956	Convair F-102	27.0
1958	Lockheed F-104	22.0
1958	Republic F-105	48.0
1959	Convair F-106	35.0
1963	McDonnell F-4*	54.0
1967	Convair F-111A*	91.0

*Two-place aircraft.

SOURCE: J. C. Fahey, *U.S. Army Aircraft 1908–1946,* Ships and Aircraft, New York, 1946; *Jane's All the World's Aircraft* (various editions); *Aviation Week & Space Technology,* inventory issues.

TABLE A.10 Characteristics of Integrated-Circuit Gates for Electronic Computers

Year	Technology	Delay, ns	Power, mW	Speed · power product, pJ
1968	ECL-III	1.1	60	66
1970	S/TTL	3	20	60
1971	ECL-1000	2	25	50
1972	LA/TTL	10	2	20
1973	NMOS	100	0.1	10
1975	IIL	10	0.1	1
1977	eweic	0.25	2	0.5

SOURCE: *Computer Design,* 69 (July 1978).

TABLE A.11 Speeds Achieved by Experimental Rocket Aircraft

Aircraft	Date	Speed, mi/h
X-1	October 14, 1947	703.134
D-558 II	August 7, 1951	1238
	November 20, 1953	1328
X-2	July 1956	1900
X-15	September 17, 1959	1350
	August 4, 1960	2196
	April 21, 1961	3074
	May 15, 1961	3307
	June 23, 1961	3603
	November 9, 1961	4093
	June 27, 1962	4159

TABLE A.12 Number of Papers Published on Masers and Lasers

Year	Masers*	Lasers*	Optical pumping*	Light and coherence*	Total*
1957	3	—	—	—	3
1958	37†	—	—	—	37
1959	34	—	—	—	34
1960	66	—	21	—	87
1961	78	—	30	—	108
1962	53	115	37	—	205
1963	69	305	37	—	411
1964	64	655‡	59	—	778
1965	66	730	56	47	899
1966	75	1197	58	61	1391
1968	52	1330	124	97	1603

*The number of papers listed in *Physical Abstracts* for years shown.
†1958 and prior, listed as electromagnetic oscillations.
‡1964 and prior, listed as optical masers.

TABLE A.13 Steam Engine Efficiencies

Year	Steam engine	Efficiency
1698	Savery	0.5
1712	Newcomen	1.3
1770	Watt	2.8
1796	Watt	4.1
1830	Cornish	10.0
1846	Cornish	15.0
1890	Triple expansion	18.0
1910	Parsons turbine	20.0
1950	Steam turbine	30.0
1955	Steam turbine	38.0

SOURCE: Thirring, *Energy for Man,* Indiana University Press, Bloomington, IN, 1958.

TABLE A.14 Power-Generation Efficiency of Public Utilities

Year	Pounds of coal per kilowatt · hour	Kilowatt · hours per pound of coal
1920	3.05	0.328
1925	2.03	0.493
1930	1.60	0.625
1935	1.44	0.694
1940	1.34	0.743
1945	1.30	0.769
1950	1.19	0.840
1955	0.95	1.053
1960	0.88	1.136
1965	0.858	1.166
1970	0.909	1.100
1975	0.952	1.050
1976	0.948	1.055
1977	0.969	1.032
1978	0.987	1.013
1979	0.979	1.021
1980	0.9761	1.0245
1981	0.9872	1.0130
1982	0.9927	1.0073
1983	0.9875	1.0127
1984	0.9833	1.0170
1985	0.9824	1.0179
1986	0.9802	1.0202
1987	0.9727	1.0281

SOURCE: *Statistical Abstract of the United States*, Commerce Department, Washington, D.C.

TABLE A.15 Telephones per U.S. Population

Year	Number of telephones per 1000 individuals
1876	0.1
1880	1.1
1885	2.7
1890	3.6
1895	4.3
1900	17.6
1905	48.8
1910	82.0
1915	103.9
1920	123.4
1925	144.6
1930	162.6
1935	136.4
1940	165.1
1945	198.1
1950	280.8
1955	337.2
1960	407.8
1965	478.2
1966	498.7
1967	518.3
1970	584
1975	695
1976	718
1977	744
1978	769
1979	793

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.16 Conversion of U.S. Merchant Marine from Sail to Mechanical Power

Year	Tons sail, $\times 10^3$	Tons power, $\times 10^3$	Power, %
1820	1258	22	1.72
1830	1127	64	5.37
1840	1978	202	9.27
1850	3010	526	14.88
1860	4486	868	16.21
1870	2363	1075	31.27
1875	2585	1169	31.14
1880	2366	1212	33.87
1885	2374	1485	38.48
1890	2109	1758	45.46
1895	1965	2213	52.97
1900	1855	2658	58.90
1905	1962	3741	65.60
1910	1655	4900	74.75
1915	1384	5944	81.11
1920	1272	13,823	91.57
1925	1125	14,976	93.01
1930	757	13,757	94.78
1935	441	12,535	96.60
1940	200	11,353	98.27
1950	82	28,327	99.71
1955	40	26,792	99.85
1960	23	23,553	99.90

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.17 Conversion of U.S. Merchant Marine from Wood to Metal

Year	Wood, $\times 10^3$t	Metal, $\times 10^3$t	Metal, %
1885	3836	430	10.07
1890	3798	627	14.16
1895	3666	970	20.92
1900	3572	1593	30.84
1905	3607	2850	44.13
1910	3391	4117	54.83
1915	3085	5305	63.23
1920	3876	12,448	76.25
1925	2907	14,499	83.29
1930	2554	13,514	84.10
1935	2185	12,469	85.08
1939	2473	12,159	83.09
1945	1915	20,898	94.16
1950	1952	29,263	93.74
1955	1622	28,336	94.58
1960	1397	27,184	95.11
1965	1198	25,318	95.48

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.18 Percentage of U.S. Dwellings with Electric Power

Year	Percentage
1907	8.0
1912	15.9
1917	24.3
1920	34.7
1925	53.2
1930	68.2
1935	68.0
1940	78.7
1945	85.0
1950	94.0
1955	98.4

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.19 Energy Consumption from Various Sources*

Year	Coal	Petroleum	Natural gas	Hydroelectric	Wood	Nuclear
1850	219	—	—	—	2138	—
1860	518	3	—	—	2641	—
1870	1048	11	—	—	2893	—
1880	2054	96	—	—	2851	—
1890	4062	156	257	22	2515	—
1900	6841	229	252	250	2015	—
1910	12,714	1007	540	539	1765	—
1915	13,294	1418	673	691	1688	—
1920	15,504	2676	827	775	1610	—
1925	14,706	4280	1212	701	1533	—
1930	13,639	5895	1969	785	1455	—
1935	10,634	5668	1974	831	1397	—
1940	12,535	7730	2726	917	1358	—
1945	15,972	10,110	3973	1486	1261	—
1950	12,913	13,489	6150	1440	1171	
1955	11,540	17,524	9232	1407	1037	—
1960	10,140	20,067	12,699	1657	832	—
1965	11,908	23,241	16,098	2058	577	—
1970	12,922	29,614	22,029	2650	425	200
1975	12,800	32,700	20,000	3125	—	1900
1976	13,700	35,200	20,400	2954	—	2100
1977	14,000	37,200	29,900	2300	—	2700
1978	13,900	38,000	20,000	2902	—	3000
1979	15,100	37,000	19,900	2876	—	2800
1980	15,400	34,200	20,400	3100	—	2700
1981	15,900	31,900	19,900	3100	—	3000
1982	15,300	30,200	18,500	3600	—	3100
1983	15,900	30,100	17,400	3900	—	3200
1984	17,100	31,100	18,500	3700	—	3600
1985	17,500	30,900	17,800	3400	—	4100
1986	17,300	32,200	16,700	3400	—	4500
1987	18,000	32,900	17,700	3100	—	4900
1988	18,800	34,200	18,600	2600	—	5700
1989	18,900	34,200	19,400	2900	—	5700
1990	19,100	33,600	19,400	2900	—	6200

*In trillions of Btu.

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.20 Deep Versus Surface Coal Mining

Year	Deep coal mining*	Surface coal mining*
1914	421.4	1.28
1915	439.8	2.83
1920	559.8	8.86
1925	503.2	16.9
1930	447.7	19.8
1935	348.7	23.6
1940	417.6	43.2
1945	467.6	110.0
1950	392.8	123.5
1955	343.5	121.2
1960	284.9	130.6
1965	332.7	179.4
1970	338.8	264.1
1975	292.8	355.6
1976	294.9	383.8
1977	266	425
1978	242	423

*In millions of short tons.
SOURCE: *Mineral Yearbook,* U.S. Bureau of Mines, Washington, D.C., 1980.

TABLE A.21 Power Output of Ultra High Frequency (UHF) Transistors

Year	Output power, W	Frequency, MHz
1963	1.5	500
1964	3	400
1965	20	435
1967	20	400
1975	60	960
1985	120	500
1988	200	425
1989	300	300

TABLE A.22 Maximum Thrust of Liquid-Propellant Rocket Engines

Year	Engine	Thrust, lb
1942	AL-1000(JP)	1000
1943	X35AL-6000(AJ)	6000
1945	CORPORAL-E-HW	19,000
1948	XLR-10-RM-2	20,750
1949	XLR-59-AJ-1	90,000
1952	XLR-43-NA-3	120,000
1953	XLR-71-NA-1	240,000
1956	XLR-83-NA-1	415,000
1960	XLR-109-NA-3	500,000
1961	F-1	1,500,000
1963	F-1A	1,522,000

SOURCE: R. W. Clarke, "Innovation in Liquid Propellant Rocket Technology," unpublished doctoral dissertation, Stanford University, Stanford,.CA, 1968.

TABLE A.23 Automobiles in the United States

Year	Number of automobiles, $\times 10^3$	Number of automobiles per capita
1900	8	0.000105
1905	77	0.000918
1915	2332	0.023192
1920	8131	0.076371
1925	17,481	0.150916
1930	23,035	0.187159
1935	22,568	0.177351
1940	27,465	0.208140
1945	23,793	0.194692
1950	40,334	0.266699
1955	52,136	0.317316
1960	61,559	0.342009
1965	75,258	0.388298
1970	89,200	0.437579
1975	106,700	0.501532
1976	110,400	0.514815
1977	112,300	0.519768
1978	116,600	0.535172
1979	118,400	0.526094
1980	121,600	0.533991
1981	123,100	0.535345
1982	123,700	0.532797
1983	126,200	0.538635
1984	128,100	0.542009
1985	131,900	0.553119
1986	135,400	0.562624
1987	137,300	0.565439
1988	141,300	0.576615
1989	144,400	0.583788

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.24 U.S. Space Payloads Injected into Orbit

Year	Successes	Failures	Total
1957	0	1	1
1958	7	10	17
1959	11	8	19
1960	16	13	29
1961	29	12	41
1962	52	7	59
1963	38	8	46
1964	57	7	64
1965	63	7	70
1966	73	4	77
1967	58	3	61
1968	45	3	48
1969	40	1	41
1970	29	1	30
1971	32	3	35
1972	31	2	33
1973	23	2	25
1974	24	1	25
1975	28	3	31
1976	28	0	28
1977	24	2	26
1978	32	1	33
1979	16	0	16
1980	13	2	15
1981	18	1	19
1982	18	0	18
1983	22	0	22
1984	22	0	22
1985	17	1	18
1986	6	3	9
1987	8	1	9
1988	2	0	2
1989	6	0	6
1990	5	0	5

SOURCE: *TRW Space Log,* various editions.

TABLE A.25 Characteristics of Short Takeoff and Landing (STOL) and Vertical Takeoff and Landing (VTOL) Fixed-Wing Aircraft*

Year	Aircraft	Payload, lb	Empty weight, lb	Gross weight, lb	Cruise speed, mi/h	Range, mi
1947	U-1A	1051	3341	5100	103	148
1951	U-1A	1912	4680	8000	106	250
1957	U-3A	671	3154	4830	157	425
1958	C-7A	6219	18,355	28,500	125	250
1958	U-10A	1000	2010	3400	144	330
1961	O-2A	696	3212	4850	118	340
1961	B-941S	22,045	29,675	58,422	215	310
1964	C-8A	13,843	23,157	41,000	180	250
1964	XC-142A	8000	25,610	41,500	207	276
1965	OV-10A	3600	6969	9908	230	228
1965	CL-84	4600	8685	14,500	260	115
1967	X-22A	1500	11,000	14,600	213	445
1976	YC-15	62,000	105,000	216,680	449	920
1977	XV-15	3400	9750	14,460	230	512

*Data are for design payload and design gross weight, most economical cruise speed, and range with given payload and gross weight at the given cruise speed.

TABLE A.26 Characteristics of Helicopters*

Year	Aircraft	Payload, lb	Empty weight, lb	Gross weight, lb	Cruise speed, mi/h	Range, mi
1947	UH-13	465	1435	2200	69	125
1948	HH-43B	2939	4604	9150	87	91
1949	UH-19A	1605	5013	8100	83	157
1952	CH-21A	1200	8266	10,855	87	174
1953	CH-21B	2249	8786	13,500	80	135
1953	CH-37A	7510	20,690	31,000	100	69
1954	CH-34	3980	7630	13,000	85	145
1956	UH-1A	800	3950	5864	101	72
1958	CH-46	2400	11,585	21,400	130	320
1962	CH-47A	10,367	17,878	33,000	106	108
1962	CH-54A	23,590	19,110	38,000	95	130
1963	CH-3C	2400	12,248	22,050	125	313
1964	CH-53A	8000	22,444	39,713	150	130
1965	AH-1G	1927	6073	9500	172	357
1967	CH-47C	6400	21,464	46,000	158	230
1968	UH-1N	3161	5549	10,000	115	273
1969	OH-58	866	1464	2768	114	299
1969	AH-1J	1790	6610	10,000	207	359
1974	YUH-61A	5924	9750	17,962	167	370
1974	YCH-53A	30,000	32,048	65,828	173	306
1974	YUH-60A	4077	10,624	16,825	167	373
1975	YAH-64	2284	9500	13,950	180	359

*Data are design payload and design gross weight, most economical cruise speed, and range with given payload and gross weight at the given cruise speed.

TABLE A.27 Activities of U.S. Scheduled Air Carriers (Foreign and Domestic Flights)

Year	Revenue passenger-miles	Available seat-miles	Revenue aircraft-miles
1950	10,243	16,842	477
1955	24,351	39,584	780
1956	27,625	43,674	896
1957	31,261	51,059	976
1958	31,499	53,115	973
1959	36,372	59,247	1,030
1960	38,863	65,567	998
1961	39,831	71,857	970
1962	43,760	82,612	1,010
1963	50,362	94,845	1,095
1964	58,494	106,316	1,189
1965	68,676	124,320	1,354
1966	69,889	137,844	1,482
1967	98,720	174,733	1,742
1968	113,997	216,524	2,147
1969	125,420	250,846	2,385
1970	131,719	264,904	2,418
1971	135,652	279,823	2,378
1972	152,406	287,411	2,376
1973	161,957	310,597	2,448
1974	162,919	297,006	2,258
1975	162,810	303,006	2,241
1976	178,988	322,822	2,320
1977	193,219	345,566	2,419
1978	236,900	369,000	2,520
1979	261,700	416,000	2,788

SOURCE: *Aviation Week & Space Technology*, various issues. All figures in millions.

