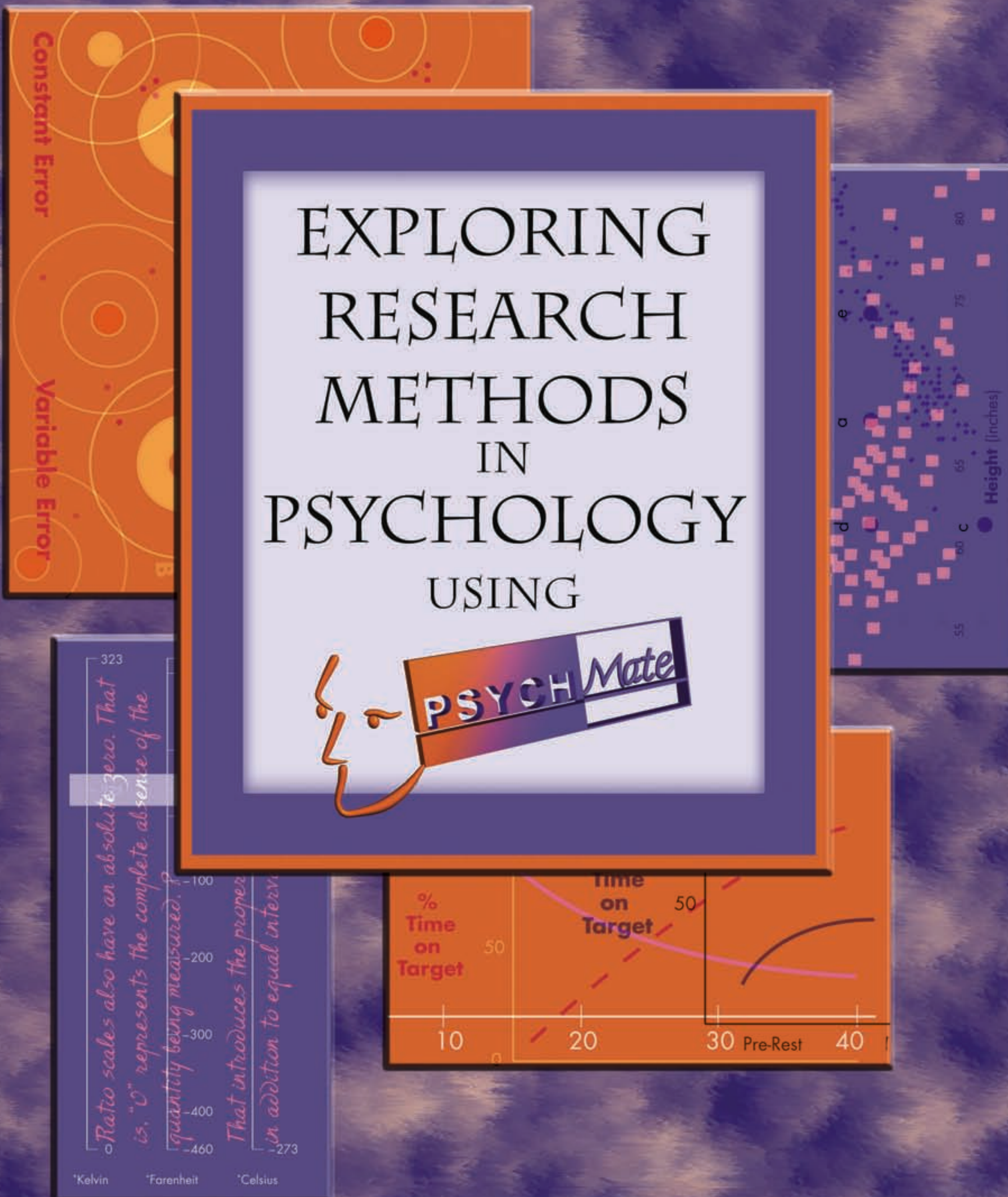
EXPLORING RESEARCH METHODS IN PSYCHOLOGY USING PSYCHMATE



JAMES D. ST. JAMES
MILLIKIN UNIVERSITY

Table of Contents

Chapter 1

Science and Psychology

The intent of this textbook is to acquaint you, the student, with science, emphasizing the application of science to psychological issues. I begin with a definition of science turning not to a dictionary but to a scientist. Ragnar Granit was awarded the Nobel Prize in Physiology in 1967 (along with George Wald and Haldan Hartline) for his work on the electrophysiology of vision.

> Science is the art of acquiring knowledge in such a manner that coherent structures of understanding can be erected on the basis of a critical evaluation of evidence.
>
> Ragnar Granit
> *The Purposive Brain* (p. 21)

Consider the parts of Granit's definition. Science is an art—that is, it is in part a creative act. Scientific theorizing sometimes involves leaps of imagination as daring as any fine art. But science has its specialty, just like any other art—it is "the art of acquiring knowledge." How does that differ from other ways of acquiring knowledge (such as everyday experience)? It differs because scientific knowledge is acquired "in such a manner that coherent structures of understanding may be constructed." That is, we build up better, more complete, more coherent theories and explanations of the nature of the world. But there is a right way to do that. Proposed explanations (hypotheses, theories) are evaluated by examining the evidence. That evaluation of evidence must be a *critical* examination. We do not just accept evidence; we check it out and see if it meets the standards for scientific reliability.

