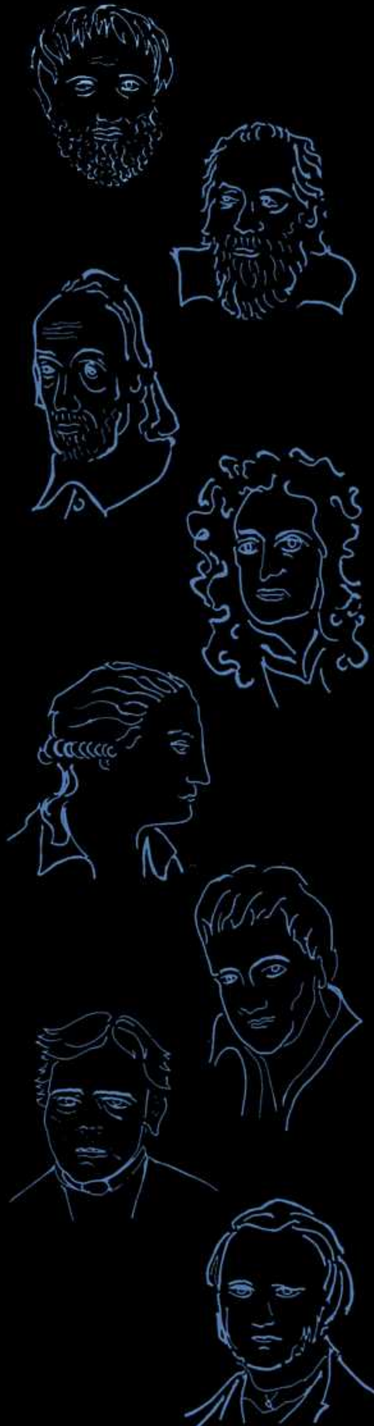
Wiley Series in Probability and Statistics



The Subjectivity of Scientists and the Bayesian Approach

S. James Press and Judith M. Tanur

The Subjectivity of Scientists and the Bayesian Approach

WILEY SERIES IN PROBABILITY AND STATISTICS
TEXTS, REFERENCES, AND POCKETBOOKS SECTION

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

The Subjectivity of Scientists and the Bayesian Approach

S. JAMES PRESS

University of California, Riverside

JUDITH M. TANUR

State University of New York, Stony Brook

Illustrated by Rachel D. Tanur

A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This book is printed on acid-free paper. ♾

Published simultaneously in Canada.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:

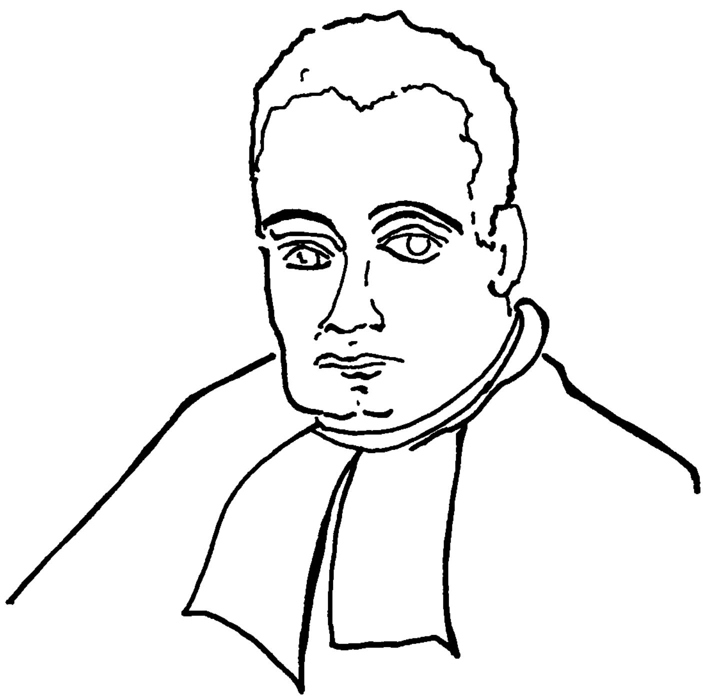
Press, S. James
The subjectivity of scientists and the Bayesian approach / by S. James Press and Judith M. Tanur.
p. cm.—(Wiley series in probability & statistics. Texts, references, and pocketbooks section)
Includes bibliographical references and indexes.
ISBN 0-471-39685-0 (cloth : alk. paper)
1. Research—Methodology. 2. Bayesian statistical decision theory. 3. Subjectivity. I. Tanur, Judith M. II. Title. III. Wiley series in probability and statistics. Texts, references, and pocketbooks section.
Q180.55.M4 .P73 2001
507′.2—dc21

00-043921

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To our spouses and children



The Reverend Thomas Bayes

Contents

Preface ix

1. Introduction 1

2. Selecting the Scientists 17

3. Some Well-Known Stories of Extreme Subjectivity 23

3.1 Introduction, 23
3.2 Johannes Kepler, 23
3.3 Gregor Mendel, 26
3.4 Robert Millikan, 34
3.5 Cyril Burt, 37
3.6 Margaret Mead, 43

4. Stories of Famous Scientists 49

4.1 Introduction, 49
4.2 Aristotle, 51
4.3 Galileo Galilei, 60
4.4 William Harvey, 71
4.5 Sir Isaac Newton, 81
4.6 Antoine Lavoisier, 95
4.7 Alexander von Humboldt, 110
4.8 Michael Faraday, 121
4.9 Charles Darwin, 128
4.10 Louis Pasteur, 143
4.11 Sigmund Freud, 156

4.12 Marie Curie, 166
4.13 Albert Einstein, 177
4.14 Some Conjectures About the Scientists, 189

5. Subjectivity in Science in Modern Times: The Bayesian Approach 199

Appendix: References by Field of Application for Bayesian Statistical Science 225

Bibliography 231

Subject Index 249

Name Index 267

Preface

We expect that both professional scientists and the general public will benefit from reading this book, not only to become better informed, but also for pure enjoyment. Our point of view in this book is that knowledge is accumulated by continually updating our understanding of a phenomenon through the merging of current findings with previous information. We believe that this is how scientists have worked informally in the past, and how scientists might work more formally in the future.