TABLE A.28 U.S. Total Gross Consumption of Energy Resources by Major Sources[a]

	Coal							
Year	Anthracite	Bituminous and lignite	Natural gas, dry[b]	Petroleum[c]	Total fossil fuels	Hydropower[d]	Nuclear power	Total gross energy inputs[e]
				Historical Years				
1947	1224.2	14,599.7	4518.4	11,367.0	31,709.3	1459	—	33,168.3
1948	1275.1	13,621.6	5032.6	12,558.0	32,487.3	1507	—	33,994.3
1949	957.6	11,673.1	5288.5	12,120.0	30,039.2	1565	—	31,604.2
1950	1013.5	11,900.1	650.0	13,489.0	32,552.6	1601	—	34,153.6
1951	939.8	12,285.3	7247.6	14,848.0	35,320.7	1592	—	36,912.7
1952	896.6	10,971.4	7760.4	15,334.0	34,962.4	1614	—	36,576.4
1953	711.2	11,182.1	8156.0	16,098.0	36,147.3	1550	—	37,697.3
1954	683.2	9512.2	8547.6	16,138.0	34,881.0	1479	—	36,360.0
1955	599.4	11,104.0	9232.0	17,524.0	38,459.4	1497	—	39,956.4
1956	609.6	11,340.8	9834.4	18,624.0	40,408.8	1598	—	42,006.8
1957	528.3	10,838.1	10,416.2	18,560.0	40,352.6	1568	1.2	41,911.8
1958	482.6	9607.6	10,995.2	19,214.0	40,299.4	1740	1.5	42,040.9
1959	477.5	9595.9	11,990.3	19,747.0	41,810.7	1695	2.2	43,507.9
1960	447.0	9967.2	12,698.7	20,067.0	43,179.9	1775	5.5	44,960.4
1961	403.8	9809.4	13,228.0	20,487.0	43,928.2	1628	17.0	45,573.2
1962	381.0	10,159.7	14,120.8	21,267.0	45,928.5	1780	23.0	47,731.5
1963	361.0	10,722.0	14,843.0	21,950.0	47,876.0	1740	33.0	49,649.0
1964	365.8	11,295.0	15,647.5	22,385.8	49,694.1	1873	34.0	51,601.1
1965[f]	327.5	12,030.0	16,136.1	23,209.3	51,702.9	2050	38.0	53,790.9
				Projected Years				
1970	309.0	14,251.0	19,374.0	27,275.0	61,209.0	2193	874.0	64,276.0
1975	280.0	16,865.0	22,360.0	31,875.0	71,380.0	2422	1803.0	75,605.0
1980	250.0	19,290.0	25,455.0	13,978.0	80,973.0	3026	4076.0	88,075.0

[a]In trillions of Btu.

[b]Excludes natural gas liquids.

[c]Petroleum products including still gas, liquefied refinery gas, and natural gas liquids.

[d]Represents projections of outputs of hydropower and nuclear energy converted to theoretical energy inputs at projected rates of pounds of coal per kilowatt · hour at central electric stations. Excludes inputs for power generated by nonutility plants, which are included within the other consuming sectors.

[e]Gross energy is that contained in all types of commercial energy at the time it is incorporated in the economy, whether the energy is produced domestically or imported. Gross energy comprises inputs of primary fuels (or their derivatives) and outputs of hydropower and nuclear power converted to theoretical energy inputs. Gross energy includes the energy used for the production, processing, and transportation of energy proper.

[f]Preliminary data.

SOURCE: Bureau of Mines Information Circular 8364, "An Energy Model for the United States, Featuring Energy Balances for the Years 1947 to 1965 and Projections and Forecasts."

TABLE A.29 Measurements of the Velocity of Light

Experimenter	Year	Velocity, km/s	Standard deviation, km/s
Corni-Helmert	1875	299,990	300
Michelson	1879	299,910	75
Newcomb	1883	299,860	45
Michelson	1883	299,853	90
Perrotin	1902	299,901	104
Rosa-Dorsey	1906	299,784	15
Michelson	1927	299,798	22
Mittelstaaedt	1928	299,786	15
Michelson, Pease, and Pearson	1933	299,774	6
Anderson	1937	299,771	15
Huttel	1937	299,771	15
Anderson	1941	299,776	9
Bergstrand	1951	299,793.1	0.32
Mackenzie	1953	299,792.4	0.5

SOURCE: *Encyclopedia Brittanica,* 1964.

TABLE A.30 Accuracy of Time Measurement

Year	Inventor	Error, s/day
1656	Huygens—first pendulum clock	10
1721	Graham—cylinder escapement	5.5
1726	Harrison—temperature compensation	1
1761	Harrison—bimetallic strip; reduced friction	0.22
1835	Robinson—barometric compensation	0.1
1893	Reifler	0.01
1923	Shortt—free pendulum in vacuum	0.0035

SOURCE: H. T. Pledge, *Science Since 1500,* Philosophical Library, New York, 1947; and *Encyclopedia Brittanica,* 1964.

TABLE A.31 Historical Data on General Aviation

Year	Number of aircraft	Aircraft shipments	Fuel consumed, $\times 10^6$ gal	Flying time, $\times 10^6$ h
1957	65,289	6118	213	10.6
1958	67,839	6414	209	11.3
1959	68,727	7689	221	11.9
1960	76,549	7588	246	13.0
1961	80,632	6778	257	13.4
1962	84,121	6697	264	14.0
1963	85,088	7569	285	14.8
1964	88,742	9336	307	15.4
1965	95,442	11,852	378	16.2
1966	104,706	15,747	486	18.9
1967	114,186	135,677	541	21.6
1968	124,237	13,698	610	22.9
1969	130,806	12,457	690	29.8
1970	131,743	7,283	759	26.0
1971	131,148	7466	734	25.5
1972	145,010	9774	977	27.0
1973	153,540	13,645	786	30.0
1974	161,502	14,165	886	31.4
1975	168,475	14,057	907	32.0
1976	178,300	15,447	1021	33.9
1977	184,300	16,920	1176	35.8
1978	198,700	16,456	NA*	39.4

*NA, not available.

SOURCES: *Aerospace Facts and Figures*; *Aviation Week & Space Technology*; and *Statistical Abstract of the United States*, Bureau of the Census, Washington, D.C.

TABLE A.32 National Income and Telephone and Radio Usage in 42 Countries

Country	GNP per capita	Telephones per capita	Radios per capita
Australia	7472	0.41	0.739
Austria	8522	0.324	0.291
Belgium	10,751	0.316	0.414
Bolivia	745	0.023	0.096
Brazil	1015	0.041	0.147
Canada	8459	0.617	0.997
Chile	740.7	0.0206	0.0766
China	360	0.0052	0.0469
Colombia	646	0.0540	0.112
Costa Rica	1598	0.0687	0.189
Cuba	894	0.0330	0.216
Czechoslovakia	406.1	0.189	0.259
Denmark	699	0.016	0.324
El Salvador	699	0.016	0.321
Finland	7013	0.425	0.458
France	9484	0.329	0.328
Federal Republic of Germany	11,558	0.374	0.337
Greece	3376	0.248	0.924
Honduras	497	0.00552	0.0468
Indonesia	219	0.00241	0.0379
Ireland	3466	0.161	0.295
Jordan	1080	0.0240	0.240
Kenya	376	0.00968	0.0346
Korea	1264	0.0294	0.114
Malaysia	1264	0.0294	0.114
Morocco	729	0.0111	0.0794
Netherlands	10,184	0.419	0.295
New Zealand	4821	0.552	0.874
Panama	1191	0.0858	0.153
Paraguay	712	0.0144	0.0623
Peru	410	0.0239	0.123
Singapore	3468	0.195	0.176
Spain	3569	0.257	0.251
Sweden	11,178	0.716	0.394
Switzerland	15,192	0.654	0.337
Taiwan	6775	0.0994	0.0880
Tanzania	208	0.00454	0.00184
Thailand	486	0.00820	0.122
Turkey	576	0.0320	0.0988
United Kingdom	6276	0.415	0.717
United States	9654	0.743	2.064
Venezuela	3053	0.0645	0.384

SOURCE: Calculated from data in *The Encyclopedia Brittanica Yearbook*, 1980.

TABLE A.33 Residential Consumption of Electricity in the United States

Year	Electrical consumption, kilowatt · hours per year per customer
1912	264
1917	268
1920	339
1925	396
1930	547
1935	677
1940	952
1945	1229
1950	1845
1955	2773
1960	3854
1965	4933
1970	7066
1971	7300
1972	7700
1973	8100
1974	7900
1975	8200
1976	8400
1977	8700
1978	8800
1979	8800
1980	9000
1981	8800
1982	8700
1983	8800
1984	9000
1985	8900
1986	9100
1987	9200
1988	9500
1989	9500

SOURCES: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.34 Installed Horsepower in Motor Vehicles (Automobiles, Trucks, Buses, and Motorcycles) and Total Number of Motor Vehicles in the United States

Year	Total horsepower, $\times 10^3$	Number of vehicles
1900	100	8000
1909	7714	312,000
1919	230,432	7,576,888
1920	280,900	9,239,161
1929	1,424,980	26,704,825
1930	1,470,568	26,749,853
1939	2,400,000	31,009,977
1940	2,511,312	32,453,233
1950	4,403,617	49,161,691
1952	5,361,586	53,265,406
1955	6,632,121	62,688,700
1960	10,366,880	73,868,565
1961	10,972,210	75,958,200
1962	11,930,000	79,173,300
1963	12,713,712	82,713,700
1965	14,306,300	90,357,000
1967	16,152,371	96,930,900
1969	18,075,000	105,096,600
1970	19,325,000	108,407,300
1971	20,732,000	113,000,000
1972	21,736,000	118,800,000
1973	23,029,000	125,700,000
1974	23,224,000	129,900,000
1975	23,752,000	132,949,000
1976	24,339,000	138,500,000
1977	25,025,000	142,400,000
1978	25,892,000	148,778,000
1979	26,617,000	154,100,000
1980	27,362,000	155,900,000
1981	27,909,000	158,500,000
1982	28,852,000	159,600,000
1983	29,662,000	163,900,000
1984	30,117,000	166,200,000
1985	30,972,000	171,700,000
1986	30,893,000	176,200,000
1987	31,488,000	179,000,000
1988	32,415,000	184,400,000

SOURCES: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.35 Highest Electrical Transmission Voltages Used in North America

Year introduced	Voltage, kV
1890	11
1896	33
1904	69
1907	110
1910	140
1912	150
1922	220
1934	287
1953	345
1964	500
1966	735
1969	765

SOURCE: Edison Electric Institute, New York, N.Y.

TABLE A.36 Aircraft Speeds*

Date	Aircraft	Speed, mi/h
August 27, 1939	Heinkel He 17	180
August 27, 1940	Caproni-Campini CC-2	176
April 5, 1941	Heinkel He 280	400
May 15, 1941	Gloster E28/39	362
July 27, 1942	Messerschmitt Me 262A-1a	540
October 1, 1942	Bell XP-59	404
July 1943	Republic X0-47H	490†
December 1943	Arado Ar-234B	461
January 8, 1944	Lockheed XP-80	500
July 1944	Gloster Meteor F-1	410
August 5, 1944	Republic XP-47J	504†
December 1944	Heinkel He 162	465
July 17, 1945	Gloster Meteor F-4	590
August 20, 1969	Modified F7F	478†

*The airplanes were flown under other than official conditions.
†Propeller-driven aircraft.

TABLE A.37 Official Aircraft Speed Records

Date	Country	Speed, mi/h
Nov. 12, 1906	France	25.66
Oct. 26, 1907	France	32.75
May 20, 1909	France	34.06
Aug. 23, 1909	France	43.38
Aug. 24, 1909	France	46.18
Aug. 28, 1909	France	47.84
Apr. 23, 1910	France	48.21
July 10, 1910	France	66.18
Oct. 29, 1910	United States	68.20
Apr. 12, 1911	France	69.47
May 11, 1911	France	74.42
June 12, 1911	France	77.67
June 16, 1911	France	80.91
June 21, 1911	France	82.73
Jan. 13, 1912	France	90.20
Feb. 22, 1912	France	100.22
Feb. 29, 1912	France	100.94
Mar. 1, 1912	France	103.66
Mar. 2, 1912	France	104.33
July 13, 1912	France	106.12
Sept. 9, 1912	United States	108.18
June 17, 1913	France	119.24
Sept. 27, 1913	France	119.24
Sept. 29, 1913	France	126.67
Feb. 7, 1920	France	171.04
Feb. 28, 1920	France	176.14
Oct. 9, 1920	France	181.86
Oct. 10, 1920	France	184.36
Oct. 20, 1920	France	187.98
Nov. 4, 1920	France	192.01
Dec. 12, 1920	France	194.52
Sept. 26, 1921	France	205.22
Sept. 21, 1922	France	211.90
Oct. 13, 1922	United States	222.97
Feb. 15, 1923	France	233.01
Mar. 29, 1923	United States	236.59
Nov. 2, 1923	United States	236.59
Nov. 4, 1923	France	266.58
Dec. 11, 1924	France	278.48
Sept. 5, 1932	United States	294.38
Sept. 4, 1933	United States	304.98
Dec. 25, 1934	France	314.319
Sept. 13, 1935	United States	352.388
Nov. 11, 1937	Germany	379.626
Mar. 30, 1939	Germany	463.917
Apr. 26, 1939	Germany	469.220
Nov. 7, 1945	Great Britain	606.255
Sept. 7, 1946	Great Britain	615.778
June 19, 1947	United States	623.738
Aug. 20, 1947	United States	640.663

TABLE A.37 Official Aircraft Speed Records (*Continued*)

Date	Country	Speed, mi/h
Aug. 25, 1947	United States	650.796
Sept. 15, 1948	United States	670.981
Nov. 19, 1952	United States	698.505
July 16, 1953	United States	715.745
Sept. 7, 1953	Great Britain	727.624
Sept. 25, 1953	Great Britain	735.702
Oct. 3, 1953	United States	752.943
Oct. 29, 1953	United States	755.149
Aug. 20, 1955	United States	822.266
Mar. 10, 1956	Great Britain	1132.136
Dec. 12, 1957	United States	1207.600
May 16, 1958	United States	1404.090
Oct. 31, 1959	U.S.S.R.	1483.850
Dec. 15, 1959	United States	1525.965
Nov. 22, 1961	United States	1606.480
July 7, 1962	U.S.S.R.	1665.900
May 1, 1965	United States	2070.102
July 26, 1976	United States	2193.16

TABLE A.38 Accuracy of Astronomical Angular Measure

Year	Astronomer	Accuracy, arc measure
127 B.C.	Hipparchus	20 minutes
1430 A.D.	Ulugh Beg	10 minutes
1580	Tycho Brahe	4 minutes
1725	Flamsteed	10 seconds
1800	Piazzi	1.5 seconds
1838	Bessel	0.3 seconds
1870	Auwers	0.1 seconds
1900	Newcomb	0.03 seconds
1935	Van Maanen	0.003 seconds

SOURCE: Pledge, H. T., *Science Since 1500,* Philosophical Library, New York, 1947.

TABLE A.39 Accuracy of Measurements of Mass

Year	Accuracy, grain
1550	0.1
1644	0.05
1653	0.03
1825 (steel knife edge; agate bearing)	0.01
1870 (vacuum weighing)	0.001
1930	0.00015

SOURCE: Pledge, H. T., *Science Since 1500,* Philosophical Library, New York, 1947.

TABLE A.40 Efficiency of Incandescent Lights

Year	Efficiency, lm/W
1888	2
1906	4
1907	5
1908	8
1915	10
1916	12
1932	14
1940	16
1959	17

SOURCE: *G. E. Bulletin* TP-110, General Electric Co., Schenectady, N.Y.

TABLE A.41 Efficiency of 40-W Fluorescent Lights

Year	Efficiency, lm/W
1938	1400
1939	1850
1941	2100
1946	2300
1949	2400
1952	2450
1954	2550
1954	2600
1958	2750
1959	2900
1962	3000
1963	3300
1967	3350

SOURCE: *G. E. Bulletin* TP-110, General Electric Co., Schenectady, N.Y.

TABLE A.42 Growth of the Cable Television Industry

Year	Households with TV, $\times 10^3$	Households with cable TV, $\times 10^3$(%)	TV stations	Cable systems
1952	15,300	14 (0.09)	108	70
1953	20,400	30 (0.15)	198	150
1954	26,000	65 (0.25)	402	300
1955	30,700	150 (0.5)	458	400
1956	34,900	300 (0.9)	496	450
1957	38,900	350 (0.9)	519	500
1958	41,424	450 (1.1)	556	525
1959	43,950	550 (1.3)	566	560
1960	45,750	650 (1.4)	579	640
1961	47,200	725 (1.5)	553	700
1962	48,855	850 (1.7)	571	800
1963	50,300	950 (1.9)	581	1000
1964	51,600	1085 (2.1)	582	1200
1965	52,700	1275 (2.4)	588	1325
1966	53,850	1575 (2.9)	613	1570
1967	55,130	2100 (3.8)	626	1770
1968	56,670	2800 (4.4)	642	2000
1969	58,250	3600 (6.1)	673	2260
1970	59,350	4500 (7.7)	686	2490
1971	60,227	5300 (8.8)	688	2639
1972	61,856	6000 (9.7)	690	2841
1973	64,602	7300 (11.3)	692	2991
1974	68,504	8700 (12.7)	694	3158
1975	70,520	9800 (13.8)	693	3506
1976	71,460	10,800 (14.8)	701	3681
1977	73,100	11,900 (16.1)	697	3832
1978	74,700	13,000 (17.1)	708	3997
1979	76,240	14,100 (20.1)	723	4150
1980	80,776	15,500 (19.2)	1011	4225
1981	82,368	18,300 (22.2)	1038	4375
1982	83,527	21,000 (25.1)	1065	4825
1983	82,239	25,000 (30.4)	1106	5600
1984	83,698	30,000 (35.8)	1138	6200
1985	85,000	31,300 (36.8)	1182	6844
1986	86,000	37,500 (43.6)	1235	7600
1987	87,000	41,000 (47.1)	1290	7900
1988	89,000	43,800 (49.2)	1362	8500
1989	NA*	47,800 (NA)	NA	NA

*NA, not available.
SOURCE: *Statistical Abstract of the United States*, Bureau of the Census, Washington, D.C.

