MIDS DATASCI W201: Research Design and Applications for Data Analysis - Weber, Ambrose, Lee, and Rivera (Spring 2018)

Steve Weber
Brooks Ambrose
Yoonjung Lee
Michael Rivera

This page intentionally left blank

This page intentionally left blank

TABLE OF CONTENTS

Burton, R. 2009. On Being Certain. Chapters 1 and 2

This page intentionally left blank

1
The Feeling of Knowing

I AM STUCK IN AN OBLIGATORY NEIGHBORHOOD COCKTAIL party during the first week of the U.S. invasion of Iraq. A middle-aged, pin-striped lawyer announces that he'd love to be in the front lines when the troops reach Baghdad. "Door-to-door fighting," he says, puffing up his chest. He says he's certain he could shoot an Iraqi soldier, although he's never been in a conflict bigger than a schoolyard brawl.

"I don't know," I say. "I'd have trouble shooting some young kid who was being forced to fight."

"Not me. We're down to dog-eat-dog."

He nods at his frowning wife, who's anti-invasion. "All's fair in love and war." Then back to me. "You're not one of those peacenik softies, are you?"

"It wouldn't bother you to kill someone?"

"Not a bit."

"You're sure?"

"Absolutely."

He's a neighbor and I can't escape. So I tell him one of my father's favorite self-mocking stories.

During the 1930s and '40s, my father had a pharmacy in one of the tougher areas of San Francisco. He kept a small revolver hidden beneath the back cash register. One night, a man approached, pulled out a knife, and demanded all the money in the register. My father reached under the counter, grabbed his gun, and aimed it at the robber.

"Drop it," the robber said, his knife at my father's throat. "You're not going to shoot me, but I *will* kill you."

For a moment it was a Hollywood standoff, *mano a mano*. Then my father put down his gun, emptied out the register, and handed over the money.

"What's your point?" the lawyer asks. "Your father should have shot him."

"Just the obvious," I say. "You don't always know what you're going to do until you're in the moment."

"Sure you do. I know with absolute certainty that I'd shoot anyone who was threatening me."

"No chance of any hesitation?"

"None at all. I know myself. I know what I would do. End of discussion."

MY MIND REELS with seemingly impossible questions. What kind of knowledge is "I know myself and what I would do"? Is it a conscious decision based upon deep self-contemplation or is it a "gut feeling"? But what is a gut feeling—an unconscious decision, a mood or emotion, an ill-defined but clearly recognizable mental state, or a combination of all these ingredients? If we are to understand how we know what we know, we first need some ground rules, including a general classification of mental states that create our sense of knowledge about our knowledge.

For simplicity, I have chosen to lump together the closely allied feelings of certainty, rightness, conviction, and correctness under the all-inclusive term, the *feeling of knowing*. Whether or not these are separate sensations or merely shades or degrees of a common feeling isn't important. What they do share is a common quality: Each is a form of metaknowledge—knowledge about our knowledge—that qualifies or colors our thoughts, imbuing them with a sense of rightness or wrongness. When focusing on the phenomenology (how these sensations *feel*), I've chosen to use the term the *feeling of knowing* (in italics). However, when talking about the underlying science, I'll use *knowing* (in italics). Later I will expand this category to include feelings of familiarity and realness—qualities that enhance our sense of correctness.

EVERYONE IS FAMILIAR with the most commonly recognized *feeling of knowing*. When asked a question, you feel strongly that you know an answer that you cannot immediately recall. Psychologists refer to this hard-to-describe but easily recognizable feeling as a tip-of-the-tongue sensation. The frequent accompanying comment as you scan your mental Rolodex for the forgotten name or phone number: "I know it, but I just can't think of it." In this example, you are aware of knowing something, without knowing what this sense of knowing refers to.

Anyone who's been frustrated with a difficult math problem has appreciated the delicious moment of relief when an incomprehensible equation suddenly *makes sense*. We "see the light." This *aha* is a notification from a subterranean portion of our mind, an involuntary all-clear signal that we have grasped the heart of a problem. It isn't just that we can solve the problem; we also "know" that we understand it.

Most *feelings of knowing* are far less dramatic. We don't ordinarily sense them as spontaneous emotions or moods like love or happiness; rather they feel like thoughts—elements of a correct line of reasoning. We learn to add 2+2. Our teacher tells us that 4 is the correct answer. Yes, we hear a portion of our mind say. Something within us tells us that we "know" that our answer is correct. At this simplest level of understanding, there are two components to our understanding—the knowledge that 2+2=4, and the judgment or assessment of this understanding. We know that our understanding that 2+2=4 is itself correct.