This book is intended to expose you to some of the basics of how to do the science of psychology. Much of what follows applies to science in general. But while all of the sciences can recognize themselves easily in Granit's definition, they differ vastly in the methods particular to their subject matter. Even within a science the methods can vary widely (in psychology, compare the methods of psychophysics and operant conditioning; in biology, compare the methods of population genetics and molecular biology). Within psychology, we have methods for studying memory and methods for investigating the effectiveness of therapy. Neither this nor any other textbook can introduce you to all of those methodologies. Nor are most individual psychologists acquainted with more than a few of them. The laboratory section of this course should acquaint you with specific research methods in psychology. The discussions of library research (see Chapter 4) will help you find information about areas beyond those covered in this textbook or in your particular course.

Students in the research methods course often ask, “Why do I need to know this stuff? I plan to be a clinical psychologist or a counselor, not a researcher.” But practitioners in even the most applied areas of psychology must remain aware of advances of knowledge in their area. Even if you do not go on to do research yourself (and most of you will not), you must be able to *read* research if you are to remain professionally competent. The “critical evaluation of evidence” requires that you have a reasonable familiarity with the methods of the science of psychology and an appreciation for what can (and cannot) be inferred from data. You probably realize that this course cannot give you a complete education in research methodology. You may ask, “So why bother?” For one thing, extravagant claims are common in the areas of mental health, therapy, and counseling. If you know something about what constitutes reliable data, you can at least detect the most egregious blunders. And that makes you better armed than most people. The more you know about how reliable information is gathered, the better you can determine when information does not meet the standards needed for reliability.

A concept currently in vogue in higher education is “critical thinking.” Much current writing and thinking in higher education concerns how we (faculty) can promote critical thinking in you, our students. This discussion recognizes a fundamental issue for education. Too often professors teach, and students learn, as if the process of education was a transfer of information from professor to student: “Memorize this stuff and we will be satisfied.” But that is more and more seen as a hollow process. If students are to be successful in college and beyond, they need to be able to learn on their own. Learning on your own means doing critical thinking—evaluating evidence yourself and reaching conclusions based on evidence. One purpose of this book is to teach you the type of “critical evaluation of evidence” that is implicit in Granit’s definition. Critical evaluation of evidence is not the only kind of critical thinking, but it is an important one.

An unfortunate aspect of teaching students about science is that much of the layman’s “common knowledge” of science is a badly distorted caricature of the real process of generating reliable information. Part of teaching about science is teaching what it is not. These distortions will be discussed as the appropriate topics come up later in this book.

One distortion that needs to be dealt with here is the widespread belief that science is characterized by the cold application of a rigid logic—the scientific method.[1] Related to that is the notion that science is somehow antithetical to an artistic spirit. But science is a highly creative endeavor, and those students who fear that imagination will have to be sacrificed for some mechanical application of logic may be pleased to learn that it just isn't so. The development of scientific theories requires you to have thoughts that no one ever had before! Science requires discipline and clarity, but that discipline and clarity must be in the hands of persons of piercing imagination. Logic is used to confirm discoveries, not to make them (a theme to which I will return at length). Artists also require discipline and clarity in order to make their imaginative products available to others through performance, painting, and the like. They must know what they are trying to do (clarity) and how to do it (gained by discipline). Of course, achievements in the fine arts are not judged against a standard of evidence. It would be absurd to ask whether abstract expressionism is "correct." That marks a difference between art and the sciences. But the fact that scientific evidence plays no part in judging artistic achievements does not imply that imagination plays no role in science.

Relationships among Variables

Science is concerned with the relationships among variables. *Such relationships provide the starting point for many scientific investigations.* Why is recognition typically better than recall? Why are males faster at rotation of mental images and more likely to be dyslexic than females? Here, a known relationship becomes a fact to be explained. We have at best a poor idea of how memory works if we cannot explain why recognition is (usually) better. We are missing a piece of the puzzle of sex differences if we cannot account for differences in imagery.[2]

Relationships among variables also provide us with tests of our explanations for various phenomena. Theories and hypotheses usually predict certain relationships among variables, and evaluation of those hypotheses often rests on whether or not the predicted relationship holds. In 1973, Ewald Tulving and D. M. Thomson (see *PsychMate 2.7: Recall, Recognition, and Encoding Specificity*) showed that a prediction from one type of explanation for the superiority of recognition—that subjects would be unable to recall words that

[1] Mr. Spock of *Star Trek* is a fine caricature of this view of science.