TABLE A.43 Magnitude of Engineering Projects

Project	Year of completion	Final cost, current dollars
Santee Canal	1800	7.5×10^5
Lancaster turnpike	1794	4.2×10^5
Middlesex Canal	1803	6.0×10^5
National road	1818	1.4×10^6
Erie Canal	1825	9.0×10^6
London-Holyhead highway	1828	3.5×10^6
Welland Canal	1829	7.7×10^6
Miami and Erie Canal	1832	1.6×10^7
Washington-Baltimore telegraph	1844	3.0×10^4
Marseilles Canal	1847	8.7×10^6
B&O Canal	1850	1.5×10^7
Panama Railroad	1855	8.0×10^6
Union navy ironclad boats	1862	9.1×10^5
Atlantic cable	1866	1.2×10^7
Suez Canal	1869	7.72×10^7
Mont Cenis tunnel	1871	1.5×10^7
Welland Canal	1872	2.2×10^7
Brooklyn Bridge	1883	1.5×10^7
Panama Canal (French)	1890*	2.2×10^8
Corinth Canal	1893	1.16×10^7
Manchester Canal	1894	7.43×10^7
Panama Canal (U.S.)	1914	2.83×10^8
Welland Canal	1932	1.3×10^8
Oakland Bay Bridge	1936	7.7×10^7
German V-2 project	1944	4.0×10^7
Manhattan project	1945	2.0×10^9
USAF ballistic missile project	1958	3.0×10^9
Project Apollo	1970	2.4×10^{10}
Trans-Alaska pipeline	1976	6.0×10^9
Interstate highway system	1982	7.6×10^{10}
North sea oil	1987	4.0×10^{10}

*Year construction halted.

TABLE A.44 U.S. Aircraft Production

Year	Production	Year	Production
1912	45	1946	36,418
1913	43	1947	17,739
1914	49	1948	9839
1915	178	1949	6137
1916	411	1950	6200
1917	2148	1951	7532
1918	14,020	1952	10,640
1919	780	1953	13,112
1920	328	1954	11,478
1921	437	1955	11,484
1922	363	1956	12,408
1923	743	1957	11,943
1924	377	1958	10,938
1925	789	1959	11,076
1926	1186	1960	10,237
1927	1995	1961	9054
1928	3346	1962	9308
1929	6193	1963	10,125
1930	3437	1964	12,492
1931	2800	1965	15,349
1932	1396	1966	19,886
1933	1324	1967	19,141
1934	1615	1968	19,414
1935	1710	1969	16,481
1936	3010	1970	10,943
1937	3773	1971	10,390
1938	3623	1972	12,693
1939	5856	1973	16,081
1940	12,813	1974	16,436
1941	26,289	1975	16,620
1942	47,675	1976	17,605
1943	85,433	1977	19,077
1944	95,272	1978	19,960
1945	48,912		

SOURCE: *Aerospace Facts and Figures,* Aerospace Industries Association, Washington, D.C.

TABLE A.45 Characteristics of Reciprocating Aircraft Engines

Year	Engine	Horsepower	Weight, lb	Displacement, in^3
1902	Wright Brothers	12	179	410
1902	Manley	52	151	549
1912	Mercedes	80	312	443
1913	Gnome	100	270	783
1914	Rolls-Royce Eagle	350	880	1241
1914	Curtiss OX5	90	390	568
1916	Hispano-Suiza A	150	467	718
1917	Liberty	420	857	1650
1921	Bristol Jupiter	485	775	1253
1922	Curtiss D-12	435	680	1210
1926	Pratt & Whitney Wasp	600	685	1344
1929	Wright Cyclone 9	1525	1469	1823
1934	Rolls-Royce Merlin	1730	1450	1649
1936	Bristol Hercules	1980	2115	2360
1937	Pratt & Whitney Double Wasp	2050	2390	2800
1944	BMW	1700	1940	2550
1945	Pratt & Whitney Wasp Major	3250	3670	4363
1948	Wright Turbo-compound	3400	3675	3347

SOURCE: *Encyclopedia Britannica,* 1964.

TABLE A.46 Tractive Effort, Steam and Diesel, of U.S. Railroads

	Tractive effort, $\times 10^6$ lb	
Year	Diesel	Steam
1935	3	2178
1936	4.1	2135
1937	7	2127
1938	9.9	2092
1939	19.1	2041
1940	32.1	2006
1941	50	1994
1942	78	2010
1943	100.6	2045
1944	147	2057
1945	210	2028
1946	248	2018
1947	326	1914
1948	455	1816
1949	618	1632
1950	808	1463
1951	1018	1272
1952	1207	964
1953	1337	723
1954	1405	546
1955	1489	389
1956	1578	255
1957	1644	176
1958	1671	99
1959	1715	55
1960	1728	20
1961	1726	8.7

TABLE A.47 Inanimate Nonautomotive Horsepower*

Year	Factories†	Mines	Railroads	Merchant ships, powered	Sailing vessels	Farms	Total
1850	1150	60	586	325	400	—	2521
1860	1675	170	2156	515	597	—	5113
1870	2453	380	4462	632	314	—	8241
1880	3664	715	8592	741	314	668	14,694
1890	6308	1445	16,980	1124	280	1452	27,589
1900	10,309	2919	24,501	1663	251	4009	43,652
1910	16,697	4473	51,308	3098	220	10,460	86,256
1920	19,422	5146	81,082	6508	169	21,443	133,770
1930	19,519	5620	109,743	9115	100	28,610	172,707
1940	21,768	7332	92,361	9408	27	57,472	188,368
1950	32,921	8500	110,969	23,423	11	57,533	233,357
1955‡	35,579	30,768	60,304	24,155	5	207,742	358,553
1960	42,000	34,700	46,856	23,890	2	237,020	384,468
1965	48,400	40,300	43,838	24,015	2	269,822	426,377
1970	54,000	45,000	54,000	22,000	1	288,500	463,501
1971	56,000	45,000	56,000	21,000	—	300,000	478,000
1972	57,000	46,000	57,000	22,000	—	305,000	487,000
1973	58,000	46,000	57,000	21,000	—	308,000	490,000
1974	59,000	46,000	61,000	21,000	—	315,000	502,000
1975	60,000	47,000	62,000	22,000	—	318,000	509,000
1976	61,000	47,000	64,000	22,000	—	324,000	518,000
1977	62,000	47,000	62,000	23,000	—	328,000	522,000
1978	65,000	48,000	64,000	25,000	—	335,000	537,000

*In units of 1000 hp.
†Includes electric motors.
‡Beginning with 1955, the data is not strictly comparable with earlier years.

TABLE A.48 Mill Consumption of Natural versus Synthetic Fibers

Year	Natural*	Synthetic*
1950	5340	1491
1951	5571	1471
1952	4957	1464
1953	4966	1502
1954	4526	1484
1955	4815	1851
1956	4825	1685
1957	4444	1745
1958	4207	1702
1959	4781	2065
1960	4657	1817
1961	4506	2064
1962	4629	2418
1963	4465	2788
1964	4615	3174
1965	4923	3512
1966	5016	3989
1967	4747	4244
1968	4489	5306
1969	4256	5500
1970	4012	5500
1971	4185	6345
1972	4091	7566
1973	3819	8665
1974	3412	7700
1975	3141	7416
1976	3543	8053
1977	3296	8889
1978	3162	9236
1979	3195	9465

*In millions of pounds.

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.49 Installed Hydroelectric-Generating Capacity

Year	Capacity, $\times 10^3$ kW
1902	1140
1912	2794
1920	4804
1925	5150
1930	9650
1935	10,399
1940	17,304
1945	15,892
1950	18,674
1960	33,180
1965	44,490
1970	55,751
1971	56,000
1972	56,000
1973	62,000
1974	64,000
1975	66,000
1976	68,000
1977	69,000
1978	71,000
1979	75,000
1980	76,000
1981	77,000
1982	79,000
1983*	81,000

*Change in definition after 1983 makes subsequent data not strictly comparable, hence it is not included.

SOURCE: *Historical Statistics of the United States,* U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.50 Installed Steam-Generating Capacity

Year	Capacity, $\times 10^3$ kW	
	Conventional	Nuclear
1902	1847	—
1912	8186	—
1920	14,635	—
1925	32,429	—
1940	37,138	—
1945	45,248	—
1950	61,495	—
1955	101,698	—
1960	149,161	—
1965	205,423	—
1970	260,000	6
1971	295,000*	—
1972	309,000*	—
1973	340,000*	—
1974	338,000	32
1975	353,000	40
1976	368,000	43
1977	388,000	53
1978	400,000	54
1979	412,000	55
1980	425,000	56
1981	440,000	61
1982	452,000	63
1983†	456,000	67

*Separate figures for nuclear and conventional steam plants are not available for these years.

†Change in definition after 1983 makes subsequent data not strictly comparable, hence it is not included.

SOURCE: *Historical Statistics of the United States*, U.S. Bureau of the Census, Washington, D.C.; and *Statistical Abstract of the United States*, U.S. Bureau of the Census, Washington, D.C.

TABLE A.51 Maximum Steam Turbine Capacity

Year	Capacity, $\times 10^3$ kW
1904	5
1905	9
1906	12.5
1912	20
1915	45
1924	50
1925	60
1928	94
1929	208
1953	217
1955	217.26
1956	260
1957	275
1958	350
1960	500

TABLE A.52 Maximum Hydroturbine Capacity

Year	Capacity, $\times 10^3$ kW
1916	7.5
1919	13.5
1920	17.5
1921	30
1922	31.25
1926	40
1928	40.6
1929	45
1930	66.7
1932	77.5
1936	82.5
1945	108
1957	167
1963	204
1966	220

TABLE A.53 Milestones in Space Flight

Year	Nation	Event
1957	U.S.S.R.	Dog in orbit
1961	U.S.S.R.	Dog returned alive
1961	U.S.S.R.	First man in orbit
1962	U.S.S.R.	First dual launch
1964	U.S.S.R.	Three-man crew
1965	U.S.	330 h in orbit
1965	U.S.	First orbital maneuver
1965	U.S.S.R.	First space walk
1966	U.S.	Rendezvous with second craft
1967	U.S.	First docking
1968	U.S.	Manned circumlunar
1969	U.S.S.R.	17 days in orbit
1969	U.S.	Manned lunar landing
1970	U.S.S.R.	First crew transfer
1973	U.S.	Skylab
1977	U.S.S.R.	Salyut
1982	U.S.S.R.	210 days in orbit
1982	U.S.	Shuttle
1986	U.S.S.R.	Mir
1987	U.S.S.R.	326 continuous days in orbit
1990	U.S.S.R.	Japanese journalist aboard Mir

TABLE A.54 Characteristics of a Select Number of Computers, 1963 to 1965

Computer	Operations per second	Seconds per dollar lease cost
	1963	
IBM 7040	21,420	44.54
IBM 7044	67,660	23.98
Philco 1000	6811	65.63
GE 215	5246	89.07
IBM 1440	1412	183.40
	1964	
CDC 3200	195,256	51.96
GE 205	1775	311.8
RCA 3301	126,761	44.54
NCR 315-100	6164	155.9
Burroughs B5500	376,275	20.78
	1965	
SDS 92	19,140	239.8
CDC 3100	118,462	77.94
DDP-116	2176	677.7
GE 625	224,374	15.20
PDP-7	68,497	103.9
NCR 590	4288	519.6
ASI 6240	33,177	155.9
Raytheon 520	28,118	207.8
IBM 360/75	3,560,854	11.81

SOURCE: Data collected by Kenneth J. Knight, University of Texas at Austin.

TABLE A.55 Chronology of Major Inventions and Innovations, 1850 to 1970

Product or Process	Invention	Innovation
DDT	1874	1942
Electric precipitation	1884	1909
Synthetic detergent	1886	1930
Ballpoint pen	1888	1946
Zipper	1891	1923
Diesel locomotive	1895	1913
Safety razor	1895	1904
Magnetic tape recorder	1898	1937
Gyrocompass	1900	1909
Radio	1900	1918
Tungsten carbide	1900	1926
Hardening of liquid fats	1900	1909
Fluorescent lamp	1901	1938
Float glass	1902	1957
Automatic drive in vehicles	1904	1939
Helicopter	1904	1936
Bakelite	1904	1910
Stainless steel	1904	1912
Neoprene	1906	1932
Silicones	1910	1946
Antiknock gasoline	1912	1935
Plexiglass	1912	1935
Watertight cellophane	1912	1926
Catalytic cracking of petroleum	1915	1935
Continuous hotstrip	1920	1923
Cotton picker (Campbell)	1920	1942
Insulin	1920	1927
Kodachrome	1921	1935
Self-winding wristwatch	1922	1928
Television	1923	1936
Cotton picker (Rust)	1924	1941
Hydraulic clutch	1924	1946
Power steering	1925	1930
Radar	1925	1934

TABLE A.55 Chronology of Major Inventions and Innovations, 1850 to 1970 *(Continued)*

Product or Process	Invention	Innovation
Crease-resisting fabric	1926	1932
Continuous steelcasting	1922	1952
Nylon, perlon	1927	1938
Jet engine	1928	1941
Penicillin	1928	1943
Sulzer loom	1928	1945
Synthetic light polarizer	1928	1932
Rocket	1929	1944
Cyclotron	1929	1937
Freon refrigerants	1930	1931
Polyethylene	1933	1937
Phototypesetting	1936	1954
Cinerama	1937	1953
Titanium	1937	1944
Xerography	1937	1950
Electronic digital computer	1939	1943
"Terylene" polyester fiber	1941	1955
Shell molding	1941	1948
Streptomycin	1943	1944
Chordane, Aldrin, and Dieldrin	1944	1947
Long-playing record	1945	1948
Transistor	1947	1951
Oxygen steel making	1949	1952
Hovercraft	1955	1968
Semisynthetic penicillin	1957	1959
Wankel rotary piston engine	1957	1967
Moulton bicycle	1959	1963
Prevention of rhesus hemolytic disease	1961	1967

SOURCE: Sahal, D., "Invention, Innovation, and Economic Evolution," *Technological Forecasting & Social Change*, **23**:213–235 (1983).

TABLE A.56 Sensitivity of Photographic Materials, 1839 to 1966

Year	Process*	Time, s†
1839	Daguerreotype	2400
1834	Talbot's Calotype	120
1856	Wet collodion	30
1880	Gelatin	5
1888	Eastman's American film	1/12
1920	Improved techniques	1/25
1930	New sensitizing dyes	1/64
1938	New technology	1/500
1956	New technology	1/1600
1966	New technology	1/4800

*Qualifying notes:
1. No specific products are given after 1856; later data may not apply to all contemporary products on the market.
2. No specific lens, emulsion, or processing combinations are given; hence, times may not be strictly comparable.
3. The assumption of full sunlight exposure disregards other practical radiation sources for which films may have been optimized (e.g., tungsten lamps or flash bulbs).
4. Image quality standards are not given; hence, times may not be strictly comparable.

†Exposure time in full sunlight at f/16 lens aperture.

SOURCE: Personal correspondence, Mr. Robert Anwyl, Eastman Kodak Company, Rochester, N.Y.