Nonscientists will benefit from reading the book by gaining an increased understanding of how scientists have really worked—an understanding that is different from what they've traditionally been taught. Nonscientists reading this book will surely improve their understanding of the many ways that personal beliefs and biases find their way into supposedly objective scientific results and conclusions. After having read the book, laypeople will be likely to review reported results of scientific studies with a more critical, questioning, and skeptical eye, an important stance in today's increasingly complex society. [See Nelkin (1995) for additional discussion of how the personal biases of scientists affect the way they sometimes communicate their results to the public.]

This book will also be interesting and useful reading for science specialists. They will be reminded of how important it is to treat the data from their experiments with great care and respect, and they will see how their educated scientific beliefs can be introduced into their research in ways that they can use profitably without compromising their results. We believe that most specialists are not likely to have seen the common threads of tenacity of belief in the face of contradictory data that have run through the work of so many of the most famous scientists in history, even such giants as Sir Isaac Newton, Galileo Galilei, and Albert Einstein. These icons of science have had such keen awareness and understanding of the basic principles underlying their disciplines that their intuition and scientific judgment generally helped to lead them to correct conclusions.

Statisticians will find that the book integrates the world of inference into the general framework of science in a way that demonstrates the interdisciplinary nature of their field. They will also be reminded of the broadly pervasive role played by subjectivity.

All readers are likely to benefit from the final chapter. It is somewhat more technical than the earlier chapters because it addresses the subject of *Bayesian statistical analysis* of scientific data. The subject is treated at an introductory, elementary level that is designed to provide an appreciation of how specialists and nonspecialists alike might benefit from adopting such a viewpoint. We believe that this chapter, especially in the context of the rest of the book, makes the volume a useful supplementary text for an undergraduate or graduate course in Bayesian analysis.

This book has had a gestation period of over 12 years. During that time we have learned to work together and help each other's understanding of the issues we were tackling. But we have been helped by others as well. Discussions with colleagues, friends, and family members and the suggestions they made helped us to give more careful thought to the fundamental issues discussed in this book. We are grateful to Gordon Kaufman, John Kimmel, Frank Lad, Teddy Seidenfeld, Arnold Zellner, and several anonymous referees for helpful suggestions for improving the manuscript. We are especially grateful to Rachel Tanur for producing the drawings of the scientists. We are also grateful for the financial support and encouragement given to us by the Department of Statistics, University of California, Riverside, and the Department of Sociology, State University of New York, Stony Brook.

Oceanside, California S. JAMES PRESS
Montauk, New York JUDITH M. TANUR
August 7, 2000

REFERENCE

Nelkin, Dorothy (1995). *Selling Science: How the Press Covers Science and Technology*, rev. ed., New York: W.H. Freeman.

CHAPTER 1

Introduction

This is a book about science, about scientists, and about the methods that scientists use. We will show that the most famous scientists in history have all used their hunches, beliefs, intuition, and deep understanding of the processes they studied, to one extent or another, to arrive at their conclusions. The reader will see that the oft-expressed notion that science is "objective" is only partially true; in fact, science is really a combination of both subjective and objective views.

In this book we tell the stories of 12 of the most famous scientists throughout history, from Aristotle, the philosopher who lived during the era of the ancient Greeks, to Albert Einstein, who lived during the twentieth century. In each case, we discuss how the scientist's preconceived beliefs informally influenced his or her scientific conclusions.

In Chapter 2 we explain that we did not choose these particular 12 scientists to study; they were selected for us. In Chapter 3 we tell the stories of five other very celebrated cases of famous scientists who may have stepped over the lines of acceptable scientific practice. Such overstepping arose either because their convictions about the correctness of their ideas led them to see or accept as accurate only what their theory predicted, or from their zeal to convince the world of the scientific merits of their work.

A major portion of the book is devoted to the stories of our 12 most famous scientists in history (Chapter 4). We examine the lives of these people, their scientific contributions, and the ways in which they used their beliefs together with the results of their scientific experiments to carry out their research. In a final section of Chapter 4 we conjecture about what these scientists might have done had modern methods been available to them for combining their preconceptions about the processes they were studying with their experimental data.

Finally, in Chapter 5, we examine how subjectivity is being used in science in modern times. That discussion focuses on *Bayesian statistical science*. But first, in this introductory chapter, we provide some definitions and background for what is to come.

1.1 SUBJECTIVITY AND OBJECTIVITY

Because we will be using the words *subjectivity* and *objectivity* frequently in this book, we must first explain how we are using these terms. We intend to use them in several ways. But we will always be thinking about observational data, about the distinction between beliefs held by a scientist about a phenomenon prior to collecting the data, and beliefs held by the scientist after the data have been collected and analyzed.