[2] Sex differences in cognition are about evenly balanced between those favoring males and those favoring females. Males are faster at rotation of mental images (see *PsychMate 1.3: Rotation of Mental Images*), but are also about four times as likely to be dyslexic.

they could not recognize—was false. Because the relationship did not hold, the hypothesis was probably wrong.

Issues of how we can test theories or hypotheses are addressed throughout this book—indeed, theories and hypotheses are how we produce reliable knowledge. My aim is that, after studying this book, you will find that you think more clearly about variables and their relationships.

I begin with a brief delineation of ways that people know about things, and some of the limitations to those ways of knowing. Next is a discussion at some length of the issue of *control*, which is a basic issue for science. That topic will recur frequently throughout this book. Next is a comparison between scientific discovery and the process of a criminal trial. Like all analogies, this one does not lead to a perfect parallel, but it does lead to some insight. The chapter ends with an example of a real, scientific detective story.

Ways of Knowing

Science is but one of the ways that we gain knowledge of the world. It is used by few people, and they (us!) use it sparingly.[3] It is, however, the only way that produces reliable knowledge. When I write that science produces reliable knowledge, I do not mean perfect knowledge. Any knowledge derived from experience (and there is no other kind) is subject to error. But there are many things we can do to rule out error, and those things make up science. It is worth contrasting it with other ways of knowing.

Casual observation and anecdotal evidence.

Probably most of what we know of the world comes through casual observation. By *casual* observation I mean simply what we notice as we go about our daily lives. Casual observation has many pitfalls. Humans are not usually very good observers—especially when they are not prepared to observe. The problems of eye-witness testimony are well known (Elizabeth Loftus, 1993, provides an introduction to this literature). A serious problem with casual observation is that you are not usually expecting to make an observation—you may not be looking in the correct direction or paying attention to the appropriate part of the scene before you.

In addition to errors of observation, errors of memory plague casual observation. We often store in memory our *interpretation* of something we see

[3] When you decide which bread to buy or which soap, are you making an informed, scientific decision? I doubt it. Most of our everyday decisions are grotesquely unscientific, even among scientists.

or hear, rather than the actual experience.[4] Jacques Barzun (2000), a cultural historian, put the affair nicely: "The mind is an impressionable organ rather than a recording instrument" (p. 224). *PsychMate 1.8: Change Blindness* discusses at greater length the problems of casual observation.

To make matters worse, observers may be deliberately misled. A magician's greatest tool is misdirection—getting people to look away at a critical point in a magic trick. Of course, this is aided enormously by the combination of our poor peripheral vision, as demonstrated in *PsychMate 1.1: The Filling-In of Blind Spots: Induced Scotomas,* and our usual unawareness of how poor our peripheral vision is. In the heyday of the "materialist" spiritualist mediums, who produced "ectoplasmic" forms of the dead, the séances were typically held in near-darkness—not the best conditions for investigation![5] Chemists and physicists have an advantage over psychologists in that nature does not try to deceive (though it also doesn't try very hard to reveal itself!). But psychologists, insofar as they study humans, have subjects who sometimes lie and cheat.[6]

Anecdotal evidence is evidence by story-telling. If someone tells you about the behavior of a friend and speculates about why the person behaved that way, that is anecdotal evidence about that behavior. Unfortunately, we can seldom rely on such evidence. Anecdotal evidence is usually *somebody else's* casual observation! That does NOT mean that it is wrong. It does mean that it is weak evidence at best and meaningless unless there is supporting evidence. That having been said, a good anecdote can help make an argument vivid and memorable. Many popular discussions of disorders such as dyslexia begin with a description of a single case of the disorder—an anecdote. The discussions then go on to present information. In such a case, the anecdote serves a legitimate purpose—it puts a human face on the disorder and serves as an "advanced organizer" (David Ausubel, 1960) that lets the reader or listener know what is coming, hence improving comprehension and memory (see *PsychMate 2.6: Organization in Memory as an Aid to Recall*).

Though casual observation and anecdotal evidence can seldom provide reliable evidence, they do often provide the hints needed to begin a more

---

[4] William Brewer (1977) presented a fine example of this with a study of what he called "pragmatic implications." He presented subjects with sentences such as "The clumsy chemist had acid on his coat." After a few minutes, subjects cued with "The clumsy chemist..." frequently falsely recalled "*spilled* acid on his coat." Of course, the pragmatic (practical) implication of a clumsy chemist having acid on his coat is that he spilled it, but that was not stated.

[5] For a fascinating account of the materialist mediums, and some photographs of ectoplasmic materializations, see Carl Murchison's (1927) *The Case For and Against Psychical Belief,* especially the chapters by Joseph Jastrow and Harry Houdini.

[6] Lying, incidentally, is not limited to humans. Guy Woodruff and David Premack (1979) reported experiments in which they induced chimpanzees to lie—though not verbally.

formal investigation. Casual observation (whether by yourself or someone else) can *suggest* the importance of variables and the relationships among those variables, but it cannot establish them.

Authority.