TABLE A.57 Characteristics of U.S. Fighter Aircraft

Aircraft	First flight	Turn rate, °/sec		Sea level climb rate, ft/min	Takeoff roll, ft	Maintenance hours per flight hour	Mean flying hours between failures	Range, mi	Payload, lb	Maximum mach number	Cruise speed, knots	Radar range, mi	Number of simultaneous targets	Number of missiles		Range of missiles, mi		Gun†
		Instantaneous	Sustained											Dogfight	BVR*	Dogfight	BVR	
F80	1944	15	10	6000	2500	20	1	435	500	0.65	379	0	0	0	0	0	0	1
F84	1946	20	10	900	2500	20	1	294	366	0.8	465	0	0	0	0	0	0	1
F86	1947	22	11	6000	1800	20	1	362	480	0.9	496	0	0	0	0	0	0	1
F89	1949	20	10	8000	3800	30	0.5	450	500	0.75	404	5	1	0	6	0	5	1
F94	1950	20	10	9000	3000	35	0.5	450	500	0.7	452	5	0	0	0	0	0	1
F100	1953	18	10	9000	2900	35	0.5	460	500	1.4	513	5	0	2	0	2	0	1
F101	1954	18	10	39,000	2700	35	0.5	500	1000	1.6	473	10	1	0	2	0	6	0
F102	1953	16	10	18,000	2300	35	0.5	956	1000	1.2	521	10	1	0	6	0	6	0
F104	1954	16	10	40,000	2800	35	0.5	600	1000	2	500	10	1	2	2	2	6	0
F106	1955	16	10	22,000	2300	29	0.7	400	1000	1.8	500	10	0	4	0	2	0	1
F8	1956	16	10	42,000	3200	27	0.6	633	1000	2	564	15	1	0	6	0	6	0
F5A	1959	16	8.5	28,000	2500	12	2.5	255	706	1.4	500	10	0	2	0	2	0	1
F4E	1967	16	10	30,000	2000	35	1.1	451	1000	2.2	512	20	1	4	4	2	10	1
F14	1971	16	12	53,000	1200	35	0.7	1738	2000	2	500	50	4	4	4	4	30	1
F5E	1971	18	10	32,000	2000	10	3	120	5000	1.6	500	20	0	2	0	4	0	1
F15	1972	19	14	62,000	850	33	2	2500	2000	2.4	500	50	1	4	4	4	20	1
F16	1974	19	14	58,000	1100	19	2.4	296	3000	1.8	500	30	1	2	0	4	0	1
F18	1978	21	13	50,000	1000	12	3.2	500	3000	1.8	500	40	1	2	2	4	20	1
F20	1982	22	14	58,000	1300	10	4	300	3000	2	500	30	1	2	2	4	20	1

*BVR, beyond visual range.
†Guns: 1 = yes, 0 = no.
NOTE: The values given have been obtained from open literature sources. In some cases, these may differ from the actual values.

TABLE A.58 Japanese Camera Patent Applications, Percent Market Share, and Total Imports

Year	Number of patent applications	Market share, %	Japanese imports, $\times 10^3$	Total imports, $\times 10^3$
1964	—	68.96	366.2	531
1965	22	71.78	366.1	510
1966	35	85.98	282.9	329
1967	50	87.10	263.9	303
1968	69	76.32	217.5	285
1969	105	79.56	272.1	342
1970	144	77.10	252.9	328
1971	113	87.10	441.6	507
1972	121	81.54	534.1	655
1973	202	82.13	565.9	689
1974	193	80.19	546.9	682
1975	221	85.57	431.3	504
1976	167	68.67	517.8	754
1977	85	74.45	996.2	1338

TABLE A.59 German Typewriter Patent Applications, Percent Market Share, and Total Imports

Year	Number of patent applications	Market share, %	German imports, $\times 10^3$	Total imports, $\times 10^3$
1964	—	4.58	27.9	609.3
1965	17	3.67	34.7	946.7
1966	17	3.48	42.9	1234.4
1967	28	3.10	39.7	1280.7
1968	27	4.46	60.8	1363
1969	23	6.03	83.8	1390.6
1970	22	9.25	130.4	1410.4
1971	22	6.81	108.2	1588.9
1972	28	6.07	112.8	1858.8
1973	23	7.83	132.7	1694.4
1974	17	4.87	133.7	2745.1
1975	26	10.26	147.7	1439.6
1976	15	6.97	134.4	1928.3
1977	9	5.11	140.4	2746.7

TABLE A.60 Japanese Watch Patent Applications, Percent Market Share, and Total Imports

Year	Number of patent applications	Market share, %	Japanese imports, $\times 10^3$	Total imports, $\times 10^3$
1964	—	0.15	0.9	609.3
1965	4	0.20	1.9	946.7
1966	6	0.15	1.8	1234.4
1967	18	0.34	4.4	1280.7
1968	15	0.35	4.8	1363
1969	15	0.63	8.8	1390.6
1970	42	1.89	26.6	1410.4
1971	81	4.85	75.6	1558.9
1972	106	17.24	320.5	1858.8
1973	88	6.66	112.9	1694.4
1974	131	15.90	436.4	2745.1
1975	109	8.79	126.5	1439.6
1976	173	11.25	216.9	1928.3
1977	33	9.51	261.2	2746.7

TABLE A.61 Seasonal Energy Efficiency Rating (SEER) of Reciprocating Compressors for Air Conditioners

Year	SEER
1981	7.8
1982	8.2
1983	8.4
1984	8.6
1985	8.7
1986	8.8
1987	9.0
1988	9.15
1989	9.25
1990	9.4

SOURCE: Data provided by Dr. Dean Ruwe, Copeland Corporation, Sidney, Ohio.

TABLE A.62 Use of Built-In-Test (BIT) in Line Replaceable Units (LRUs) in U.S. Air Force Aircraft

Year	Aircraft	Percent of LRUs
1963	C-141	10
1967	F-111	18
1968	C-5A	24
1975	F-15	25
1977	EF-111	25
1979	E-3A	27
1980	E-4B	30
1985	B-1B	38
1986	C-5B	35

SOURCE: AFTEC Logistics Assessment Center, Wright-Patterson AFB, OH.

TABLE A.63 Recording Density of IBM Mainframe Computer Hard Disk Drives

Year	IBM disk drive	Recording density, bits per square inch
1957	350	2.1×10^3
1963	1311	5.1×10^4
1965	2311	1.5×10^5
1966	2314	2.2×10^5
1971	3330	7.8×10^5
1973	3340	1.7×10^6
1976	3350	3.1×10^6
1979	3370	7.7×10^6
1981	3380	1.2×10^7

SOURCE: Freeman, R. C., "The Future of Peripheral Data Storage," *Mini-Micro Systems,* 175–179 (February 1982).

TABLE A.64 Recording Density of IBM Magnetic Tape Storage Device

Year	IBM tape unit	Recording density, bits per square inch
1953	726	1.4×10^3
1955	727	2.8×10^3
1959	729 III	7.3×10^3
1962	729 V	1.1×10^4
1965	2401-1	1.4×10^4
1966	2401-4	2.9×10^4
1973	3420-4	1.1×10^5

SOURCE: Freeman, R. C., "The Future of Peripheral Data Storage," *Mini-Micro Systems,* 175–179 (February 1982).

TABLE A.65 Market Share of Electronic Telephone Switching Systems

Year	Penetration, × 10^2 %
1965	0.00009
1966	0.00015
1967	0.00157
1968	0.00707
1969	0.01205
1970	0.02234
1971	0.03578
1972	0.05746
1973	0.08495
1974	0.11700
1975	0.14995
1976	0.18482
1977	0.23308
1978	0.28858
1979	0.35238
1980	0.42244
1981	0.47951

SOURCE: Lee, J. C., and K. W. Lu, "Algorithm and Practice of Forecasting Technological Substitutions with Data-Based Transformed Models," *Technological Forecasting & Social Change*, **36**:401–414 (1989).

TABLE A.66 Performance of Gas Turbines

Year	Engine*	Brake hp	Weight, lb	Volume, ft^3	Specific fuel consumption, lb fuel/hp · h
1955	Standard 1s/60	60	133	6.05	1.45
1957	Standard 1s/250	250	350	19.8	1.20
1969	Solar T62%	105	70	3.32	1.1
1975	Solar T39	40	94	7.2	0.95
1976	Williams Int'l	153	85	4.5	0.9
1986	Tierney	75	127	1.88	1.07

*Small engines, 50 to 250 hp.
SOURCES: A. W. Judge, *Small Gas Turbines and Free Piston Engines* (1960); and *Gas Turbine International*, **10(1)** (Jan–Feb 1969). Manufacturers' catalogs and literature.

TABLE A.67 Performance of Rotary Engines

Year	Engine*	Brake hp	Weight, lb	Volume, ft^3	Specific fuel consumption, lb fuel/hp · h
1959	KKM 250	40	48.4	0.5	0.675
1962	Curtiss Wright RC1-60	103	192	5.26	0.6
1965	Curtiss Wright RC2-60	206	261	5.92	0.57
1967	KKM 612	115	283	2.96	0.5
1968	Toyo Kogyo Model 0813	128	224	8.5	0.51
1969	Toyo Kogyo Model 0820	110	268	8.4	0.5
1986	John Deere 1007R	100	175	4.66	0.42

*Small engines, 50-250 hp.

SOURCES: "Transport Engines of Exceptionally High Specific Output," *Proceedings International Society of Mechanical Engineers,* **183(38)** (1968–1969); K. Yamamoto, "Rotary Engines," Toyo Kogyo Co. Ltd., 1969; and John Deere Co. product literature.

TABLE A.68 Substitution of Pressurized-Cabin Passenger Aircraft for Unpressurized-Cabin Aircraft

	Millions of seat · miles flown		
Year	Unpressurized	Pressurized	Total
1946	3.5	0.1	—
1947	3.5	1.9	5.4
1948	3.2	3.2	6.4
1949	3.2	4.0	7.2
1950	3.1	4.6	7.7
1951	2.7	5.8	8.5
1952	2.4	8.0	10.4
1953	2.2	9.9	12.1
1954	1.7	11.6	13.4
1955	1.8	12.8	14.6
1956	1.9	14.6	16.5
1957	0.9	18.8	19.8
1958	0.8	20.3	21.2
1959	1.4	26.2	28.1
1960	0.4	33.3	33.7

TABLE A.69 Substitution of Jet Aircraft for Propeller Aircraft, U.S. Airlines

	Millions of seat · miles flown		
Year	Propeller	Jet	Total
1959	1817	993	2810
1960	1721	1649	3370
1961	1473	2387	3860
1962	1290	2840	4130
1963	1160	3000	4160
1964	994	3796	4790
1965	775	4615	5390
1966	960	6960	7920
1967	630	8980	9610
1968	730	11,050	11,780
1969	110	12,230	12,340
1970	90	15,220	15,310

TABLE A.70 Substitution of First Generation Turbofan Engines for Turbojets in American Passenger Airlines

Year	Turbojet engines	Turbofan engines	Total
1959	1393	160	1533
1960	2777	367	3144
1961	3151	927	4078
1962	3257	1379	4636
1963	3304	1805	5109
1964	3237	2706	5943
1965	3271	3810	7081
1966	3310	5697	9007
1967	3335	8180	11,515

NOTE: After 1967, second generation turbofan engines continued the substitution.

TABLE A.71 Substitution of Second Generation Turbofan Engines for First Generation Turbofan Engines in American Passenger Airlines

Year	First generation	Second generation	Total
1969	6821	201	7022
1970	7557	712	8269
1971	7904	1192	9096
1972	8056	1731	9787
1973	8273	2654	10,927
1974	8554	3854	12,408
1975	8679	4978	13,657

TABLE A.72 Substitution of Treated Railroad Cross Ties for Untreated Ties

Year	Untreated ties*	Treated ties*	Total*
1920	49,037	37,792	86,829
1921	50,450	36,072	86,522
1922	46,012	40,630	86,642
1923	42,779	41,656	84,435
1924	38,582	44,490	83,072
1925	32,627	50,090	82,717
1926	25,188	55,558	80,746
1927	23,280	62,963	86,243
1928	20,254	64,331	84,585
1929	17,240	64,724	81,964
1930	14,796	54,529	69,325
1931	12,598	41,851	54,449
1932	10,030	30,107	40,137
1933	11,389	26,618	38,007
1934	11,764	32,367	44,131
1935	11,321	33,939	45,260
1936	10,911	38,206	49,117
1937	10,064	29,674	49,738
1938	7919	34,589	42,508
1940	6628	38,698	45,326
1941	6205	43,872	50,077
1942	5309	47,932	53,241
1943	4522	44,822	49,344
1944	3564	47,695	51,259
1945	2967	43,657	46,624
1946	2479	37,671	40,150
1947	2286	37,920	40,206
1948	2191	38,281	40,472
1949	1728	31,198	32,926
1950	1538	31,553	33,091
1951	1653	30,384	32,457
1952	1321	32,910	34,231
1953	1318	32,144	33,462
1954	1197	24,531	25,728
1955	683	26,490	27,173
1956	475	26,848	27,323
1957	626	24,497	25,123
1958	296	17,426	17,722
1959	190	18,077	18,267

*Number of ties in thousands.
SOURCE: *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.73 Substitution of Mechanical Loaders for Hand Loading in U.S. Coal Mining

Year	Hand loading*	Mechanical loading*	Total*
1925	496,939	6243	503,182
1926	545,899	10,545	556,444
1927	482,885	16,500	499,385
1928	459,397	21,559	480,956
1929	467,859	37,862	514,721
1930	400,702	46,982	447,684
1931	315,595	47,562	363,157
1932	254,252	35,817	290,069
1933	280,317	37,821	318,138
1934	297,145	41,433	338,578
1935	301,549	47,177	348,726
1936	343,985	66,977	410,962
1937	330,280	83,500	413,780
1938	233,045	85,093	318,138
1939	246,421	110,712	357,133
1940	269,734	147,870	417,604
1941	292,411	186,667	459,078
1942	282,587	232,903	515,490
1943	260,687	249,805	510,492
1944	244,489	274,189	518,678
1945	205,118	262,512	467,630
1946	175,617	245,341	420,958
1947	193,072	298,157	491,229
1948	164,206	295,806	460,012
1949	109,447	222,376	331,823
1950	120,119	272,725	392,844
1951	113,791	302,051	415,842
1952	87,431	268,994	356,425
1953	71,222	278,329	349,551
1954	46,142	242,970	289,112
1955	52,794	290,671	343,465
1956	58,372	307,402	365,774
1957	54,912	305,737	360,649
1958	43,308	243,573	286,881
1959	39,703	243,731	283,434
1960	39,182	245,706	284,888
1961	37,416	235,350	272,766
1962	40,296	240,970	281,266
1963	43,015	259,241	302,256
1964	40,707	281,101	321,808
1965	36,029	296,632	332,661
1966	28,243	310,281	338,524
1967	19,219	329,914	349,133
1968	14,755	329,387	344,142
1969	11,700	335,431	347,131
1970	9599	329,189	338,788
1971	4992	270,896	275,888
1972	2974	301,129	304,103

*Thousands of net tons produced.
SOURCE: *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition,* U.S. Bureau of the Census, Washington, D.C.

TABLE A.74 Substitution of Basic Oxygen Process for Open Hearth Process in U.S. Steel Industry

Year	Open hearth process*	Basic oxygen process*	Total†
1958	77	1	79
1959	83	2	85
1960	88	3	91
1961	85	4	89
1962	84	6	89
1963	90	9	98
1964	99	15	114
1965	95	23	118
1966	85	34	119
1967	71	41	112
1968	66	49	115
1969	61	60	121
1970	48	63	111
1972	35	75	110
1974	36	82	117
1975	22	72	94
1976	24	80	103
1977	20	77	97
1978	21	84	105
1979	19	83	103
1980	13	68	81
1981	14	73	87
1982	6	45	51
1983	6	52	58

*Millions of tons of steel produced.
†The two processes may not add to total because of rounding.
SOURCES: American Iron and Steel Institute, *Annual Statistics Report*, Washington, D.C.; U.S. Department of the Interior, Bureau of Mines, *Mineral Yearbook*, Washington, D.C.