In common usage, certain entities are seen to have only a *subjective reality*, in that those entities are constructs based on views and beliefs that were formed out the human mind. On the other hand, the term *objective reality* is used for entities that exist outside the minds of individuals, in that they exist in the world regardless of whether a person perceives them. For example, the opinions people hold about a political or social issue are personal beliefs that have only a subjective reality, whereas our starting point with respect to external reality is that the moon, the sun, and the planets would all exist regardless of whether or not human beings perceived them. But objective entities need not be corporeal. For example, all the scientific laws that govern the behavior of the physical world would exist regardless of their codification by human beings or humans' belief in them.

The crux of our usage makes these terms specific to the scientific endeavor. When anything is observed or measured by human beings, human perception and sensory mechanisms are involved, and the resulting observations that are collected then depend on subjectively based distortions of objective reality. Further, after entities in the objective world are observed (perceived) by a human being, either through the senses and/or assisted by measuring instruments, the resulting measurements (called *data*) are interpreted by a human being in a subjective way that reflects the person's own experience, understanding, and preconceived notions and beliefs about the entities, the object, phenomenon, or construct being measured. The interpretation of such data is also affected by the person's state of mind and state of senses at the time of the interpretation. We use the term *subjective*, or *subjectivity*, to refer to preexisting views or beliefs about entities that influence both the gathering of data and their interpretation.

Broadening our interpretation of the term, we will also use *subjectivity* to mean a person's intuition, belief, and understanding about some proposition or hypothesis prior to that person collecting observational data that bear on the proposition, or prior to that person obtaining information about the hypothesis. Because the views about the hypothesis that we are talking about are those that a particular person held prior to that person having collected data, or they are his or her views prior to being told about the values of data that have been collected, and because these views may well differ across individuals, we call those views or beliefs *subjective*.

To define *objectivity* in an experimental context, we appeal to a (grossly oversimplified) description of the old-fashioned textbook image of how science proceeds and how scientists behave. In this image, science and scientists are *objective* in the following sense. A hypothesis is developed (the passive voice here is used to signify that little attention is paid to the origin of the hypothesis) and the scientist designs a study to test this hypothesis. After data gathering, whether by designed experiment or by carefully carried out observation in a nonexperimental setting, the scientist dispassionately evaluates the results and their implications for the hypothesis. If the results support the hypothesis, the scientist writes up the study for publication; if they do not, the scientist, again dispassionately, abandons the hypothesis as being wrong, and either revises the hypothesis in light of the new findings and repeats the cycle, or goes on to other concerns.

The subjectivity that we intend to demonstrate that scientists use routinely is both less and more than the opposite of this kind of mythical objectivity. We surely do not suggest that scientists ignore the objective reality that is "out there," nor that they ignore the guidance offered by the results of their investigations. And surely historians of science have long since made it clear that scientists are considerably more human than our oversimplified portrait of dispassionate automaton would indicate. Scientists care about their work, care about their results, and even care about the recognition that will come to them with successful discoveries. These factors often drive the way in which they carry out their research and the way they report it to the world. Despite these factors, some observers of scientific methodology still describe it as "objective."

Strongly held personal beliefs and hypotheses about scientific phenomena, the stuff of what we refer to as subjectivity, are sometimes so strongly held that a scientist will, under their influence, announce confirmatory results from experiments not yet carried out. Such a practice is clearly fraudulent, as is introducing major alterations of data from actual experiments to make them conform to a subjectively held theory. But we need to understand that the dividing line between normal practice and fraud is sometimes a fine one. For example, normal statistical analysis of scientific data often requires the analyst to decide to drop some data points because they lie so far away from what is expected that they seem to be aberrations or mistakes rather than meaningful data that should be considered along with the bulk of the other observations (such data points are sometimes called *outliers*), or because they are poorly measured.

But what we mean by the subjectivity of scientists is deeper than these understandably human traits. Some scientists seem to be particularly opinionated and stubborn, not unlike some nonscientists. These scientists develop hypotheses based on strongly held, preconceived notions of how the world operates, and they sometimes (usually, unconsciously) design studies to prove these notions rather than merely to test them. In cases in which the results of their investigation are contrary to what the hypothesis would predict, a scientist is sometimes more likely to doubt that the data are accurate than to conclude that the theory is incorrect. Such a scientist will redesign the study and persist in trying to find data that prove the hypothesis, sometimes for years, sometimes for a lifetime.

1.2 SUBJECTIVE AND OBJECTIVE PROBABILITY

There is another important and related sense in which we use the terms *subjective* and *objective* in this book; it relates to their meanings in relation to *subjective* and *objective probability*. This use of the terms is discussed in great detail in Chapter 5. For the time being, we merely state that *subjective probability* refers to an individual scientist's degree of belief about the chance of some event occurring. *Objective probability* refers to the mathematical or numerical probability or chance of some event occurring.

1.3 RATE-OF-DEFECTIVES EXAMPLE

The definition of objectivity employed by some philosophers requires that the statement be testable by anybody. Subsumed in this position is the notion that everyone must have the same interpretation of observational data. But we believe that such uniformity of interpretation is rarely the case. We believe that different people come to observational data with differing preexisting views that induce them to construe the interpretation of the data somewhat differently, depending on their preconceived biases. It will be illuminating to begin our discussion of subjectivity in scientific methodology with an example that illustrates how different scientific observers, usually unwittingly, bring their own beliefs and biases—their subjectivity—to bear on the interpretation of scientific data. We see in the example below how different observers of the same data can proceed very differently and thus come away with very different interpretations of them. We call this illustration the *rate-of-defectives example*, and we refer to it later.