> However little authority may be depended upon, it possesses, nevertheless, a name of honor, and habit is more strongly inducive to error than authority; but popular prejudice is more forceful than either of them. For authority merely entices, habit binds, popular opinion makes men obstinate and confirms them in their obstinacy.
>
> Roger Bacon
> *Opus Majus* (1267/1928, p. 10)

Much of what you believe about the world has come from authority—someone you recognize (correctly or not) as more knowledgeable than you. We *must* gain much of our knowledge from authority. We don't live long enough to become experts in all the areas that affect our lives. We entrust the care of our bodies and cars to our physicians and mechanics, often with little personal understanding of how they were fixed. We rely on scientific authorities for (sometimes confusing) advice about what to eat or what not to eat.

The problem with authority as a way of knowing is that not all authority can be trusted. Many persons who set themselves up as authorities are either honestly mistaken or outright charlatans.

In the sciences, we must rely on research reported by others. We cannot independently verify every discovery. (This issue is addressed again in the discussion of replication of research findings in Chapter 7.) And while outright fraud is probably relatively rare in the sciences, experimental results can be wrong for many reasons other than dishonesty. Perhaps the best approach to authority as a source of knowledge is summed up in a phrase that came out of the nuclear arms negotiations between the United States and the Soviet Union during the Cold War—"Trust, but verify."

Religious revelation.

An important source of knowledge for many people is religious revelation. In a sense, this is knowledge given by the highest authority—the word of God. But there is a problem with religious revelation; most of it is wrong. I assert that boldly because it is undoubtedly so. Many religious claims, based on religious revelation, contradict each other. Therefore, they cannot all be correct. The problem is in knowing *which* religious revelations are correct. Though religious issues are of considerable importance to people's lives (Is there a heaven? How do I get there?), science offers little by way of help in deciding among religious claims. The essential problem is that most religious claims are not subject to falsification, which is discussed in more detail in Chapter 2. Science cannot dispute religious revelation. It is simply powerless to decide among competing religious claims. Consider, for example, how you would test the competing claims concerning infant baptism versus adult baptism—a very serious doctrinal dispute among Christians with both sides claiming Biblical sanction (religious revelation) for their views.[7]

Systematic observation.

While casual observation serves as our principal way of knowing about the world, we gain far greater reliability in our knowledge when we observe systematically. Essentially, this means observation that is conscious and deliberate (as opposed to accidental). Systematic observation implies that some variables of interest have been identified, and that those variables are being measured or classified. There is a contemporaneous recording of the data so memory lapses are less of a problem. With systematic observation, then, we enter the realm of science.

Experimentation.

Experimentation relies upon, but also improves upon, systematic observation. One improvement is to make events happen when we can best observe them. When we rely on naturally-occurring phenomena, they may not oblige us by happening when we are looking, when we have the right measuring instruments, or when other conditions are controlled. But when we *cause* the phenomenon to happen and know when it will happen, we have aides to observation at hand (a microscope to see something small, or a computer to record a reaction time) and can control for other factors that might affect the phenomenon.

A second improvement is that we actively manipulate a variable or variables to see what effect they will have. We are no longer at the mercy of waiting around until an event occurs—we make it happen.

Both systematic observation and experimentation represent a scientific approach to gaining knowledge. They have real checks on the quality of the

[7] An Internet search for "infant versus adult baptism" will quickly find the arguments on both sides.

evidence and, hence, produce *reliable* knowledge. This does not mean that they are invariably correct. Indeed, the history of science is a history of frequent error. Nevertheless, a better and better approximation to what O. Kempthorne (1976) called a "validated model of the world" (p. 32) is a tremendously valuable asset.

Robert S. Woodworth (1938) was one of the first psychologists to comment on the difference between the "experimental method" and the "correlational method," with the latter based on systematic observation but lacking experimental manipulation. He also pointed out that the correlational method "*does not directly study cause and effect*. The experimentalist's independent variable is antecedent to his [sic][8] dependent variable; one is the cause (or part of the cause) and the other effect. The correlationist studies the interrelation of different effects" (p. 3, emphasis added). You will often see this phrased as "correlation does not imply causality."[9] Indeed, it does not. But the basic problem for correlational studies—based on systematic, scientific observation and measurement, but lacking experimental manipulation of at least one variable—is that it is usually easy to suggest alternative explanations of cause and effect. Experiments do a better job of ruling out alternative explanations. (Thomas Cook and Donald Campbell (1979) provide an excellent discussion of the difficulties of *causality,* what it means that one thing causes another, and how to establish causality.)

Note that the role of experimentation is to rule out two kinds of alternative explanations. One kind of alternative explanation is the kind that challenges the integrity of the experiment itself (confounds or threats to internal validity—see Chapter 5). Another kind of alternative explanation challenges the theory being tested—i.e., alternative explanations for the phenomenon.

---

[8] "Sic" is from the Latin meaning "thus." It indicates that the original is quoted exactly as written, and is used where the original was ungrammatical or, as here, is now dated. It is no longer customary to assume that all persons are male. The square brackets—[ ]—indicate something has been added to a quote.