TABLE A.75 Substitution of Catalytic Cracking for Thermal Cracking in the U.S. Petroleum Refining Industry

	Capacity, thousands of barrels per day		
Year	Thermal cracking	Catalytic cracking	Total
1937	2193	02	2195
1938	2228	120	2348
1939	2114	24	2138
1940	2168	116	2284
1941	2197	155	2352
1942	2284	172	2456
1943	2289	252	2541
1944	1214	330	1544
1945	2199	876	3075
1946	2239	960	3199
1947	2361	1122	3483
1948	2495	1286	3781
1949	2347	1402	3749
1950	2408	1722	4130
1951	2340	1823	4163
1952	2369	2075	4444
1953	1592	2512	4104
1954	1362	2828	4190
1955	1124	3318	4442
1956	957	3668	4625
1957	860	3894	4754

TABLE A.76 Substitution of Pelletized Iron for Iron Ore*

Year	Pellets†	Ore†	Total†
1949	10	36,753	36,763
1950	39	41,338	41,377
1951	86	48,865	48,951
1952	66	24,876	24,942
1953	350	49,679	50,029
1954	556	30,801	31,357
1955	721	44,108	44,829
1956	3032	47,575	50,607
1957	4220	39,560	43,780
1958	5654	22,203	27,857
1959	5513	18,436	23,949
1960	7574	28,725	36,299
1961	9968	20,539	30,507
1962	10,164	20,122	30,286
1963	13,297	9808	23,105
1964	14,836	19,979	34,815
1965	16,649	19,857	36,506
1966	17,488	20,533	38,021
1967	20,185	15,367	35,552
1968	24,913	12,787	37,700
1969	28,780	13,394	42,174
1970	28,803	12,155	40,958
1971	27,530	9576	37,106
1972	30,388	8736	39,124
1973	35,365	10,627	45,992

*Substitution is for Lake Superior district only. Substitution figures are based on iron content (61.5 percent iron in pellets and 51.5 percent iron in ore).
SOURCE: U.S. Department of the Interior, Bureau of Mines, *Mineral Yearbook*, Washington, D.C.
†Thousands of tonnes (long tons).

TABLE A.77 Market Shares in U.S. Beverage Can Market

Year	Three-piece steel, %	Two-piece steel, %	Two-piece aluminum, %
1963	94.5	—	5.5
1964	90.7	—	9.3
1965	90.1	—	9.9
1966	88.2	—	11.8
1967	84.8	—	15.2
1968	84.5	—	15.5
1969	81.9	—	18.1
1970	80.0	—	20.0
1971	72.8	—	27.2
1972	77.4	—	22.6
1973	72.3	—	27.7
1974	65.4	—	34.6
1975	54.8	6.9	38.3
1976	43.4	11.7	44.9
1977	35.3	14.2	50.4
1978	26.6	18.1	55.3

SOURCES: For 1963 to 1974, *Industry Reports: Metal Cans, U.S. Department of Commerce, Washington, D.C. and the U.S. Government Printing Office (for various years). For 1975 to 1978, Can Manufacturers Institute, Washington, D.C.*

TABLE A.78 Substitution of Natural Gas in U.S. Ammonia Production

Year	Capacity, thousands of short tons	
	Natural gas	Total
1931	110	640
1932	110	640
1933	110	640
1934	110	640
1935	110	640
1936	110	640
1937	110	640
1938	110	640
1939	110	640
1940	160	640
1941	160	690
1942	160	854
1943	490	1254
1944	682	1443
1945	737	1501
1946	780	1740
1947	780	1740
1948	802	1762
1949	856	1816
1950	899	1859
1951	1031	2020
1952	1496	2080
1953	1932	2576
1954	2754	3398
1955	3090	3734
1956	3505	3949
1957	3818	4975
1958	3892	5139
1959	NA	NA*
1960	4364	5521
1961	5035	6248
1962	5701	6914
1963	6678	7891
1964	7454	8667
1965	9382	10,593
1966	11,551	12,762
1967	14,543	15,354
1968	16,676	17,217

*NA, not available.
SOURCE: W. H. C. Simmonds, National Research Council of Canada, private communication.

TABLE A.79 Adoption of Computers in Municipal Police Departments*

Year	Departments, %
1960	9
1963	18
1967	40
1971	44
1974	56

*Cities over 25,000 population.
SOURCE: *Municipal Year Book,* International City Managers Association, 1968 through 1975.

TABLE A.80 Sensitivity of Infrared Detectors in the 3 to 5 μm Band

Year	D*	Material
1956	2.5×10^{9}	InSb
1959	1.3×10^{10}	InSb
1961	5.1×10^{10}	InSb
1963	1.25×10^{10}	InSb
1964	6.0×10^{10}	InSb
1965	1.4×10^{10}	InSb
1966	5.2×10^{10}	InSb
1967	4.0×10^{10}	InSb
1969	7.0×10^{10}	InSb
1971	1.0×10^{11}	InSb
1976	1.0×10^{11}	InSb
1978	1.3×10^{11}	InSb
1978	1.5×10^{11}	PbS
1979	7.5×10^{10}	SiSe
1983	2.0×10^{11}	InSb
1984	1.0×10^{11}	InSb

*In units of $cm(Hz)^{(1/2)}W^{-1}$.
NOTE: The maximum theoretical limit is 2.0×10^{12}.

TABLE A.81 Sensitivity of Infrared Detectors in the 8 to 12 μm Band

Year	D*	Material
1958	1.5×10^{7}	HgCdTe
1959	3.5×10^{9}	GeHg
1963	4.0×10^{9}	GeHg
1964	7.8×10^{9}	GeHg
1965	7.0×10^{9}	GeHg
1966	4.0×10^{10}	GeHg
1967	4.0×10^{10}	HgCdTe
1969	5.0×10^{10}	HgCdTe
1971	5.0×10^{10}	HgCdTe
1973	4.0×10^{10}	HgCdTe
1975	4.0×10^{10}	HgCdTe
1976	4.0×10^{10}	HgCdTe
1976	4.0×10^{10}	PbSnTe
1983	4.0×10^{10}	HgCdTe
1984	6.0×10^{10}	HgCdTe

*In units of $cm(Hz)^{(1/2)}W^{-1}$.

NOTE: The maximum theoretical limit is 4.0×10^{11}.

TABLE A.82 Power and Energy Density of Hydrogen-Oxygen Fuel Cells

Year	Reciprocal power density, lb/kW	Time, h	Energy density, kW · h/lb
1962	80	555	6.938
1965	70	1000	14.286
1968	245	1500	6.122
1968	38	1944	51.158
1970	23	2550	110.870
1972	19	3472	182.74
1973	16	6500	406.25
1976	15	8055	537.00
1980	10	10,000	1000.00

TABLE A.83 Substitution of Continuous Mining Machines in U.S. Bituminous Coal Industry*

Year	Production by continuous miners†	Total production†
1952	8215	357,931
1953	11,830	349,551
1954	16,435	289,112
1955	27,460	343,465
1956	39,907	365,714
1957	53,793	360,649
1958	56,373	386,884
1959	65,792	283,434
1960	77,928	284,888
1961	84,321	272,766
1962	90,174	281,266
1963	104,350	302,256
1964	124,677	321,809
1965	141,938	332,001
1966	155,053	338,524
1967	165,571	349,133
1968	163,816	344,142
1969	172,642	347,132
1970	169,897	338,788
1971	152,943	275,888
1972	178,375	304,105
1973	178,600	299,353

*Underground mining only.
†Thousands of net tons.
SOURCE: *Minerals Yearbook*, U.S. Department of the Interior, Bureau of Mines, Washington, D.C., various editions.

TABLE A.84 Substitution of Mechanical Loaders in U.S. Bituminous Coal Industry*

Year	Production using mechanical loaders†	Total production†
1940	147,870	417,604
1941	186,667	549,078
1942	232,903	515,490
1943	249,805	510,492
1944	274,189	518,678
1945	262,512	467,630
1946	245,341	420,958
1947	298,157	491,229
1948	295,806	460,012
1949	272,276	331,823
1950	272,725	392,844
1951	302,051	415,842
1952	268,994	356,425
1953	278,329	349,551
1954	242,970	289,112
1955	290,671	343,465
1956	307,402	365,774
1957	305,737	360,649
1958	243,573	286,881
1959	243,731	283,434
1960	245,706	284,888
1961	235,350	272,766
1962	240,970	281,266
1963	259,241	302,256
1964	281,101	321,808
1965	296,632	332,661
1966	310,281	338,524
1967	329,914	349,133
1968	329,387	344,142
1969	335,431	347,131
1970	329,189	338,788
1971	270,896	275,888
1972	301,129	304,103
1973	297,384	299,354

*Underground mining only.
†Thousands of net tons.
SOURCE: *Minerals Yearbook*, U.S. Department of the Interior, Bureau of Mines, Washington, D.C.

TABLE A.85 Use of Float Glass in the Flat Glass Industry

Year	Float glass, %*
1967	16
1968	23
1969	29
1970	37
1972	50
1975	78

*Expressed as percent of flat glass.
SOURCE: *U.S. Industrial Outlook,* U.S Department of Commerce, Washington, D.C., 1977.

TABLE A.86 Substitution of Natural Soda Ash for Synthetic Soda Ash in Alkali and Chlorine Production

Year	Natural soda ash, %*
1971	37
1972	44
1973	47
1974	53
1975	53
1976	75
1977	81

*Expressed as percent of all soda ash production.
SOURCE: *U.S. Industrial Outlook,* U.S. Department of Commerce, Washington, D.C., 1977.

TABLE A.87 Substitution of Kraft Pulping Process in Paper Production

Year	Kraft pulp production*	Total pulp production*
1920	189	3822
1925	410	3962
1930	950	4360
1935	1468	4926
1940	3748	8900
1950	7506	14,849
1955	11,577	20,740
1960	15,034	25,316
1965	21,146	42,216
1970	29,408	42,216

*Production in thousands of tons.
SOURCE: Guthrie, J. A., *An Economic Analysis of the Pulp and Paper Industry,* Washington State University Press, Pullman, WA, p. 42, 1972.

Appendix

B

Precursor Events in Several Technologies

Table	Subject
B.1	Events in the History of Automotive Fuel Injection
B.2	Events in the History of Electronic Engine Control
B.3	Events in the History of Electronic Ignition
B.4	Events in the History of Electronic Distributors
B.5	Events in the Recent History of Automotive Carburetors
B.6	Events in the History of Plastic Automobile Body Shells
B.7	Events in the Recent History of Steel Automobile Body Shells
B.8	Events in the History of High-Strength Low-Alloy (HSLA) Steel
B.9	Events in the History of Plastic Automobile Structural Parts
B.10	Events in the History of Automobile Engine Turbocharging
B.11	Events in the History of Diesel Passenger Car Engines
B.12	Development of Several Aluminum Alloys for Aircraft Use
B.13	Events in the Recent History of Aluminum-Lithium Alloys

The events in all the tables except Table B.12 were extracted from entries in the *Engineering Index*. All the tables were collected as part of the research described in Martino, J. P., "Using Precursors as Leading Indicators of Technological Change," *Technological Forecasting & Social Change,* **32**:341–360 (1987), which was supported by the National Science Foundation.

TABLE B.1 Events in the History of Automotive Fuel Injection

Year	Events
1968	A fluidic control for fuel injection is demonstrated and showed that fluidic control is feasible but further development needed.
1971	Borg Warner demonstrates that electronic fuel injection, using electronic control based on several engine parameters, reduces exhaust emissions.
	Lucas Automotive of Great Britain develops electronically controlled fuel injection system that will make it possible to meet U.S. 1974 emissions regulations.
1972	The Bosch electronic fuel-injection unit is adapted to Renault R17 TS with a high-performance engine to meet U.S. 1974 emissions requirements.
	Electronic fuel injection is proposed as a means for reducing pollution, making starting easier, and eliminating the choke, throttle, and dieseling.
	A digitally controlled fuel injection system in which engine speed and throttle angle are used as indexes to a table of fuel flow rates stored in a digital Metal Oxide Semiconductor (MOS) Read-Only Memory (ROM) (open loop system) is demonstrated.
	Bendix describes electronic fuel injection using control law on engine response surface.
1973	A control law for fuel injection based on engine speed, throttle angle, and an engine response surface is developed.
	Bendix announces the development of an oxygen sensor for closed-loop control of fuel injection to minimize emissions.
	The Bosch K-Jetronic fuel injection system uses measurements of air flow to control fuel injection rate.
	Tests on Bosch closed-loop exhaust emission control system with electronic fuel injection show that low emissions and fuel economy can be obtained. Durability of the system has not yet been demonstrated.
	Bendix describes design of electronic fuel injection control. Discusses problems of operating in automotive environment and issues of mass production and cost.
1974	Pneumatic control combined with electronic switching provides low-cost fuel injection, reducing emissions while increasing engine power.
	Bendix presents tutorial on design of electronic controls for fuel injection.
	Norwegian experiments demonstrate electronically controlled fuel injection using general purpose digital computer to replace the pulse-forming network in a Bosch Jetronic fuel-injection system.
	Bosch announces electronic fuel injection in a closed-loop system that monitors exhaust gas composition.
	Fiat announces experiments with electronically controlled fuel injection in gasoline (Otto-cycle) engines, using minicomputer as part of experimental apparatus.
1975	Bendix announces electronic control of fuel injection using feedback control from an exhaust oxygen sensor to attain near-stoichiometric fuel-air mixture.
	Cadillac Seville incorporates Oldsmobile V-8 engine with modifications for fuel injection and electronic control.
	Bendix produces Electronic Control Unit for electronic fuel injection, recently introduced on a U.S. passenger car.
	Cadillac introduces electronic control of fuel injection utilizing Bendix ECU.

TABLE B.1 Events in the History of Automotive Fuel Injection (*Continued*)

Year	Events
1976	Peugeot, Renault, and Volvo introduce new V6 engine with fuel injection to meet power and emissions requirements.
	Gulton Industries announces automated testing of pressure transducers for fuel injection systems, lowering production costs.
	Bendix announces pressure-monitoring system for use in testing electronic fuel injection controls, resulting in increases in production yields. The new test equipment draws heavily on Bendix's experience with aircraft flightline check-out equipment.
1977	Volvo introduces feedback system that measures exhaust gas composition and controls Bosch fuel injection to achieve acceptable emissions without EGR or air injection, which will appear on 1977 model Volvo 240 for sale in California.
	A corona-discharge air-mass flow transducer for control of electronic fuel injection is demonstrated.
	Experimental demonstration shows that a gasoline engine with fuel injection produces higher power, has better fuel economy, and less emissions than an engine with a carburetor.
1978	Bendix announces prototype microprocessor electronic fuel-injection controller that can achieve equal emissions, better driveabilty, and the same fuel economy as current electronic analog fuel-injector controls.
	An *interindustry emission control* (IIEC-2) car, utilizing electronic control of fuel injection, spark timing, and EGR is tested; it is installed in the 1977 Ford Granada.
	GM announces plans to phase out carburetors and replace them with electronically controlled throttle-body fuel injection. The reason is to improve driveability and reduce fuel consumption and emissions.
	Porsche introduces turbocharged six-cylinder engine with Bosch K-Jetronic fuel-injection system. The engine is based on racing engine designed in 1964.
	Tests reported on alternative means for reducing emissions, including Bendix electronically controlled fuel injection.
	A microcomputer-based ignition timing, EGR control, and fuel-injection control is demonstrated on Ford's IIEC-2 Concept Car.
1979	Bendix proposes electronically controlled multipoint fuel injection for eight-cylinder engines as means for achieving 0.4 gal/mi of nitrogen oxides. This will require more frequent measurements and computations.
	Bendix announces multipoint electronic fuel injection as means of meeting pollution and fuel economy standards. Prototype is available but is not yet fleet tested.
	Ford and Cadillac use electronic fuel injection to meet pollution requirements. Ford central fuel injection will be incorporated in 1980 models. Cadillac Eldorado and Seville will incorporate GM system.
	Chrysler carburetor uses feedback from oxygen sensor in exhaust manifold to control air-fuel mixture.
	A standard microprocessor is used in place of custom Large Scale Integration (LSI) for electronic control of fuel injection in prototype.
	An ultrasonic fuel injection is demonstrated; it has low cost and wide tolerance, and it is durable.

TABLE B.1 Events in the History of Automotive Fuel Injection (*Continued*)

Year	Events
1980	An air-mass flow measurement system that functions in automotive environment to improve performance of fuel injection is demonstrated.
	Bosch demonstrates hot-wire anemometer for measuring air-mass flow for better control of fuel injectors.
	Toyota develops a prototype single-point fuel-injection system with electronic control. It controls 2.0-L six-cylinder engine.
	Nissan demonstrates electronic engine controller that includes fuel-injection control.
	GM develops *throttle-body injection* (TBI) and introduced into 6.0-L Cadillac Eldorado and Seville; the 10 psi low-pressure injector (claimed to be a significant achievement) has electronic control.
	Chrysler introduces single-point, electronically controlled fuel injection as part of electronic engine control.
1982	Mitsubishi Motors develops electronically controlled fuel injection using airflow sensor based on von Karman vortex street for demonstration only, not for production prototype. Fuel response is greatly improved.
1983	Automatic honing of fuel-injection nozzles, using cubic boron nitride as abrasive, produces better finish and more accurate injection at economic cost.
	Nissan introduces electronic engine control that controls fuel injection and other variables to minimize emissions. A microprocessor-controlled fuel-injection system was developed that injects fuel at intake valve of each cylinder only while that valve is open.
1984	Electronic and mechanical fuel injection appears on Mercedes-Benz 190 E 2.3-16.
	Several U.S. cars have electronically controlled fuel injection.
1985	Lower cost fuel injectors to be introduced in 1987 models.
1986	Mazda RX-7 rotary engine has fuel injection.