Before we present the example, however, we quote Ian Hacking (1965, p. 217), who noted that "[o]ne of the most intriguing aspects of the subjectivist theory, and of Jeffreys' theory of probability, is use of a fact about Bayes' theorem to explain the possibility of learning about a set-up from a sequence of trials on it. The fact seems to explain the possibility of different persons coming to agree on the properties of the set-up, even though each individual starts with different prior betting rates and different initial data." Here Hacking refers to the comforting fact that although different observers of the same data may have differing interpretations of them, eventually, with a sufficiently large number of trials, the differing prior views of different observers about the same data will generally disappear as the data begin to dominate all personal views about the underlying process.

Let us suppose that you collect 100 observations from an experiment. We can refer to these observations as *data points*. You then send these data to five scientists located in five different parts of the world. All five scientists receive the same data set, that is, the same 100 data points. (Note that for purposes of this example, the subjectivity involved in originally deciding what data to collect and in making the original observations themselves is eliminated by sending the same "objective" data to all five scientists.) Should we expect all five of the scientists to draw the same conclusions from these data?

Our answer to this question is a very definite "no." But how can it be that different observers will probably behave very differently when confronted with precisely the same data? Part of the thesis of this book is that the methodology of science requires that inferences from data be a mixture of both subjective judgment (theorizing) and objective observation (empirical verification). Thus, even though the scientists all receive the same observational data, they come to those data with differing beliefs about what to expect and about how to proceed. Consequently, some scientists will tend to weight certain data points more heavily than others and consider differing aspects of the data as more consequential. Different scientists are also likely to weight experimental errors of measurement differently from one another. Moreover, scientists may decide to carry out formal checks and statistical tests about whether the phenomenon of interest in the experiment was actually demonstrated (to ask how strongly the claimed experimental result was supported by the data). Such tests are likely to have different results for different scientists, because different scientists will bring different assumptions to the choice of statistical test. More broadly, scientists often differ on the mathematical and statistical models they choose to analyze a particular data set, and different models usually generate different conclusions. Different assumptions about these models will very often yield different implications for the same data.

These ideas that scientists can differ about the facts are perhaps surprising to some of us. Let us return to our 100 observations and five scientists to give a very simple and elementary example, with the assurance that analogous arguments will hold generally for more realistic and more complicated situations.

Hypothetically, suppose there is a special type of machines that produces a *certain component that we can call a "groove joint," a component required* for the hard drives of desktop computers. It is common knowledge in the computer industry that because groove joints are very difficult to fabricate, such machines generally produce these components with about a 50% defective rate. That is, about half the groove joints produced by a given machine will have to be discarded because they are defective. The machines that produce groove joints are very expensive.

The hypothetical South Bay Electronics Company suspects that its newly purchased machine may be producing groove joints at a defective rate different from the industry norm, so it decides to test the rate at which the new machine produces defectives. It first fabricates 100 groove joints on the new machine (each groove joint is fabricated independently of every other groove joint) and examines each one to classify it as "good" or "defective." South Bay records the sequence of 100 groove joints produced by the new machine as: G, G, D, . . . , with "G" representing "good" and "D" representing "defective." The company finds that there were 90 defective groove joints in this batch, but the quality control staff still doesn't know the long-run rate of defectives for their machine. They decide to send the results representing the 100 tested groove joints to five different scientists to ask them for their own estimates of the long-run rate of defectives for this machine. We shall call this long-run rate of defectives p, bearing in mind that p can be any number between zero and one.

The sequence of G's and D's are the data you send to the five scientists (three women and two men) in five different locations around the world to see how they interpret the results. You tell them that you plan to publish their estimates of the long-run rate of defectives and their reasoning behind their estimates in a professional scientific journal. Thus their reputations are likely to be enhanced or to suffer in proportion to their degrees of error in estimating p. As we shall see, it will turn out that they will all have different views about the long-run rate of defectives after having been given the results of the experiment.

Scientist 1 is largely a *theorist* by reputation. She thinks that $p = 0.5$ no matter what. Her line of reasoning is that it just happened that 90% of the first 100 groove joints were defective, that there was a "run" of "defectives." Such an outcome doesn't mean that if the experiment were to be repeated for another 100 trials (produce yet another 100 groove joints) the next 100 trials wouldn't produce, say, 95 defectives, or any other proportion of defectives. Scientist 1 has a very strong preconceived belief based upon theory that groove joints are about equally likely to be good or defective ($p = 0.5$) in the face of real data that militate against that belief. For her, unless told otherwise, all machines produce defective groove joints for roughly half of their output, even if many runs of many defectives or many good groove joints just happen to occur.

Scientist 2 has the reputation for being an *experimentalist*. He thinks that $p = 0.9$, because that is the proportion of defectives found in the batch. (This estimate of p is what statisticians call the *sample estimate* (and under certain conditions, they also call it the *maximum likelihood estimate*). Scientist 2's definition of the best estimate available from the data is the fraction of defectives actually obtained. While scientist 1 believed strongly in theory, scientist 2 is ready to abandon theory about the production of defective groove joints in favor of strong belief in data, regardless of theory.

When he actually carries out experiments in his research, which is only occasionally, scientist 3 is an extremely *thorough and careful experimentalist*. He carefully examines the actual sequence of outcomes and decides to ignore a run of 50 straight defectives that occurred, feeling that such a run must be a mistake in reporting since it is so unlikely. This reduces the number of available data points (the sample size) from 100 down to 50, and out of those 50, 40 were defective. So his conclusion is that $p = 40/50 = 0.8$. Scientist 3 has taken the practical posture that many scientists take of weighting the observations so that some observations that are believed to be errors are discarded or down-weighted in favor of those thought to be better measurements or more valid in some experimental sense.