[9] This phrase is sometimes (especially in Introductory Psychology and Statistics textbooks) taken as implying that when we analyze data using a correlation we cannot make causal inferences. A correlational study—one that establishes a relationship between variables but does not manipulate them—does not imply causality. But a correlational study may or may not be analyzed using the statistical technique of correlation. And an experiment—which does manipulate the variables—may well be analyzed with a correlation. For example, Sternberg's measure of the speed of search of short-term memory (see *PsychMate 2.2: Scanning Short-term Memory*) relied directly upon the slope of the regression line, when the correlation between RT and memory set size was very high.

Control

One of the most important concepts in regard to scientific investigation is that of *control*. We speak of *control groups*, *control conditions*, and simply *controls*. They are related by the fact that they all address one of the fundamental aspects of scientific investigation—ruling out alternative explanations.

The following discussion owes much to Edwin G. Boring's (1954) classic discussion "The nature and history of experimental control." I will begin by quoting at some length his account of an early experiment with very modern controls.

> The concept of control is pretty old and was quite obvious once the Renaissance had turned men's thought from theological fiat to experiment as the means for penetrating into nature's secrets. Here is a story that makes the whole matter clear.
>
> In 1648 the Torricellian vacuum was known to physics in general and to Pascal in particular. This is the vacuum formed at the upper closed end of a tube which has first been filled with mercury and then inverted with its lower open end in a dish of mercury. The column of mercury falls in the tube until it is about 30 in. high and remains there, leaving a vacuum above it. Pascal was of the opinion that the column is supported by the weight of the air that presses upon the mercury in the dish (he was right; the Torricellian tube is a barometer) and that the column should be shorter at higher altitudes where the weight of the atmosphere would be less. So he asked his brother-in-law, Perier, who was at Clermont, to perform for him the obvious experiment at the Puy-de-Dôme, a mountain in the neighborhood about 3000 ft. ("500 fathoms") high as measured from the Convent at the bottom to the mountain's top. On Saturday, September 19th, 1648, Perier, with three friends of the Clermont clergy and three laymen, two Torrecellian tubes, two dishes and plenty of mercury, set out for the Puy-de-Dôme. At the foot they stopped at the Convent, set up both tubes, found the height of the column in each to be 26 old French inches plus 3 1/2 Paris lines (28.04 modern inches), left one tube set up at the Convent with Father Chastin to watch it so as to see whether it changed during the day, disassembled the other tube and carried it to the top of the mountain, 3000 ft. above the Convent and 4800 ft. above sea-level. There they set it up again and found to their excited pleasure that the height of the mercury column was only 23 French inches and 2 Paris lines (24.71 in.), much less than it was down below just as Pascal had hoped it would be. To make sure they took measurements in five places at the top, on one side and the other of the mountain top, inside a shelter and

> outside, but the column heights were all the same. Then they came down, stopping on the way to take a measurement at an intermediate altitude, where the mercury column proved to be of intermediate height (26.65 in.). Back at the Convent, Father Chastin said that the other tube had not varied during the day, and then, setting up their second tube, the climbers found that it too again measured 26 in. 3 1/2 lines. These are reasonable determinations for these altitudes, showing about the usual one inch of change in the mercury column for every 1000 ft. of change in altitude.
>
> In this experiment there was no elaborate design, and it took place 195 years too soon for the experimenters to have read John Stuart Mill's Logic, but the principle of control and of the Method of Difference is there. How important it was for them to have left a barometer at the base of the Puy-de-Dôme to make sure that changes in the tube that they carried up the mountain were due to elevation and not to general atmospheric changes or to other unknown circumstances! How wise of the party at the top to have made the measurement under as many different conditions as they could think of with altitude constant! How intelligent of them to take a reading on the way down and thus to turn the Method of Difference into the Method of Concomitant Variation! (pp. 577-578)

Indeed, these were wise decisions. A barometer's reading can change over the course of a day as a result of changes in the atmosphere (though that was unknown to the participants). That might have led to an erroneous conclusion but for the knowledge that the barometer at the foot of the mountain had not changed. Furthermore, showing that the readings were unaffected by changes in location that were not accompanied by changes in altitude helped to test what I will later call the "boundary conditions" of the phenomenon.

The Method of Difference and Method of Concomitant Variation mentioned in Boring's quote are two of the formal scientific methods proposed by James Stuart Mill, the great British philosopher, in his *A System of Logic, Ratiocinative and Inductive* published in 1843. Mill was not the first to describe these scientific methods—being anticipated by David Hume in1739 and Francis Bacon in 1620—but Mill wrote when science was becoming more formalized and developed the ideas more completely. Mill proposed several methods of scientific inquiry, which are outlined briefly.