TABLE B.2 Events in the History of Electronic Engine Control

Year	Events
1972	A demonstration model (cigar-box size) of digital-memory electronic controller for fuel flow is shown. It is tested on the Triumph 2.5 PI and demonstrates emissions quality just short of the 1975 federal standards.
1973	Transducer designs are presented for parameters needed by electronic engine controllers, including throttle angle, crankshaft speed and position, air-mass flow, and exhaust quality.
1974	GM presents theoretical analysis of closed-loop control of automotive engines to maintain proper fuel-air ratio.
	The British produce a prototype of an engine speed limiter that cuts off spark when the engine speed exceeds a critical value and signals the driver to shift to higher gear. This allows the engine to be run at full throttle without damage.
	Bendix proposes closed-loop electronic control of automobile engines to improve fuel economy. Sensors, actuators, and multiplexing circuits must be developed to achieve this goal.
1975	Chrysler introduces analog electronically controlled "lean-burn" engine with air to fuel ratio of 18:1 and meets the federal emission specifications for nitrogen oxides (NO_x) without catalytic converter or air pump. A closed-loop feedback control system for automobiles, which optimizes fuel economy and power output, was designed.
1976	Ford announces an air-mass flow meter based on the vortex shedding principle; it allows better control of fuel to air ratio for emissions control.
	Chrysler introduces an analog electronic control for its lean-burn spark management system. Analog Integrated Circuit (ICs) and discrete components are utilized. The system measures throttle position, rate of change of throttle, temperature, crankshaft position, engine speed, and engine load to compute the proper spark advance.
1977	Bosch proposes single central computer for control of automotive functions.
	A University of Louisville researcher designs a microprocessor-based control for fuel-air mixture to achieve fuel economy.
	Fairchild announces two new chips for analog-to-digital conversion of measured automotive parameters; they are intended for electronic engine control.
	United Technologies demonstrates microprocessor engine-control system utilizing engine speed and air-mass flow to control emissions; the results are superior to the use of air density as the input variable.
	Ford demonstrates an experimental electronic engine control utilizing ultrasonic air-mass flowmeter and a control law that optimizes engine performance on the basis of measured engine parameters.
	Bunker Ramo develops a production prototype of a rugged manifold vacuum sensor for electronic engine control; it is suitable for the automotive environment.
1978	The interindustry emission control (IIEC-2) car, utilizing electronic control of fuel injection, spark timing, and Exhaust Gas Recirculation (EGR), is tested and is installed in the 1977 Ford Granada.
	Bendix announces the availability in production quantities of sensors for measuring manifold absolute pressure, crankshaft position, throttle position, engine temperature, and exhaust gas recirculation for use in electronic engine control.

TABLE B.2 Events in the History of Electronic Engine Control (*Continued*)

Year	Events
	Ford reports that the biggest part of the cost of electronic engine control is not the microprocessor but the sensors and actuators. Reductions in the cost of these are needed.
	Ford demonstrates experimental knock-detection sensor suitable for use in electronic engine control.
	Ford combines the Electronic Engine Control (EEC-I) spark advance and EGR control, originally introduced on Lincoln Versailles, with a closed-loop fuel-air mixture control obtained from the 2.3-L engine used on Bobcat and Pinto models sold in California. The result is the EEC-II, which will be used in all states on 5.8-L V-8 and is optional on 1979 Mercury Marquis and California Ford LTD.
	Chrysler will replace the analog control with a digital control of the spark on its 1979 models.
1979	Bosch announces new electronic system combining electronic fuel injection and electronic ignition using microcomputer control.
	The Varajet II for Pontiac 2.5-L L-4 engine has electronic control of the Rochester carburetor to adjust the fuel mixture to reduce pollution.
	Custom digital IC for cruise control is used on the Audi.
	GM introduces (micro)computer-controlled catalytic converter (C-4) system; it controls carburetor fuel flow on the basis of exhaust gas measurements.
	Nissan introduces microcontroller on its 2.8-L engine, which controls fuel injection, spark timing, exhaust gas recirculation, and the auxiliary air intake when idling.
	Ford EEC-I controls ignition timing, fuel injection, EGR flow, and thermactor air to meet the 1980s emission requirements. Total electronic control reduces maintenance of mechanical parts.
1980	The U.S. Department of Transportation proposes electronic control of engines as a means of improving fuel efficiency and reducing emissions.
	Bosch announces a prototype electronic engine control combining fuel injection and electronic ignition.
	Toyota reports the development of electronic control for timing and fuel injection for small cars.
	American Microsystems develops a custom microprocessor specifically for automobile engine controls. The chip set can be manufactured at reasonable costs, is reliable, and works in the automobile environment.
	GM introduces an electronic emissions control that has integral self-test and diagnostics.
	GM introduces electronically controlled throttle body injection on 6.0-L engines for Cadillac Eldorado and Seville.
	American Motors Corporation describes a single-chip microcomputer system to control and protect the three-way-catalyst-equipped six-cylinder vehicles.
	Fiat develops a controller using microcomputer that controls the engine and transmission, and has greater fuel economy, lower emissions, and better driving ability.
	Feasibility study of multivariate linear-quadratic control theory for feedback control of automobile engines shows that it can produce superior driveability and can avoid torque sag at lean operating conditions.

TABLE B.2 Events in the History of Electronic Engine Control (*Continued*)

Year	Events
	Ford's EEC-III electronic engine control controls the ignition timing, exhaust gas recirculation, fuel, evaporative emission control canister purge, and secondary air.
	Chrysler has a new single-point, fully electronically controlled system to control fuel injection. The electronic ignition and electronic spark advance portions are derivatives of devices, with demonstrated reliability, in use for several years.
	Cadillac introduces an electronic engine control that integrates fuel injection, electronic spark timing, EGR switching, and cold fast and curb idle-speed regulation.
	Nissan demonstrates an electronic engine controller that includes fuel injection control and controls fuel-air mixture, spark ignition timing, EGR, and idle speed. It is marketed as ECCs.
1981	GM reports that custom ICs can reduce cost and increase the reliability of electronic engine controls.
	Motorola develops an engine controller based on its widely used MC6805 microcomputer that uses phase-locked loop for engine timing of low-pressure fuel injection based on the crankshaft angle.
1982	Nippon Electric Company introduces the MU PD7811 single chip for engine control, replacing several chips.
	Nissan Motors and AMP introduce a fiber optics system to replace the conventional electronic control wiring in automobiles. It provides noise immunity and reduced harness size.
1984	The Volvo prototype measures wheel speeds, detects wheel spin, and reduces engine torque by controlling the fuel injectors and turbocharger (if there is one).
	Motorola reports over $50 million has been spent on developing electronic systems to improve fuel economy and reduce emissions.
	Toyota introduces the lean combustion system, which controls air to fuel ratio to improve fuel economy.
1985	Bendix prototypes a lower-cost fuel injector to be utilized in 1987 model cars.
	GM obtains a British patent for a fuel injection rail designed to reduce turbulence.
	The Austin Rover 1.6-L engine with electronically controlled fuel injection is demonstrated.
	Motorola optimizes the 6805 microprocessor for automotive applications. The MC6805S2 has dual timers, suitable for ignition control. MC6805K2 has on-chip Electronically Erasable Programmable Read-Only Memory (EEPROM) for storing data such as mileage or cumulative engine hours while the engine is shut off. Adaptive control by the microprocessor is demonstrated.
	It is proposed that sensors of crankshaft position, fuel flow, oxygen, knock, pressure, and temperature for on-engine measurements as input to a central controller be developed.
	Chrysler, working in conjunction with suppliers, achieves 120,000-mi reliability for electronic engine controls and other automotive electronic parts.
	Ford announces EEC-IV, a two-chip system that includes 16-bit microprocessor and memory with both *read-only memory* (ROM) and *random access memory* (RAM), developed jointly with Intel.

TABLE B.2 Events in the History of Electronic Engine Control (*Continued*)

Year	Events
1986	A GM subsidiary, Rochester Corp., that is a carburetor manufacturer begins production of U.S.-designed electronically controlled fuel-injection system; it will compete with German Bosch, which is currently the leading supplier.
	A University of Pisa professor demonstrates a closed-loop microprocessor system for control of fuel flow. Control is based on a combination of specific fuel consumption and work done per cycle.
	Hitachi develops a single-point injection system integrated into the throttle body; it controls fuel flow in response to measured air mass flow (hot wire anemometer).
	Mazda RX-7 rotary engine has 8-bit microcomputer to control spark timing, fuel injection flow rate, and emissions control.
1987	All 1987 Ford models will incorporate a custom IC in the electronic engine control module that cuts parts count and reduces circuit board size of the power supply to ¼ of its former size.

TABLE B.3 Events in the History of Electronic Ignition

Year	Events
1970	*Wireless World* publishes "do-it-yourself" construction article on 12-V transistorized ignition; the parts lists is provided, and the operation are described.
1974	A study by Joseph Lucas Ltd. shows that performance of ignition systems, including transistorized ones, is limited by inadequate performance of mechanical elements of the system, including breaker points. The mechanical sophistication must be decreased and electronic sophistication increased.
	Ford introduces transistorized ignition on all 400 Cubic Inch Displacement (CID) and 460 CID engines, as well as all 200 through 351 CID engines to be sold in California.
	Ball Brothers introduces capacitative-discharge ignition system using unijunction transistor to the filter point bounce and buffer point opening.
	A capacitative-discharge electronic ignition, utilizing standard electronic components including an IC timer, filament transformer, and rectifier is demonstrated.
	RCA introduces high-voltage, reliable, low-cost silicon power transistor for transistorized ignition.
1975	Continental produces industrial (nonautomotive) IC spark engine, utilizing electronic ignition for reduced emissions.
	A high-energy ignition system is developed for automotive engines. It controls the dwell time and primary current in the ignition coil, needs no adjustment for the lifetime of the unit, and is suited for both breaker and transistorized ignition.
	A multispark system that electronically generates 200 sparks during each combustion cycle, improving starting in arctic climates, is demonstrated.
1976	Chrysler introduces an analog electronic control for its lean-burn spark management system. Analog ICs and discrete components are utilized. The system measures throttle position, rate of change of throttle, temperature, crankshaft position, engine speed, and engine load to compute the proper spark advance.
1977	GM introduces the MISAR ignition control system on the Oldsmobile Toronado; it is the first microprocessor in a production automobile.
1978	A high-voltage thyristor is developed in Switzerland to provide stability at high-voltage and high-current capacity, along with high-power gain for electronic ignition.
	Chrysler develops the LSI digital ignition control suited for mass production. A fusible-link *programmable read-only memory* (PROM) device allows programming during assembly for four, six, or eight cylinders.
	The Buick turbocharged V-8 includes a closed-loop electronic spark-timing control.
	The Oldsmobile Toronado is in production with electronic spark control system. Magnetically sensed crankshaft position, engine r/min, manifold vacuum, and coolant temperature are utilized by a digital microprocessor to compute spark advance.

TABLE B.3 Events in the History of Electronic Ignition (*Continued*)

Year	Events
1979	Fully electronic ignition is available for two- and four-cylinder engines.
	Bosch reports that electronic ignition advance can realize the full potential for fuel economy only if the fuel to air ratio is simultaneously optimized.
	A pressure sensor is developed to provide information for spark advance and knock detection for electronic ignition.
1980	Fuji develops a microprocessor-based controller that measures six engine parameters and computes optimum spark advance and dwell time. It uses a double winding coil and high-voltage diode instead of a mechanical distributor. In tests, it improves fuel consumption by 3.4 percent.
	The new Audi Quattro incorporates electronic ignition and turbocharger.
1981	Bendix introduces an electronic ignition that maintains high spark energy even at low speed (starting) or when input voltage is low (weak battery).
	Mitsubishi develops a hybrid ignition system; it uses a magnetic pickup to generate voltage-waveform signals that replace the mechanical breaker and that are used by the electronic ignition.
1983	Ford uses the Hall effect sensor for ignition timing on EEC-IV, which gives a square-wave signal more appropriate for digital control than the previously used variable reluctance sensors.
1984	A single-chip ignition controller is developed.

TABLE B.4 Events in the History of Electronic Distributors

Year	Events
1974	Lumenitron Ltd. reports that since 1967 it has successfully used optoelectronic replacements for breaker points. Over 10,000 systems are in use, some with over 150,000 mi of service.
	Motorola Semiconductor presents a design for an electronic breakerless inductive system that reduces emissions and increases spark plug life.
	Ball Brothers introduces a capacitative-discharge ignition system that uses a unijunction transistor to the filter point bounce and to buffer point opening.
1975	RCA introduces IC for breakerless ignition.
	A breakerless ignition system utilizing a toothed metallic trigger wheel is demonstrated. Its operation is purely electronic, and the output voltage is independent of the engine speed or battery condition, improving starting in cold weather or with a low battery.
1978	Ford demonstrates distributorless ignition, which reduces Radio Frequency Interference (RFI) and provides greater range of spark advance and smaller size.
1979	Motorola demonstrates a fully electronic ignition with no mechanical distributor.
	Echlin Manufacturing develops a prototype electronic trigger for spark ignition systems utilizing the Weigand effect that fits into the conventional distributor bowl.
1980	Fuji develops a microprocessor-based controller that measures six engine parameters and computes optimum spark advance and dwell time. The controller uses double-winding coil and high-voltage diode instead of a mechanical distributor. In tests, it improved fuel consumption by 3.4 percent.
1986	Hitachi develops a completely electronic distributor with no mechanical parts.

TABLE B.5 Events in the Recent History of Automotive Carburetors

Year	Events
1976	The variable Venturi carburetors are introduced on production automobiles; the opening is electronically controlled and is a function of engine speed and load.
1979	The Pierburg P 1B is a further development of previous mechanical carburetors.
	Bosch and Pierburg design electronic controls for conventional carburetors.
	The Pontiac Varajet-II carburetor uses a microprocessor to control fuel flow on the basis of exhaust gas measurements.
	GM introduces microcomputer control of a carburetor in conjunction with a *computer-controlled catalytic converter* (C-4) system. Closed-loop control of the fuel-air mixture maintains the effectiveness of the catalytic converter. It will be introduced in 1980 models in California and in 1981 nationwide.
	Ford's Variable Venturi carburetor is mechanically controlled.
	Ford develops a carburetor with closed-loop control of the fuel-air mixture based on an oxygen sensor for a three-way catalyst.
1980	A carburetor is developed for the Honda CVCC engine; it produces a lean mixture for the regular cylinder and a rich mixture for the auxiliary combustion chamber. A mechanical linkage between throttle shafts is needed.

TABLE B.6 Events in the History of Plastic Automobile Body Shells

Year	Events
1978	Ford builds a demonstration graphite fiber reinforced plastic car; test results are reported in 1983.
	ICI America introduces resins and urethane thickeners for the production of structural-grade *sheet molding compound* (SMC). Actual field use is still needed to demonstrate its usefulness for automobile applications.
1980	Ford Motor Company reports the design, development, and manufacture of a lightweight, one-piece SMC hood for the Econoline Van.
	Laminates of steel sheet with plastic core have the same formability as steel and are lower cost and lighter weight.
1981	About 100 kg of polymers is now used in a medium-sized car.
	There are reports of several experiments with plastic for body panels and interior components.
1982	Fiat builds a demonstration car with a steel skeleton and plastic panels for the body shell. Its skeleton has only 800 welds as compared with 3000 that is typical for an all-steel car and is more resistant to corrosion, has better formability for low drag shape, and is 20 percent lighter than an all-steel car.
1983	The 1984 model Fiero bolt-on body allows tolerances of ± 0.005 in its finish and assembly of bodies; the precision is equal to or better than that of steel bodies.
1984	The Pontiac Fiero, a steel "space frame" with all-plastic body skin, is produced. The flexible plastic "Bexloy" is used in the rear and proves that plastic can match the surface finish of steel. Special painting arrangements are separate lines for space frame, plastic panels, and flexible front and rear fascias. Painting the panels before mounting them reduces paint repair work.
	The use of plastics in automobiles is three times greater than that of a decade earlier and now makes up 10 percent of a car's total weight.
1985	Finite-element methods are used to design steel chassis that carries all torsional loads, putting no stress on the plastic body.
	Successful in-plant coloring of ABS body plastic reduces inventory costs by eliminating the need to store colored plastic.
1986	Auto plants have 2000-t plastic molding machines. In the Ford Sierra, the plastic grille panel is body-color painted in the same plant as the painted metal panels, demonstrating that plastic can be used in conjunction with steel.