Scientist 4 is a *decision theorist*, driven by a need to estimate unknown quantities on the basis of using them to make good decisions. Such a scientist would be interested in minimizing the costs of making mistakes, and might perhaps decide that overestimating p is as bad as underestimating it, so the costs should be the same for making these two types of errors. Moreover, he wants to select his estimator of p to enhance his reputation, and you have told him that a correct estimate will do just that. So he decides to select his estimator such that the cost of being wrong will be as small as possible, regardless of the true value of p. (In the theory of decision making, under such circumstances he would often adopt what is called a *quadratic loss function*.) If scientist 4 were to adopt the subjective belief that all values of p are equally likely and then used a result from probability theory (called *Bayes' theorem*; see discussion of this theorem in Chapter 5), his resulting estimate of p would be $p = 91/102 = 0.892$.

Scientist 5 may be said to be *other-directed*. She badly wants the recognition from her peers that you said would depend on the accuracy of her estimate. She learns from you the names of the other four scientists. She then writes to them to find out the estimates they came up with (and for their own reasons, they send her their results). Having obtained the results of the other scientists, scientist 5 decides that the best thing to do would be to use their average value. (Robert Millikan, winner of a Nobel Prize in Physics, did similar averaging of earlier reported results—from entirely separate experiments in that case—in his oil-drop experiment to measure the charge on an electron; see Section 3.4.) So for scientist 5, the estimate of $p = (0.5 + 0.9 + 0.8 + 0.892)/4 = 0.773$. Or perhaps, learning that scientist 1 made an estimate of p that was not data dependent, scientist 5 might eliminate scientist 1's estimate from her average, making the subjective judgment that it was not a valid estimate. Then scientist 5 would have an estimate of $p = (0.9 + 0.8 + 0.892)/3 = 0.864$. Scientist 5's strategy is used by many scientists all the time, so that the values they propose will not be too discrepant with the orthodox views of their peers.

So the five scientists came up with five different estimates of p starting with the same observed data. All the estimates are valid. Each scientist came to grips with the data with a different perspective, using somewhat different procedures, and with a different belief about p.

The values of p found by the five scientists thus are as follows:

Scientist	1	2	3	4	5
Estimated Value of p	0.500	0.900	0.800	0.892	0.773 or 0.864

But what is the "true" value of p?

In empirical science we never know with certainty what is true and what is not. We have only educated beliefs and the results of observations and experiments. But note that with the exception of scientist 1, who refuses to be influenced by the data, all the scientists agree that p must be somewhere between 0.773 and 0.900, agreeing that the machine is turning out defective groove joints at a long-run rate of greater than 0.5.

Suppose that the machine had produced 1000 groove joints instead of the 100 groove joints we just considered, but with analogous results of 90% defectives. Would this have made any difference to the scientists? Well perhaps it might have, for example, to scientist 1, because people differ as to the point at which they will decide to switch positions from assuming that the machine produces defectives at typical rates despite the experimental outcome results, to a position in which they are willing to assume that the machine has a higher long-run defective rate. One scientist may switch after 9 defectives out of a batch of 10 groove joints, another after 90 out of 100, whereas another may not switch until there are perhaps 900 defectives out of a batch of 1000 groove joints, or might insist on an even more extensive experiment with even stronger evidence. In any case the scientists may still differ in their views about the long-run rate of defectives for the machine for a very wide variety of reasons, only a few of which are mentioned above.

1.4 DIVERSITY IN THE SCIENTIFIC METHODOLOGY PROCESS

The rate-of-defectives example we just discussed was very elementary, but it was science nevertheless. Similar interpretive and methodological issues arise in all branches of the sciences. To examine the issues more broadly, we step back to the origins of science itself, and to the structure, development, and applications of various scientific methodologies.

Wolpert (1992, p. 35) suggests that "unlike technology or religion, science originated only once in history, in Greece." He asserts that "Thales of Miletos who lived in about 600 B.C., was the first we know of who tried to explain the world not in terms of myths but in more concrete terms, terms that might be subject to verification." (Miletos is a city on the Aegean coast of what is now Turkey; at that time the city was considered part of Greece.) This was the *beginning of what has come to be known as science. (In fact, it is claimed that* Thales predicted the total solar eclipse of 585 B.C., and was responsible for proving five theorems of plane geometry, including the fact that an angle inscribed in a semicircle must be a right angle.)

There is no unique definition for the term *scientific method*, a term discussed at length by Francis Bacon, Immanuel Kant, and others. But we can try to characterize the scientific methodology process. We will see that scientists have varied greatly in their methodological approaches to research, so that there has not really been a single, acceptable approach regarded as the "correct" path to scientific achievement. In the following, we use the scientists studied in Chapters 3 and 4 to illustrate these ideas.

Alexander von Humboldt (Section 4.7), a founding father of the field of physical geography, mostly observed and carefully recorded his observations. In some instances, he also hypothesized possible explanations about the mechanisms underlying his observations and showed how the observations fit these hypotheses. He was the quintessential observer. But as physical geography was not viewed in earlier years as a field that required experimentation, he rarely conducted experiments. So the vast bulk of his beliefs about the possible causes of the phenomena he observed were totally subjective, and not based on experimental verification.