First is the Method of Agreement. "If *A* is always followed by *a*, then *A* is presumably the cause of *a*" (Boring, 1954, p. 574). But this really will not do, as Mill himself noted in remarking that mere agreement would lead us to conclude that night causes day and that day causes night. Indeed, this sort of logical fallacy has a Latin name: *post hoc, ergo propter hoc* ("after it, therefore because of it").

Ambrose Bierce, an American author and wit, provided the following example of the weakness of the Method of Agreement in his *Devil's Dictionary* (first published about 1911): "Effect, n. The second of two phenomena that always occur together in the same order. The first, called a Cause, is said to generate the other—which is no more sensible that it would be for one who has never seen a dog except in pursuit of a rabbit to declare the rabbit the cause of the dog" (1946, p. 229). Mill noted that this method is strengthened considerably if we can vary *A* "at will." When we cannot, we have what we now call a correlational study. Boring put it this way: "Mere agreement does not furnish rigorous proof, although you may be limited to it when you lack the voluntary variation of events—the independent experimental variable—and are reduced to description only. For this reason the establishment of causal relations in biography, history, geology, paleontology, and even astronomy is less sure than in experimental science" (1954, p. 574).

Mill's second method is the Method of Difference. "If *A* is always followed by *a* and not-*A* is always followed by not-*a*, then *A* is certainly the cause of *a*" (Boring, 1954, p. 574). If *A* is dichotomous (that is, it has only two values—present or not present), then we have the Method of Difference. But when *A* can take on a range of values (for example, varying the amount of study time in a memory experiment, or varying dosages of a drug), then we use the Method of Concomitant Variation. Here the value of *a* is shown to vary concomitantly with changes in the amount of *A*. It is precisely this method that was used in Pascal's experiment, when readings were taken with the barometer at three different altitudes.

Pascal's experiment also illustrates what I will term "control conditions." The barometer left at the base of the mountain served as one kind of control. The repeated readings taken at varying heights and at varying locations at the same height provided another. Other uses of "control" refer not to control conditions for explicit comparison, but rather to any techniques used to avoid variation in experimental conditions. Testing all subjects at the same time of day or calibrating laboratory instruments against known standards are examples of this kind of control. These I will refer to as "experimental controls."

The use of a "control group" in its modern sense is actually relatively recent. R. L Solomon (1949) wrote, "If one is interested solely in the concept of Control Group in the history of psychology, Twentieth Century psychology contains that history. We have not been able to find a single case of the use of control group design, as we use it today, before the year 1901. Control group designs seem to have awaited the development of statistical concepts which allow for the characterization of group performances in terms of measures of central tendency; and psychologists seem to have been slow to combine statistical sophistication with experimental design" (p. 137). Solomon noted that early experiments on training typically used a pretest followed by training and then a posttest with the difference between pre- and posttest scores indicating the amount of learning. According to Solomon, the first adequate experiment on training was that of Winch (1908) who used two groups of subjects.[10] One group received a pretest, then training, then a posttest. The other group was the control group. They received the pretest, then *no* training, then the posttest. Boring notes the interesting coincidence that "the statistical techniques for measuring the significance of group differences were being invented or discovered just about the time group controls came into general use" (p. 585). Indeed, "Student's" paper detailing the *t*-test for mean differences was published the same year as Winch's experiment on training (though Winch did not take advantage of Student's test).[11]

An analogy

A useful analogy to the process of scientific discovery is the process of a criminal trial. Both processes are ultimately concerned with finding the truth, and both processes have more than a few pitfalls. In a criminal proceeding, a prosecutor, based on evidence developed by the police, presents a theory of the crime to a jury or judge. That theory is that the accused committed the crime in some specific manner. Note that it is up to the prosecutor to provide evidence in favor of his or her case (theory). That is the sense in which the accused is "innocent until proven guilty"—if the prosecution does not make its case, the jury must acquit. As a case is presented, the judge acts as a referee and can rule out certain evidence (such as most hearsay testimony) that does not meet the proper standards.

The role of the attorney for the accused is to try to demonstrate any logical or evidential flaws in the prosecutor's case. The judge or jury then makes a decision concerning guilt. A verdict of "guilty" indicates that they found the

[10] "The control group in transfer experiments was employed in a small way by Thorndike and Woodworth (1901)..." (Woodworth, 1938, p. 178).

[11] "Student" was the pseudonym of Edwin Gossett, a pioneer of statistics, who was employed by the Guinness brewery of St. James' Gate, Dublin, Ireland, makers then and now of Guinness Stout and other fine products of the brewer's art. Gossett was an early and successful developer of what is now termed *statistical quality control*. Company rules forbade him from publishing under his own name, (lest his publications somehow reflect poorly on the company?), so he was required to use a pseudonym. He chose marvelously—what better for a scientist than "Student"? The small community of researchers in the newly-developing field of inferential statistics was mostly British, and was, of course, well aware of the identity of "Student."

evidence for the prosecution's case against the defendant persuasive "beyond a reasonable doubt." A verdict of "not guilty" is more complex. It may indicate that the jury did not find the evidence persuasive at all, and they believe the defendant to be innocent. It may, however, merely reflect that the jury did not find the evidence compelling—they may feel that it is somewhat likely that the defendant is guilty, but there is still reasonable doubt.