TABLE B.7 Events in the Recent History of Steel Automobile Body Shells

Year	Events
1980	Vitreous enamels are being tested for body protection as well as for exhaust protection.
1981	Ciba-Geigy has developed an adhesive to replace spot welding, including repairs equal in strength to spot welding.
	Finite-element design methods can reduce the weight of the steel automobile deck while meeting strength specifications. There is a 20 percent weight reduction with no cost increase, and a 23 percent weight reduction at a slight cost increase.
	There is a trend toward the use of *high-strength* (HS) steel not only for bumpers and structural members but also for body panels. In addition, surface-coated sheets are being developed and used for corrosion prevention.
1982	All Ford and Lincoln 1983 models will incorporate galvanized steel.
	All Chrysler 1983 hoods will be one-sided galvanized steel.
1983	The development of one-side galvanized steel is helping steel retain its position as the leading source of body shell material.
	Kawasaki Steel Corp. has new masking coat process for producing one-sided galvanized steel sheet.
	A pilot production line is set up in the U.S. for one-sided hot-dip galvanized steel.
	Fiat reports using hot-dipped or electrogalvanized sheets for structural parts; painted coated steel for body panels is common in Europe.
1984	Dodge Mini Ram Van uses 1½-sided galvanized steel for all exterior panels except the roof; the product should last 10 years.
1986	New plant at Dearborn will produce 700,000 t/year of anticorrosion sheet steel, which is supposed to last 10 years in service.

TABLE B.8 Events in the History of High-Strength Low-Alloy (HSLA) Steel

Year	Events
1980	Higher strength-to-weight is reportedly obtainable from unalloyed and "diet" steels.
	Vanadium HSLA steel has superior formability properties and is available in hot and cold rolled, both uncoated and zinc coated.
	Dual-phase steels have both martensite and ferrite and are formed more easily than previous high-strength steel. They pick up strength when stamped, which allows for reduced thickness and weight with no loss of strength.
1981	New HSLA steels are developed for automotive applications, including deep-drawable grades.
	There is a trend toward the use of high-strength steel not only for bumpers and structural members but also for body panels. In addition, surface-coated sheets are being developed and are used for corrosion prevention.
	Nippon Steel develops a new steel sheet with good formability and dent resistance properties. Several heat treatments give the desired high-bake hardenabilty and high anti-aging properties.
	Ford Motor Company reports that high-strength steels are offering significant competition to other materials for weight reduction.
	Nippon Steel develops high-strength sheets with good formability.
1982	HSLA steels are being used for lighter-weight bumpers.
1984	Steel sheets for automobiles must have contradictory properties: high strength, formability, as well as weldability. Nippon Steel has a new process for making such steels.

TABLE B.9 Events in the History of Plastic Automobile Structural Parts

Year	Events
1980	Fiber-reinforced plastic wheels do not yet meet current impact and burst requirements; the use of *thermoplastics* (TP) may help.
	A duplicate of a Ford 1979 U.S. passenger car is designed and fabricated using graphite-fiber-reinforced plastic as the structural material. It is laid up by hand. The overall weight savings is 565 kg or ⅓ of the car's weight.
	Design methods are developed for continuous fiber composite materials. Specific techniques are developed for laminate design and stress analysis.
	A demonstration door panel of composite material with continuous glass intrusion strap saves 10 lb and eliminates the intrusion beam. The consolidation of parts allows the cost to be competitive with steel door.
1981	Composite plastic springs are used in producing the Corvette.
1982	A fiber-reinforced plastic rear axle has been demonstrated. The weight saving is 57 percent, and the increase in cost is about $1.20/kg saved; its durability is short of expectations.
	A monotapered leaf spring has been demonstrated. It is made of a hybrid laminate, i.e., carbon-reinforced and glass-fiber-reinforced plastics.
	The demonstration model was 76 percent lighter than steel. Test results are presented.
1983	Nisshin Steel starts a pilot production of steel-plastic-steel sandwich, which is formable like steel sheet.
	Formula 1 Grand Prix racing cars by McLaren International of Surrey, England, have graphite composite skin over aluminum honeycomb for their monocoque chassis. Metal almost totally eliminated from chassis. The composite is also used for the front cowl, cockpit, and rear stabilizer.
	Reinforced plastic sheet is being stamped into rocker and timing chain covers, battery trays and oil pans, spare wheel compartments, front panels, and bumper inserts.
	A carbon-fiber-reinforced plastic rear axle has been designed as a replacement for the steel axle in a production car, resulting in a 27 percent weight reduction. It passes laboratory tests that simulate actual loads.
	The C 111 experimental car has a load-bearing floor and front end with facings of fiberglass-reinforced epoxy and a core of polyurethane foam.
1984	The Audi 100-200-Avant is produced, and its spare wheel recess is load-bearing fiber-reinforced plastic.
	Ford's plant at Milan, MI, has an automated, cost-efficient plastic manufacture and assembly facility. It will make blow-molded load floors, seat backs, and other seat components.
	GM's plastics plants, Chevrolet/Adrian, stamps plastic fender-liners. They are modeled on a sheet steel stamping line.

TABLE B.10 Events in the History of Automobile Engine Turbocharging

Year	Events
1975	Renault introduces the Formula 1 engine with an exhaust gas turbocharger.
1978	Porsche introduces a turbocharged version of the 911 engine, which has been in production since 1964 as a racing engine. The new engine includes an air-to-air intercooler that cools boost air, allowing a greater compression ratio of 7:1. A nonturbocharged version was introduced in 1975.
	Buick introduces turbocharged V-8 as an option for its current model cars.
1979	After considerable development effort, BMW has a new six-cylinder engine with exhaust-driven turbocharger, which raises the output to 221 kW.
1980	Brown Boveri has designed turbochargers for passenger car diesel engines, 30 to 100 kW range, that are based on experience obtained with commercial vehicles. The result is better performance and fuel economy, and lower emissions.
	The Audi Quattro is introduced and has electronic ignition and turbocharger.
	Saab-Scania develops a knock detector to be used in electronic controller for Saab 2.0-L turbocharged engine.
1981	The Volvo 2.1-L four-cylinder engine is developed in a more powerful version by means of turbocharging. It is designed for high torque at low and medium speeds, not for high power.
	Bendix reports that tests of Bendix supercharger on a 1.4-L gasoline engine confirms improvements in fuel economy, performance, and driveability, with acceptable emissions.
	Poll of automakers shows that all have active turbocharger or supercharger projects, and expect to introduce them on 1982 or 1983 models; in most cases, they will add them to existing engines rather than design new engines.
1982	A turbocharged version of Mitsubishi Colt 1400 is offered.
	Mitsubishi Motors introduces the Redia sedan and Cordia coupe with turbocharged engine.
1983	A newly designed control removes the compressor from the air intake when the boost is not needed; this eliminates the problem of rotational inertia of the turbocharger rotor.
	The Audi 200 touring sedan has a third-generation turbocharged engine for better fuel economy and response characteristics.
	Ford introduces an overhead-cam six-cylinder turbocharged diesel to go in its Mark VII luxury cars, Thunderbird, and Cougar, and in Continental if enough engines are available. It has a high compression ratio, and its compact design allows it to fit under the hood.
	GM will expand the use of diesels to include all-new six-passenger front-wheel-drive replacements for C-body cars.
	Several electronically fuel-injected cars will have turbochargers added in the 1984 model year.
	Ford and Lincoln-Mercury will have an intercooler on their turbochargers.
1984	Water-cooled bearing housing on turbocharger reduces carbonization of oil and will become standard on the Audi 200.
	There are reports of several successful applications of turbochargers in automobile engines.

TABLE B.11 Events in the History of Diesel Passenger Car Engines

Year	Events
1979	A new injector that is suitable for small indirect injection of passenger car engines is demonstrated.
1980	Bosch Corp., a developer of controls, reports that electronic controls for diesel engines have reached a high development state.
	There are reports that diesel engines are being used for passenger cars. Electronic controls, derived from gasoline engines, can improve the performance of diesels. Mechanically controlled fuel pumps will be replaced by electronically controlled fuel pumps.
	TRW proposes electronically controlled fuel-injection pumps for passenger diesels.
1981	An electronic fuel-injection system for passenger diesels is being designed.
1982	The Diesel Rabbit is introduced and is the most fuel-efficient car in America, i.e., 50 mi/gal.
	Bosch announces fail-safe electronics for controlling diesel engines; these shut the engine off in event of failures that would otherwise damage the engine.
1983	Ford Europe introduces diesel engines for its passenger cars; they are 1.6 L and 54 bhp.
	Germany intends to export to the United States an automobile with an optimized turbocharger, combustion chamber, and fuel injector diesel engine.
	Ford's 1984 Tempo, Topaz, Thunderbird, Cougar, Escort, Lynx, and Mark VII will have optional diesel engines.
	GM's new-for-1984 six-passenger front-wheel-drive replacements for C-body cars will have optional diesel engines.
	Cummins Diesel introduces electronically controlled fuel pump that fits within the envelope of mechanical pumps. An external plug-in *programmable read-only memory* (PROM) chip allows standardization to be achieved within sets of identical pumps by closed-loop control rather than by the previously required calibration of individual pumps.
1984	Closed-loop electronic control of EGR is introduced for passenger diesel engines in California.
1985	Toyota develops a microprocessor-based controller for automotive diesel engines that produces high-power output and good fuel economy.
	The first accurate needle lift sensor for generating a signal for the timing and duration of fuel injection for diesel engines under electronic control is developed.
	Electronic controls for barometric altitude compensation of injection pump timing for passenger diesels in California is introduced.

TABLE B.12 Development of Several Aluminum Alloys for Aircraft Use

Product name	Events
Alclad	It was commercialized in 1928 by Alcoa.
	The alloy core is covered with pure aluminum.
	Cladding provides galvanic protection even where underlying alloy is exposed, e.g., at rivets.
	Northrop Alpha, first U.S. all-metal plane, was made of Alclad (1930). Boeing Monomail, 1930, was also made of Alclad.
17ST (later redesignated 2017-T4)	Its trade name is duralumin.
	It was developed by Wilm in 1908, with the discovery of age hardening.
	Germany started producing it in 1911.
	It was commercialized by Alcoa in the United States in 1916.
	It was used in Zeppelins in World War I in 1916.
	It was used in Junkers F-13; its first flight was in 1919, and it was the first all-metal aircraft.
2020	It was developed in T6 temper to give strength with lower density. (It included lithium.)
	It was first announced in 1957 and was used in the Navy A-5; its first flight was in 1958.
	It was withdrawn because of lack of fracture toughness.
2024	Its trade name is super duralumin.
	It was developed in 1931 in T3 temper.
	The Douglas DC-3 transport was constructed almost exclusively of Alclad 2024-T3 (1935).
2224	It was registered on 05/04/78 and was developed by Boeing as an extrusion alloy for the 757 (1982) and 767 (1981).

TABLE B.12 Development of Several Aluminum Alloys for Aircraft Use (*Continued*)

Product name	Events
2324	It was registered on 05/04/78.
	The T39 temper was registered by Boeing on 12/21/78 and was developed by Boeing as a plate for the 757 (1982) and 767 (1981).
2419	It was registered on 12/12/72 and was used in the B-1B bomber (1983).
7075 (originally designated 75S)	It is a first-generation alloy.
	It was introduced in 1943 in T-3 and T-6 tempers and was used for upper wing skin and stringers of the B-50; the first flight was in 1947.
	T73 temper was patented in 1965 and was intended to improve resistance to stress corrosion.
	T73 was used in the DC-10 (1970).
7079	It is a first-generation alloy.
	It was first commercialized in the United States in 1954 and was derived from an earlier German alloy.
	It was the original material in the Navy A6, first flight 1960, and was later replaced by 7050 when corrosion problems developed.
7150	It was registered on 05/04/78 and was used as extrusions and plate for the 757 (1982) and 767 (1981).
7178	It was introduced in 1951 in T6 temper and was used on the Boeing 707, first flight in 1954.
7475	It was registered on 09/15/69.
	It was first used on the MRCA Tornado, first flight in 1974, and was later used on the F-16.

SOURCE: Data in this table was obtained primarily from *Metals and Alloys* and *Jane's All The World's Aircraft*.

TABLE B.13 Events in the Recent History of Aluminum-Lithium Alloys

Year	Events
1978	MIT develops a new "splat cooling" process for aluminum-lithium alloys, allowing the previously unmixable elements to be combined.
	Aluminum-lithium alloys are proposed by the Aluminum Company of America (Alcoa) for aircraft use because of their light weight.
1982	At the first international conference on aluminum-lithium alloys, 21 papers are presented, giving evidence of growing interest in these alloys.
1983	Alcoa develops three new aluminum-lithium alloys for aircraft use.
	Alcoa will expand its aluminum-lithium production to cast 20,000-lb ingots.
1984	The trade magazine *Metals Progress* reports that significant improvements in production capabilities and mechanical properties of aluminum-lithium alloys have been achieved in the preceding two years.
	The trade magazine *Iron Age* reports that aluminum-lithium alloys may save 15 to 18 percent of the structural weight in future aircraft.
	The second international conference on aluminum-lithium alloys is held, and 38 papers are presented, showing significant growth in interest in these alloys.
	The trade magazine *Light Metals Age* reports that first-generation aluminum-lithium alloys will replace conventional alloys 2024, 7075, and 2014. Second-generation aluminum-lithium alloys, with greater fracture toughness, will replace 2124, 7475, and 7050.
1985	The trade magazine *Aviation Week & Space Technology* reports that a modified heat treatment has proved successful in achieving adequate fracture toughness in aluminum-lithium alloys.
	Alcoa introduces commercial-quality 0.5-in aluminum-lithium plates.
	International Light Metals is developing methods for casting and forging aluminum-lithium alloys.
1986	Alcoa receives a patent on aluminum-lithium alloy process and brings the aluminum-lithium casting facility at Berwyn, PA, on-line.
	A trade magazine asserts safe recycling of aluminum-lithium alloys will challenge producers and users because of toxicity problems.
	McDonnell-Douglas is testing aluminum-lithium alloy for use on wing skins of F-15 aircraft.
	Alcoa will complete the production qualification of aluminum-lithium alloy in 1987.
	Kaiser Aluminum is casting selected aluminum-lithium alloys at its research center.

Appendix

C

User's Guide to TEKFOR

This Appendix provides a brief introduction to the program TEKFOR. TEKFOR is intended to perform the various program functions described in this book. This appendix explains the use of the program. It does not cover the theory behind the program.

The program will run under MS-DOS, and requires at least 256K of memory. The CONFIG.SYS file should include the command:

DEVICE = ANSI.SYS

for the program to achieve the intended screen display. The program will function correctly without the ANSI screen drivers. However, without them, screen clearing will not work, and the screen control characters will be displayed on the screen.

TEKFOR is distributed only for use in conjunction with this textbook. It may not be copied or distributed further without the permission of the author.

TEKFOR is completely menu-driven. All functions are selected from a menu. A menu item may be selected using the arrow keys, followed by <ENTER > , or the user may type in the single capitalized letter in each menu item. Once a menu item is selected, a lower-level menu may be reached, from which further selections may be made. When a particular function has been selected, the program requests specific items of information from the user (e.g., upper limit of a growth curve; confidence bounds desired on a regression, or whether results are to be sent to the printer).

TEKFOR performs the following functions:

1. Creating data files for use by other segments of the program
2. Sorting data and computing medians and quartiles to analyze the data from Delphi sequences
3. Fitting any of several curves to data
4. Running a KSIM simulation model

5. Maintaining files
6. Editing files

Each of these functions is briefly described below.

File Creation

When this option is selected, the user is presented with a menu listing the various TEKFOR functions. The user selects the one for which the data will be used. The user is then given the opportunity to enter the data in spreadsheet format. The data is saved in a file in the proper format for the function selected. NOTE: The data may not be used for any other TEKFOR function, since the file is not in the proper format for any other function.

Medians and Quartiles

The Delphi procedure, a means of extracting opinions from a panel of experts, requires that the panel results be fed back to the panel members after completion of each round of questionnaires. Conventionally, the feedback consists of the median and quartiles of the panel estimates. The data creation portion of TEKFOR allows the user to enter Delphi responses in any order. These responses will then be sorted, and the median and quartiles will be computed and displayed. Additional items may be added to the file as they are obtained, and medians and quartiles can be recomputed.