Albert Einstein (Section 4.13), who propounded the theory of relativity, mostly hypothesized about his observations, relying heavily on his profound understanding of physics and on the implications of mathematical models he developed for describing physical phenomena to make verifiable predictions about the universe. But he personally almost never carried out experiments that might have demonstrated the validity of his mathematical inferences and subjective speculations about the world. Other scientists, sometimes many decades later, did ultimately check and verify many of Einstein's predictions empirically.

Charles Darwin (Section 4.9), the biologist who propounded the theory of evolution, spent years in the process of observing and recording data meticulously, but then he devoted the next 23 years of his life to theorizing and hypothesizing, trying to generalize from his observations. Predictive mechanisms for the process of biological inheritance were developed by others, such as Gregor Mendel (discussed in Section 3.3).

Sigmund Freud (Section 4.11) treated many people with emotional and mental problems and recorded his subjective observations case by case. He then spent many years trying to generalize his subjective beliefs about his patients to synthesize a theoretical framework from his case studies.

Sir Isaac Newton (Section 4.5), perhaps the greatest of all scientists of all time, observed, hypothesized, developed the mathematics he needed to create mathematical models for use in quantitative prediction about the phenomena he wanted to study, built scientific equipment to improve the quality of his observations, carried out experiments designed to test his hypotheses, and then made and checked his predictions about the world. But as will be seen in Section 4.5, his strong beliefs about his theories may have led him to misrepresent some of his experimental results relating to the speed of sound.

Marie Curie (Section 4.12) was another quintessential experimenter. She carried out experiments, laboriously and painstakingly, extending over many years, experiments designed to isolate the new element, radium. Her subjective beliefs that radium was indeed an element and could be isolated from pitchblende gave her the patience to continue searching until she gathered a sufficient quantity of the precious element to establish its atomic properties.

Some scientists did not attempt to establish quantitative relationships. Rather, they carried out demonstrations to show that certain types of effects existed. For example, William Harvey (Section 4.4) provided a demonstration of the circulation of blood.

So sometimes a scientist merely makes observations; in other cases, the scientist observes and then hypothesizes but does not follow up with experiments; in other cases, the scientist observes, hypothesizes, experiments, or demonstrates, and then perhaps rehypothesizes, and reexperiments, often cycling through the investigative process many times. There is a complete spectrum of methodological behaviors followed by scientists. Most scientists labor in the vineyards to extend other people's theories and end up falling somewhere in between the extremes of the methodological investigatory spectrum. There is no set rule about what procedure every scientist follows. Francis Bacon (1561–1626), a philosopher, attempted to establish some rules about *proper* scientific methodology. His approach was experimental, relying on inductive logic. He was not convinced that mathematics should play an important role in the scientific process, so his method was qualitative. Bacon's approach to scientific methodology was not really adopted until the nineteenth century and was quite limited relative to the broader spectrum of procedures actually adopted by scientists over the centuries.

1.5 CREATIVITY IN SCIENCE

It is most usual that scientific advance starts with an idea or a question in a scientist's head about some phenomenon. [Occasionally, a scientific advance comes about serendipitously, in that the scientist notices something unanticipated but realizes that it is important and thus discovers some unexpected effect; see, for example, Roberts (1989). We confine our attention to more usual, nonserendipitous discoveries.] But how does such an idea or question arise? A scientist observes the surrounding world, as do we all, but when the scientist has observed some phenomenon, perhaps many times, a trained curiosity prompts the asking of such questions as: I wonder what caused that to happen? I wonder what would happen differently if the conditions were changed somewhat? Is the response directly proportional to the stimulus? I wonder what the role of randomness was in that phenomenon?

This process, stimulated by scientific curiosity, of raising questions about the mechanism that generates some real-world effect, sets the scientist to theorizing and formulating hypotheses about the phenomenon. The process demands that the scientist be creatively subjective, and think nonlinearly, much as the artist, composer, or creative writer behaves. The scientist must be able to leap to tentative conclusions about what to expect under certain circumstances, without having yet observed the phenomenon under those circumstances and without passing through the appropriate logical steps (linear thinking) required to arrive at deductive conclusions. This subjectively creative process is fundamental to how science is carried out. The scientist begins searching through a lifetime of personal experience, knowledge, and understanding of related phenomena. A review of the published literature on related topics reveals what is already known about this subject. If the full answers are not known, the exploratory path is open and the scientist begins to pursue the topic with vigor. Thus, the scientist proceeds with the idea, with speculation about it, and the scientific research process begins. Within this creative process, the scientist is most often driven by his or her own biases, intending to find preconceived results (results within a certain range) in the experiment. However, as Kuhn (1962, p. 35) pointed out: "[T]he range of anticipated, and thus of assimilable, results is always small compared with the range that imagination can conceive. And the project whose outcome does not fall in the narrower range is usually just a research failure, one which reflects not on nature but on the scientist."

1.6 THOUGHT (IMAGINARY) EXPERIMENTS VERSUS PHYSICAL EXPERIMENTS

Scientific research usually involves both theory and experiment. That is, a scientist, after theorizing about the mechanism that is involved in a phenomenon, typically carries out experiments that will bear on the mechanisms postulated to be involved. Good experiments are designed to demonstrate whether the hypotheses about observed phenomena were correct. Usually, not all of a scientist's ideas are verified by an experiment or series of experiments, and modifications of the original beliefs must be made and new experiments mounted to test the modified beliefs.