Let us examine the analogy to a scientist trying to persuade others of the soundness of a theory. If you present a theory of why or how some phenomenon occurs (e.g., the cause of schizophrenia or the nature of different memory systems), it is your responsibility to show that your theory is correct. The "jury" of other scientists can quite reasonably refuse to accept your theory unless you have evidence to back it up. Scientists favor evidence based on the results of experiments, as opposed to the eye-witness testimony and circumstantial evidence of a courtroom, but otherwise the logic is the same. Who acts as judge, deciding what evidence is admissible? To a considerable degree, that role is played by peers, who review articles submitted for publication and by the editors of the scientific journals. The body of scientists who deal with the area in question will serve as both defense counsel and jury. Some may reply to the published article, pointing out flaws in the "case," thus acting as defense counsel. Others will decide whether they find the case sufficiently compelling that they will accept the theory (at least provisionally). These scientists are playing the role of the jury. If they find the evidence compelling, they will make that theory a part of their intellectual equipment, and try to find ways to fit the new theory with older ones (or modify older ones to adapt to the new knowledge). They may also incorporate that theory into their teaching. But if other scientists do not find the evidence compelling, they may either reject the theory altogether (the defendant did not commit the crime) or hold a decision in abeyance until more evidence is available—there is still reasonable doubt.

There are some differences between scientific theory-testing and a criminal trial, of course. One important one is that a scientific case is never fully settled. In a criminal trial, new evidence discovered after the trial does not automatically re-open the case, but, at least in principle, a scientific theory is always subject to an appeal due to new evidence. Another difference is that a jury must reach a unanimous decision or the whole trial is just thrown out. Scientists are not required to reach agreement as a jury is. Some may find the evidence compelling and teach the theory or incorporate it into their own work. Others may reject it.

One very important way in which jury decision-making is like that of the body of scientists is in the fact that "proof" is never absolute. It is beyond a *reasonable* doubt, but not beyond all doubt. As noted in Chapter 2, empirical evidence simply cannot prove beyond all doubt. Data (evidence) can suggest, support, and even convince, but it cannot prove beyond any possible doubt.

Much of the rest of this book is about scientific detective work—finding the evidence that can both help create a coherent theory of a phenomenon and provide evidence to support that theory. Just as in police work, careful collection of evidence is required. Some leads will turn out to be false, just as most hypotheses about a phenomenon may be wrong. Sometimes the evidence will seem to be very clear, but still turn out to be wrong. While no path to knowledge is perfect, a careful respect for the quality of the evidence will eventually lead to something pretty close to the truth.

Science: An Example

It may be useful at this point to consider an illustrative example of scientific research. Because of the clarity and importance of his work, I choose John Snow's investigation of the cause of cholera during the London epidemic of 1854.[12]

Cholera is a water-borne bacterial disease usually contracted from water supplies contaminated by feces. Direct person-to-person contact rarely spreads the disease, though poor sanitation in caring for patients with the resulting diarrhea often aids the spread of the contamination. The disease produces gastro-intestinal symptoms of diarrhea, cramps, and vomiting. Patients die of dehydration—sometimes within hours. Cholera is now largely unknown in Western Europe and the United States. There, water treatment and filtering almost completely eliminate the bacterium that causes the disease. However, outbreaks of the disease can occur when flooding overwhelms water treatment plants and ground water becomes contaminated by mixing with sewerage. In poorly developed countries, cholera remains a threat. It should also be noted that nursing care, in the form of rehydration via intravenous fluids, is the normal method of treatment. At the time of the early pandemics, however, intravenous rehydration was not yet known.

Cholera began historically as an endemic disease on the Indian sub-continent.[13] It spread in a series of pandemics beginning in about 1816 when it spread across India. A second pandemic, beginning in 1829, saw it spread to

[12] This discussion is largely based on information from the UCLA medical school web site on John Snow (www.ph.ucla.edu/epi/snow.html) and the web site of the John Snow Society (www.johnsnowsociety.org).

[13] An *endemic* disease is one restricted to a particular locality, while a *pandemic* disease is one spread over a wide area with a higher rate of infection. An *epidemic* disease is one with a sudden, rapid spread.

Europe, reaching London and Paris by 1832. There were more than 23,000 deaths in England and more than 100,000 in France. By mid-century, the pandemic had spread to North America. As recently as 1994, an outbreak beginning in Peru killed 10,000 and infected more than one million. The improved methods of rehydration have reduced the death rate enormously in areas where adequate nursing care is available.