Curve Fitting

In technological forecasting, it is common to fit S-shaped curves, trends, or log-log curves to data. These curves are projected to obtain a forecast. Data for curve fitting must have been entered into a file as data pairs: time and performance.

TEKFOR allows the user to fit one of three types of S-shaped growth curves to sets of data: Pearl curve, Gompertz curve, and Floyd curve. The user may select one of these curves. After the curve has been selected, the user will be asked to supply an upper limit to the curve (and in the case of the Floyd curve, the performance of the competing technology).

TEKFOR allows the user to fit either a linear or an exponential (semilog) trend to sets of data. The user may select either type of curve to be fitted.

TEKFOR allows the user to fit a log-log fit to data, such as when plotting performance against cumulative production. When this option is selected, TEKFOR fits a straight line to the logarithms of the data.

For any of the types of curves fitted by TEKFOR, the user will be asked what percent confidence bounds are desired on the fitted curve. The user may supply any value between 0 and 100. The confidence bounds are then computed on the assumption that errors are distributed according to a Student-t distribution with $N - 2$ degrees of freedom, where N is the number of data points.

For any of the types of curves fitted by TEKFOR, the user may request extrapolation-interpolation points. The program will compute the value of the fitted curve (and the confidence bounds if that option is chosen) for each of the points. Thus, if a curve is fitted to a set of data for several past years and a forecast is desired for some future year, the future year can be supplied as an extrapolation point. The program will then compute the value of the fitted curve for that future year.

The output of the curve fitting will always be displayed on the screen. It may also be sent to the printer or stored in a file for later editing with a word processor and incorporation into a report.

KSIM

KSIM is a simple simulation model that is often used by technological forecasters as a quick means of understanding the behavior of some system of interacting elements. The user is required to supply an initial value for each of the variables in the model (values in the range of 0 to 1), a matrix of impacts of each variable on the others (values in the range of −10 to 10; some impacts may be zero), and a set of impacts from the "outside world" on each variable. The program then simulates the behavior of the system by integrating the differential equations implied by the impacts among the variables. A time plot of the behavior of the variables is printed out. The user is asked to specify the length of the run, and the number of time steps between the values printed out. The output graph will always be displayed on the screen. It may also be sent to the printer or stored in a file for later editing with a word processor.

File Maintenance

This option permits the user to see the files currently in a selected directory and to delete any of those files no longer needed. To see the files in a directory, the user must specify the drive and complete path, including directory and subdirectory.

File Editing

This option permits the user to make changes in a data file. When this option is selected, the user is asked the name of the file to be edited and the file to which the results are to be written. The results may be written to a new file, or the old file may be overwritten. One important use of this option is changing the values in a KSIM matrix between simulation runs. This allows sensitivity testing, and testing of policy options, without the need to reenter the entire matrix.

Help

At many of the prompts, the user may type the function key <F1 > to receive help on what is expected at that point. In addition, at the main menu, the user may select INFORMATION, to receive instructions about the various choices at the main menu.

Menus

TEKFOR is a completely menu-driven program. At the main menu, the user may select any of several functions. In most cases, the user will then be presented with a lower-level menu with more detailed choices. For instance, choosing INFORMATION at the main menu will lead to a menu that permits the user to select INFORMATION for a specific portion of the program. Choosing curve fitting at the main menu will lead of a menu of the different types of curves that may be fitted to a set of data.

At each menu, the user may move to a menu item using the arrow keys, then select the item by pressing <ENTER > . Alternatively, the user may press the key corresponding to the single capital letter in the name of the menu item.

User Interface

TEKFOR was written with the thought that "to err is human," and it is, therefore, intended to be "user tolerant." The program was written so that an erroneous response (e.g., supplying a letter when a number was expected) will simply result in the prompt being repeated. In addition, at virtually every prompt, two unlisted responses are available. Pressing the escape key <ESC > , immediately returns the user to the next higher menu and cancels whatever the user was doing. Thus, if at any point you feel lost, simply press <ESC > and get back to a starting point. If the user realizes that a response to a previous prompt was in error, the backslash (\) can be used to back up, one prompt at a time, until the point is reached at which the error was made. The correct response can then be given, and progress can be continued from that point. Any responses that have been "backed up" will be wiped out and must be reentered.

If your experience with TEKFOR indicates that a modification or improvement of the program would make it more useful to you, please advise the author. All suggestions for improvement will be given serious consideration, and those selected for adoption will be included in future versions.

Appendix

D

User's Guide to MAXIM

Introduction

This appendix provides a brief introduction to MAXIM. MAXIM is a microcomputer version of the University of Dayton's cross-impact model MAXIM. This guide explains the use of the program. It does not cover the theory behind the program.

MAXIM will run under MS-DOS, and requires 256K of memory. The CONFIG.SYS file must include the command:

DEVICE = ANSI.SYS

for the program to achieve the intended screen display. The program will function correctly without the ANSI screen drivers. However, without them, screen clearing will not work, and the screen control characters will be displayed on the screen.

The version of MAXIM distributed on the disk is intended for instructional and demonstration purposes. A professional version of MAXIM, and training in its use, is available from the author:

Joseph P. Martino
905 South Main Ave.
Sidney, OH 45365

MAXIM

MAXIM is a cross-impact model, intended to be used for forecasting the outcome of complex networks of events. It requires two types of input data: events which might occur and the interactions between pairs of events (i.e., cross impacts).

Events are things that may or may not occur. An event is specified by a name, a time of occurrence, and a probability of occurrence. The simplest event specification consists of a date and a probability of its ever occurring. There are two other, more complex types of events allowed by MAXIM.

The first of these is the event that may occur at any date within a span of time. This type of event is specified by a probability of ever occurring, the first year of possible occurrence, the number of years during which it may occur, and the conditional probability that each year within the span will be the year of occurrence (these conditional probabilities must sum to 1.0). (In MAXIM, the maximum length of the span of years is limited to 10.)

The second of these more complex events is the set of mutually exclusive events: At most one event in the set can occur. The set is specified by identifying which events are included; the probability that one of the events in the set will occur; the date of occurrence for each event; and for each event, the conditional probability it will be the event in the set to occur (these conditional probabilities must sum to 1.0). Sets of events are numbered in sequence: set 1, set 2, etc. An event is identified as belonging to a particular set by giving the number of the set to which it belongs when it is entered. It is not necessary to enter events in the same set one right after the other. Events in a set can be entered in any sequence and may be interspersed with events from other sets, or with events not belonging to any set.

Cross impacts are specified in terms of the changes in timing and probability of the "impacted" event. An event may impact another one either by occurring or by failing to occur. A complete specification of a cross impact includes the change in timing and probability of the impacted event from the occurrence and nonoccurrence of the "impacting" event (ordinarily there will be impacts only from occurrence or from nonoccurrence, but not from both).

MAXIM allows the user to specify events and the cross impacts among them. The user must first specify a model, then verify it, and finally run it. Output of the run is saved to several files, which may be viewed on the screen or sent to the printer. The user may save a model for later use and may load a previously saved model. The user may also modify the current model.

Specifying a model requires that the user define certain model parameters, then specify the events, followed by the cross impacts. When the user is specifying an event, the program asks for a name (limited to 20 characters, upper or lower case or mixed), its probability of ever occurring, whether the event is a member of a set, and the number of years in which it can occur. (Set members can occur in only one year; events not in a set may specify a span of time.) If the event is a member of a set, the program asks for the set number. These numbers must be assigned in sequence, starting with 1. The first member of a set is of course the first event to receive the set number. Set members may be entered at any point during the course of specifying events; all that is required is to enter the number of the set to which the event belongs. If the event can occur in more than one year or is a member of a set, the program asks for the appropriate conditional probabilities.

Specifying the cross impacts requires that the user identify by name the impacting and impacted events, and the changes from the impacts. Ordinarily these changes are entered as the new date and the new probability for the impacted event. The program then converts these changes into "impact factors," in the range from -1 to 1, which are stored for later use. Experienced MAXIM users may wish to enter the impact factors directly, and the program provides this option.

Timing or probability of events, and cross impacts, may be changed by selecting the MODIFY option at the main menu. The MODIFY option does not permit adding events to the model, only modifying events and cross impacts already in the model. Additional events and cross impacts may be added to the model by selecting the SPECIFY option at the main menu, then the ADD option.

Before the model can be run, it must be verified for internal consistency. This option can be selected when the user has completed entry of events and cross impacts. Verifying includes checking that all conditional probabilities sum to 1.0. If an error is found, the user is advised of the nature of the error. It may be corrected by using the MODIFY option.

Once the model has been verified, it can be run. The user will be asked the number of simulation runs desired.

The model output is written to several files. Each of these files can be viewed on the screen or can be sent to the printer. Output from the simulation runs includes frequency of occurrence by year for each event, total frequency of occurrence for each event, and joint frequency of occurrence for each pair of events.

At virtually every prompt, two unlisted responses are available. If the user has made an error at a previous prompt, the backslash (\) can be used to back up, one prompt at a time, until the prompt at which the error was made is reached. The correct entry can then be made, and the program can be continued from that point. All responses that have been "backed up" will be wiped out and must be reentered. At almost any prompt, the <ESC > key can be used to cancel whatever the user is doing and return to the next higher menu.

Appendix E

Uniformly Distributed Random Numbers

37898	39747	61459	73275	45206	34864	39040	40883	92042	62990	70262	90318	86785	96820	55136
30524	53118	45964	19809	75535	08882	16285	34092	32355	81635	35045	90884	22901	35583	73450
27800	43738	52742	67596	47268	02473	15502	93879	64679	79946	68709	42683	64776	00100	74439
33358	26400	99653	60431	78960	47054	44480	82161	44560	89386	76708	43215	97385	06538	63537
98341	38810	87753	46034	76321	25183	26560	11649	35389	82147	52720	11585	56169	91883	63502
88382	03543	49591	04672	98220	93609	76911	36080	46980	68083	64617	45384	61878	06950	36351
64962	70178	24014	49417	05864	40141	10510	51053	03967	72924	68993	91062	37030	24937	86899
56405	45241	47147	61373	70503	12902	22371	38700	82321	56501	34005	75818	56142	40547	93519
58619	87802	81334	69865	64407	98793	91632	06881	94715	37424	77490	04619	68748	59961	04784
97654	88378	72760	30529	24858	69441	92956	71557	97523	36379	96162	39812	68336	84092	63951
91459	09778	77510	68632	36008	80117	23427	19044	34788	10981	84000	82111	35372	19417	91116
51902	06633	42302	65647	83513	91839	21226	37242	65740	94892	62086	86320	61229	16706	81252
69170	72891	80841	50875	94445	22181	57053	30428	45198	12734	96413	71481	27768	55205	40537
09526	17990	76897	69339	02450	51759	67132	76743	05080	40688	93038	24971	45090	10233	54256
41458	50222	93770	64708	40449	84182	21567	27760	10234	91102	90017	48865	55848	88776	55986
14257	01408	89585	11486	68951	04276	07291	59379	85330	25763	04423	34196	15903	81409	30680
87080	87467	11114	52158	86051	08992	69486	04956	07428	26321	04416	55384	86796	64228	87475
46864	64740	31519	59794	84535	64964	11689	34909	51859	09724	38223	05180	15661	01890	82995
76039	90582	04504	48887	34431	15937	31822	67032	80096	51155	19066	71174	38890	15831	15765
29508	58810	03349	42106	07876	20305	45366	93485	05649	78531	54988	19270	62015	31166	67956
47823	45234	86434	42891	94660	11062	24682	49597	81439	23470	97889	83182	28600	91765	69197
43817	54274	95151	28644	28872	52220	86428	45273	51999	78713	22214	55535	00488	22723	29576
86698	10854	16155	67281	70229	17010	28609	41662	22626	77192	34680	23654	25905	40323	56290
85658	40378	76670	34594	48642	75896	43440	63795	15532	44393	97860	99372	59577	58379	41142
93181	15654	65345	11039	38020	65713	99425	65866	49401	43398	44662	49217	12691	44885	28769
02462	96829	55566	99584	97720	24597	30946	64571	98666	04585	40024	06630	38888	22095	06554
18100	52247	89935	54956	00623	07884	61437	17194	07480	04045	43502	19768	94366	49608	44312
56886	61621	05248	73097	37045	53525	33218	02336	63369	69883	38335	67778	40327	60253	46968
65309	17012	38689	73465	04771	87589	80663	47806	79641	75962	83504	68332	04500	39529	54704
61437	44737	86378	11242	50980	85792	58828	96569	45741	39852	43245	39183	91982	97807	89398

Index

ABOUT THE AUTHOR

Joseph P. Martino is a senior research scientist at the University of Dayton Research Institute, where he is responsible for directing the Institute's programs in technology forecasting. Previously, he served in the United States Air Force as director of engineering standardization, Defense Electronics Supply Center. Dr. Martino holds an M.S. in electrical engineering from Purdue University and a Ph.D. from Ohio State University.

DISK WARRANTY

This software is protected by both United States copyright law and international copyright treaty provision. You must treat this software just like a book, except that you may copy it into a computer to be used and you may make archival copies of the software for the sole purpose of backing up our software and protecting your investment from loss.

By saying, "just like a book," McGraw-Hill means, for example, that this software may be used by any number of people and may be freely moved from one computer location to another, so long as there is no possibility of its being used at one location or on one computer while it is being used at another. Just as a book cannot be read by two different people in two different places at the same time, neither can the software be used by two different people in two different places at the same time (unless, of course, McGraw-Hill's copyright is being violated).

LIMITED WARRANTY

McGraw-Hill warrants the physical diskette(s) enclosed herein to be free of defects in materials and workmanship for a period of sixty days from the purchase date. If McGraw-Hill receives written notification within the warranty period of defects in materials or workmanship, and such notification is determined by McGraw-Hill to be correct, McGraw-Hill will replace the defective diskette(s). Send requests to:

Customer Service
TAB/McGraw-Hill
13311 Monterey Ave.
Blue Ridge Summit, PA 17294-0850

The entire and exclusive liability and remedy for breach of this Limited Warranty shall be limited to replacement of defective diskette(s) and shall not include or extend to any claim for or right to cover any other damages, including but not limited to, loss of profit, data, or use of the software, or special, incidental, or consequential damages or other similar claims, even if McGraw-Hill has been specifically advised of the possibility of such damages. In no event will McGraw-Hill's liability for any damages to you or any other person ever exceed the lower of suggested list price or actual price paid for the license to use the software, regardless of any form of the claim.

McGRAW-HILL, INC. SPECIFICALLY DISCLAIMS ALL OTHER WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Specifically, McGraw-Hill makes no representation or warranty that the software is fit for any particular purpose and any implied warranty of merchantability is limited to the sixty-day duration of the Limited Warranty covering the physical diskette(s) only (and not the software) and is otherwise expressly and specifically disclaimed.

This limited warranty gives you specific legal rights; you may have others which may vary from state to state. Some states do not allow the exclusion of incidental or consequential damages, or the limitation on how long an implied warranty lasts, so some of the above may not apply to you.

Engineering

OU STS

2

Cutting edge methods for predicting the future performance of new technologies!

TECHNOLOGICAL FORECASTING FOR DECISION MAKING

THIRD EDITION

The methodology used in technological forecasting is made crystal clear in this essential reference for engineers, managers, government strategic planners, and students. Providing a careful balance of theory and practical applications, the book shows technological forecasting at work in business, industry, and government—with well-chosen examples that illustrate both the strengths and limitations of each technique.

Demonstrating the key point that *technological change can be anticipated, forecast, and managed effectively,* the book contains such special features as:

- The rationale behind each important method in current use—and why it works!
- Vital applications in business, government, and R&D planning
- Highly useful data tables that provide historical examples and case studies
- Complete listings of applicable computer programs, with 3.5" and 5.25" computer disks included
- New sections on probabilistic methods, never before available in book form
- Numerous historical examples and hands-on exercises
- And much more!

For practicing engineers, the book offers the nuts and bolts of preparing technological forecasts, including their mathematical derivations; for managers, it offers a full understanding of how the forecasts can best be applied and what problems may arise. For all readers, it is an indispensable decision-making tool that will greatly enhance their on-the-job effectiveness.

Cover Design: Richard Adelson

ISBN 0-07-040777-0

90000

9 780070 407770

McGraw-Hill, Inc.
Serving the Need for Knowledge
1221 Avenue of the Americas
New York, NY 10020