The scientist who merely hypothesizes and does not experiment is being strictly speculative; in such a case, subjective beliefs and the logic (perhaps mathematical projections or predictions) that ties them together are all that are available to justify scientific conclusions. These hypothesized beliefs may be arrived at by "thought experiments" in which the scientist reasons about the phenomenon, and based on a special understanding of how it works, tries to conclude what is likely to happen under various conditions or scenarios. Without empirical verification of the logically derived hypothetical conclusions, the scientist is likely to be wrong, at least in part. In some unusual and remarkable cases, the scientist hypothesizes correctly about the likely outcome of experiments that *could* be carried out but *haven't been yet*. That is, such a scientist correctly predicts what will happen under various conditions. One of the characteristics of the *outstanding* scientist is that these subjective judgments about the phenomenon under study turn out to be correct, even though the conclusions reached do not appear to have been arrived at through logical, step-by-step reasoning, experimental verification, or any combination thereof. There are many instances of this interesting, characteristic trait of the 12 outstanding scientists on whom we focus in Chapter 4. There, we summarize the lives, the scientific achievements, and the methodological approaches to science adopted by these very famous scientists.

Some scientists make the decision to carry out a physical experiment (as distinguished from a thought experiment) that will bear on the phenomenon of interest. Sometimes they design the experiment to capture a *qualitative* effect, such as whether a certain phenomenon will occur under predicted circumstances. For example, will a current flow if we do such and such? In other situations, the scientist designs the experiment to capture a *quantitative* effect, such as how much of something will occur under particular circumstances. For example, how much current will flow if we do such and such?

A scientist who decides to carry out a physical experiment must decide:

1. Which experiment will be carried out (can the phenomenon be observed directly, or will indirect reasoning be necessary)?
2. Which equipment will be used to carry out the experiment? The equipment decision determines how large the observational errors are likely to be, and often, whether the phenomenon can be observed at all.
3. Which hypotheses will be tested by this experiment? The scientist must decide which conditions must be controlled or kept constant during the experiment and which can be varied to generate the phenomenon. The hypothesis about the predicted phenomenon must be clearly stated in advance of the experiment, and the conditions under which the phenomenon will occur must be prespecified and designed into the experiment. If the experiment is quantitative, the scientist must determine which outcome variables must be monitored when the input variables are varied.
4. How will the hypotheses be examined? That is, how "close" to the predicted phenomenon must the experimental outcome be before the scientist will claim to have produced the anticipated effect? How should the scientist measure closeness?
5. What extraneous variables should be controlled or measured?
6. If the scientist is to repeat the experiment on several occasions, how will strange or unusual outcome observations be treated? How many times should the experiment be replicated? How will the results of various replications be combined? How sensitive must the experiment be to observe the phenomenon of interest? What if the phenomenon were to be observed only sometimes and not at others?
7. What should be done if the experimental outcomes contradict the initial hypotheses?
8. How will the results be reported?
9. Which results will be reported to the scientific community so that the work can be reviewed by peers and replicated by others to verify its correctness?

All of these choices the scientist must make are partly subjective and partly objective. While attempting to avoid obvious biases, a scientist wants to use the fundamental understanding of the underlying phenomenon to help arrive at scientific conclusions that will be meaningful. To ignore such preexisting understanding of the phenomenon under study is to fail to take advantage of the best information available to build new knowledge. It would amount to reinventing the wheel every time we wanted to make an improvement in our understanding of transportation methods. We see in Chapter 5 how Bayesian methods of analysis seek to take advantage of this pre-existing understanding by incorporating various aspects of it into the analysis of an experiment.

1.7 ANALYSIS OF RESULTS OF PHYSICAL EXPERIMENTS

The results of physical experiments that scientists carry out are observations, either qualitative or quantitative. If qualitative, they may be *descriptive*, in that they explain in qualitative terms what happened; or they may be *categorical*, in that they summarize whether the predicted phenomenon occurred. When quantitative, they are numbers obtained by observing the phenomenon under study under various conditions. Any of these forms of outcome of an experiment characterizes the results of the experiment. These are the basic raw data. After obtaining such data, the scientist must try to make sense out of them. If the data are quantitative, they must be summarized and understood visually. In any case, inferences and predictions must be made from them. Quantitatively observed data points generally contain errors of measurement. Such errors of measurement are largely random, and thus follow the laws of probability, part of the domain of statistical science. Also in the domain of statistical science is the modeling of the laws that are thought to govern the phenomenon.

The scientists discussed in Chapter 4 did not use the methods of analysis of Bayesian statistical science (or for that matter, with the exception of Einstein, any formal statistical methods). Although it is true that Bayes' theorem had been put forth in the middle of the eighteenth century, many of the methods of analysis that today we call Bayesian statistical inference had not yet been developed; and none of it was available during the lifetimes of many of our scientists. Such methodology, however, is now available for scientific analysis, and makes possible more precise inference from the results of categorical and quantitative experiments. Such methodology is discussed at length in, Chapter 5.

1.8 BLINDING AGAINST EXPERIMENTER BIAS

The subjectivity of scientists has long been well-known to scientists themselves, and consequently, they have established procedures to guard against those effects of subjectivity that may bias experiments and mislead experimenters. In biology/medicine/pharmacology, for example, modern clinical statistical trials have typically required that allocation of treatments to subjects be decided on a randomized basis with some subjects receiving an inert placebo, so that causation can be established. But the safeguard against experimenter bias is via a process called *double-blinding*. In a double-blind experiment, neither the people who diagnose the results of the clinical trials nor the subjects themselves know which subjects actually received the drug being tested for efficacy and which received the placebo. If the experimenter does not know whether the subject he or she is diagnosing has received the treatment or the placebo, his or her diagnostic judgment cannot be influenced by that knowledge. Similarly, a patient cannot attribute changes in his or her condition to a treatment if he or she is not sure whether indeed he or she is getting the treatment or getting a placebo.