John Snow (1813-1858) was a London-educated physician of broad interests. He was one of the first British physicians to pursue the use of the new anesthetics developed in the United States in the 1840s. He made improvements in administration of the anesthetics and, in 1853, administered obstetrical anesthesia to Queen Victoria during the birth of Prince Leopold. Snow began his investigations of cholera during an outbreak in London in 1848-49. During that epidemic, there were at least 250,000 cases and 53,000 deaths in England. Noting that the symptoms were intestinal, Snow argued in his 1849 book, *On the Mode of Transmission of Cholera*, that it was a water-borne disease. Medical opinion on the cause of the epidemic was largely in favor of some kind of air-borne pathogen. Note that at that time the role of bacteria in infectious disease was only beginning to be understood.

It should be noted that, especially in the poorer parts of London (and most other cities), sanitation was dreadful. Sewers were often just open trenches that ran down the middle of streets into which people dumped their "night soil" in the morning. Sewerage ran off into the River Thames untreated. The waste of cattle and horses was largely left in the streets for the rain to wash away.

Another outbreak of cholera occurred in the Soho district of London in August of 1854. Over a very hot summer, there had been a few cases of cholera in the area, but on August 31, the disease reached epidemic proportions with 56 new cases in a single day. By the next day the death toll was 70 with 143 new cases within an area of a few blocks.

Snow began to investigate the latest outbreak on his own as he had no official standing. Snow's approach has become standard for epidemiology—he drew a map. This is sometimes cited as the beginning of modern epidemiology. Snow reasoned that if the prevailing theory of transmission by a mist or miasma—essentially by bad air—were correct, there should be a relatively uniform distribution of cases along the streets of the affected areas. Snow's map plotted the number of deaths for each building in the affected area. The data initially challenged the theory of transmission by air and appeared to support Snow's own theory of transmission by contaminated water because the overwhelming majority of cases occurred within walking distance of the Broad Street pump—a hand pump available to the public. An inspection he made of the pump, however, failed to find contamination. Snow consulted the Register of Deaths and plotted more cases. In doing so he found details that seemed to undermine his theory. A 500-inmate workhouse in the neighborhood (where the homeless could find shelter and a bit of gruel in return for a loss of freedom and hours of manual labor) had only five deaths by cholera. At a large brewery adjacent to the pump, there had been no cases. Even worse, deaths were reported in the rural villages of Hampstead and Islington some distance away.

Snow went on to ask people who remained in the district about details of the outbreak. He discovered that the workhouse had its own well and did not use the Broad Street pump. Also, the workers at the brewery drank only their own product at work. When Snow visited the village of Hampstead, where a death had occurred far from the pump, relatives of the dead told him that the deceased had a jug of water delivered regularly from the Broad Street pump because she preferred its flavor. On further inquiry, he found that her niece had visited her, drank from the jug, and later died at home—in Islington.

Snow's map provided another clue. Cholera was worst around the Broad Street pump, which pumped ground water. But cases of cholera occurred in other parts of London, as well. In much of London, water was supplied to homes and businesses by one of two commercial suppliers. They pumped water from the Thames—the great river that flows through London. Although the pumping of water in this manner is not unlike much of today's water supplies to many cities, the treatment plants used today were absent and water was not treated with chlorine or filtered. The two company's operations differed in a crucial detail, however. The Lambeth Company drew its water from the Thames upstream from London, while the Southwark and Vauxhall Company drew its water from local wells and from the Thames in central London. Snow found an astonishing difference in death rates from cholera when comparing households that used water from the two companies. The death rate per 10,000 households was 37 for areas supplied by the Lambeth Company. It was 315 per 10,000 for those supplied by the Southwark and Vauxhall. Water taken from wells and from the Thames in London was regularly contaminated with London's sewerage.

Armed with these facts, Snow persuaded the Board of Guardians of St. James Parish to remove the handle from the Broad Street pump. There was an immediate reduction in cholera cases, though in par that was probably due to the fact that many people had already fled the neighborhood.

Snow did not, of course, solve the problem of the cause of cholera—it would be several decades before the bacterium *Vibrio cholerae* was identified as the culprit by the great bacteriologist Robert Koch in 1883.

Today, the John Snow pub stands on Broad Street (now Broadwick Street). Nearby, a granite slab marks the location of the original Broad Street pump.

The second edition of Snow's *On the Mode of Communication of Cholera* (1855) is a deserved classic in science in general and epidemiology in particular. It is available online through the UCLA medical school web site listed in the footnote at the beginning of this section. Both editions are republished in Snow (1845, 1855/1965).

The history of epidemiology is full of great detective stories. Careful observation and the interplay of hypothesis and data, a willingness to stick with a theory when some data appeared to contradict it, a willingness to challenge received wisdom (authority), comparisons and controls, and willingness to put a theory to experimental test (by removing the pump handle) are admirably displayed in John Snow's work. In much scientific work, only a partial answer is found, and some of the causal mechanisms may remain hidden—just as the bacterial cause of cholera remained hidden in Snow's time. But even an incomplete understanding can be enormously better than none. To know that contaminated water produces cholera can prevent the disease even when the specific bacterial pathogen remains unidentified. In the rest of this book, I will have occasion here and there to draw parallels with John Snow's investigation of cholera.