In the social sciences, the existence of "experimenter effects" [for example, Rosenthal and Jacobson (1968)] in which the person conducting an experiment is likely to find the results he or she expected, has given rise to a set of procedures by which the person actually running the experiment is kept blind to the hypothesis being tested.

In our discussion of Gregor Mendel in Section 3.3 we see that one of the reasons suggested for his data being "too good to be true" is that when one expects a certain ratio of results and can see that ratio developing as the data are counted, one may be unconsciously led to stop counting when the ratios are as expected. Present-day genetics researchers routinely blind themselves to the results of such counts as they are developing.

In his oil drop experiment to study the charge on the electron, Robert Millikan used a preferential treatment of his data. While he seemed to be choosing to eliminate only observations with large experimental error, in fact he eliminated observations that were contrary to his preconceived ideas, thus permitting his personal biases to determine his inferences about the true value of the charge on the electron. (See Section 3.4 for a discussion of how he weighted his data to conclude what he preferred.)

Recent efforts by "particle physicists" have resulted in a blinding approach to guarding against experimental bias in their field, an approach that has been called *offsets*. "Research teams in particle physics have been programming their computers to add unknown numbers called 'offsets,' to their data to make the outcome of their analyses blind . . . It is only after they experiment is over that the researchers discover what the value of the offset is" (Glanz, 2000, p. D1).

Dr. Eric Prebys, a particle physicist at Princeton University and a member of a large research team called Belle, said scientists "always want a particular outcome whatever they say. Removing all landmarks is the only sure way to insulate them from their own hopes and expectations" (quoted in Glanz, 2000, p. D4). The basic idea is to blind the researchers from seeing the final experi-

Applied Probability and Statistics (Continued)

MONTGOMERY and PECK · Introduction to Linear Regression Analysis, *Second Edition*
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product in Optimization Using Designed Experiments
NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON · Applied Life Data Analysis
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, and SUGIHARA · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
PORT · Theoretical Probability for Applications
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
RACHEV · Probability Metrics and the Stability of Stochastic Models
RÉNYI · A Diary on Information Theory
RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
*SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE, CASELLA, and McCULLOCH · Variance Components
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
THOMPSON · Empirical Model Building
THOMPSON · Sampling
THOMPSON · Simulation: A Modeler's Approach
TIJMS · Stochastic Modeling and Analysis: A Computational Approach
TIJMS · Stochastic Models: An Algorithmic Approach
TITTERINGTON, SMITH, and MAKOV · Statistical Analysis of Finite Mixture Distributions
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume I: Point Pattern and Quantitative Data
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN RIJCKEVORSEL and DE LEEUW · Component and Correspondence Analysis
VIDAKOVIC · Statistical Modeling by Wavelets
WEISBERG · Applied Linear Regression, *Second Edition*
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment
WHITTLE · Systems in Stochastic Equilibrium
*ZELLNER · An Introduction to Bayesian Inference in Econometrics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Biostatistics Section

ARMITAGE and DAVID (editors) · Advances in Biometry
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies
DUNN and CLARK · Applied Statistics: Analysis of Variance and Regression, *Second Edition*
*ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
*FLEISS · The Design and Analysis of Clinical Experiments
FLEISS · Statistical Methods for Rates and Proportions, *Second Edition*
FLEMING and HARRINGTON · Counting Processes and Survival Analysis
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LAWLESS · Statistical Models and Methods for Lifetime Data
LEE · Statistical Methods for Survival Data Analysis, *Second Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
McNEIL · Epidemiological Research Methods
McFADDEN · Management of Data in Clinical Trials
*MILLER · Survival Analysis, *Second Edition*
PIANTADOSI · Clinical Trials: A Methodologic Perspective
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOOLSON · Statistical Methods for the Analysis of Biomedical Data

Financial Engineering Section

HUNT and KENNEDY · Financial Derivatives in Theory and Practice
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance

Texts, References, and Pocketbooks Section

AGRESTI · An Introduction to Categorical Data Analysis
ANDĚL · Mathematics of Chance
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Second Edition*
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and COLTON · Encyclopedia of Biostatistics: Volumes 1 to 6 with Index
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BENDAT and PIERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BHATTACHARYA and JOHNSON · Statistical Concepts and Methods
BILLINGSLEY · Probability and Measure, *Second Edition*
BOX · R. A. Fisher, the Life of a Scientist
BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Texts, References, and Pocketbooks (Continued)

COOK and WEISBERG · Applied Regression Including Computing and Graphics
COOK and WEISBERG · An Introduction to Regression Graphics
COX · A Handbook of Introductory Statistical Methods
DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Sixth Edition*
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
*DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
FREEMAN and SMITH · Aspects of Uncertainty: A Tribute to D. V. Lindley
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
HELLER · MACSYMA for Statisticians
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
KHURI · Advanced Calculus with Applications in Statistics
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MARDIA · The Art of Statistical Science: A Tribute to G. S. Watson
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and *Nonlinear Optimization*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis
RENCHER · Multivariate Statistical Inference with Applications

*Now available in a lower priced paperback edition in the Wiley Classics Library.