The *feeling of knowing* is also commonly recognized by its absence. Most of us are all too familiar with the frustration of being able to operate a computer without having any "sense" of how the computer really works. Or learning physics despite having no "feeling" for the rightness of what you've learned. I can fix a frayed electrical cord, yet am puzzled by the very essence of electricity. I can pick up iron filings with a magnet without having the slightest sense of what magnetism "is."

At a deeper level, most of us have agonized over those sickening "crises of faith" when firmly held personal beliefs are suddenly stripped of a visceral sense of correctness, rightness, or meaning. Our most considered beliefs suddenly don't "feel right." Similarly, most of us have been shocked to hear that a close friend or relative has died unexpectedly, and yet we "feel" that he is still alive. Such upsetting news often takes time to "sink in." This disbelief associated with hearing about a death is an example of the sometimes complete disassociation between intellectual and felt knowledge.

To begin our discussion of the *feeling of knowing*, read the following excerpt at normal speed. Don't skim, give up halfway through, or skip to the explanation. Because this experience can't be duplicated once you know the explanation, take a moment to ask yourself how you feel about the paragraph. After reading the clarifying word, reread the paragraph. As you do so, please pay close attention to the shifts in your mental state and your feeling about the paragraph.

> A newspaper is better than a magazine. A seashore is a better place than the street. At first it is better to run than to walk. You may have to try several times. It takes some skill, but it is easy to learn. Even young children can enjoy it. Once successful, complications are minimal. Birds seldom get too close. Rain, however, soaks in very fast. Too many people doing the same thing can also cause problems. One needs lots of room. If there are no complications, it can be very peaceful. A rock will serve as an anchor. If things break loose from it, however, you will not get a second chance.

Is this paragraph comprehensible or meaningless? Feel your mind sort through potential explanations. Now watch what happens with the presentation of a single word: kite. As you reread the paragraph, feel the prior discomfort of something amiss shifting to a pleasing sense of rightness. Everything fits; every sentence works and has meaning. Reread the paragraph again; it is impossible to regain the sense of not understanding. In an instant, without due conscious deliberation, the paragraph has been irreversibly infused with a *feeling of knowing*.

Try to imagine other interpretations for the paragraph. Suppose

I tell you that this is a collaborative poem written by a third-grade class, or a collage of strung-together fortune cookie quotes. Your mind balks. The presence of this *feeling of knowing* makes contemplating alternatives physically difficult.

Each of us probably read the paragraph somewhat differently, but certain features seem universal. After seeing the word *kite,* we quickly go back and reread the paragraph, testing the sentences against this new piece of information. At some point, we are convinced. But when and how?

The kite paragraph raises several questions central to our understanding of how we "know" something. Though each will be discussed at greater length in subsequent chapters, here's a sneak preview.

- Did you consciously "decide" that *kite* was the correct explanation for the paragraph, or did this decision occur involuntarily, outside of conscious awareness?

- What brain mechanism(s) created the shift from not knowing to *knowing?*

- When did this shift take place? (Did you know that the explanation was correct before, during, or after you reread the paragraph?)

- After rereading the paragraph, are you able to consciously separate out the *feeling of knowing* that *kite* is the correct answer from a reasoned understanding that the answer is correct?

- Are you sure that *kite* is the correct answer? If so, how do you know?

2

How Do We Know What We Know?

PARENTS' AND TEACHERS' CUSTOMARY ADVICE FOR "NOT getting" math and physics is to study harder and think more deeply about the problem. Their assumption is that more effort will bridge the gap between dry knowledge and felt understanding. Without this assumption, we would give up every time we failed to understand something at first glance. But for those "what's the point of it all" existential moments—when formerly satisfactory feelings of purpose and meaning no longer "feel right"—history and experience have taught us differently. Logic and reason rarely are "convincing." (In this context, "convincing" is synonymous with reviving this missing "*feeling of knowing* what life is about.") Instead, we conjure up images of ascetics, mystics, and spiritual seekers—those who have donned hair shirts, trekked through the desert à la St. Jerome, huddled in caves or under trees, or sought isolation and silence in monasteries. Eastern religions emphasize a "stillness of the mind" rather than actively thinking about the missing sense of meaning.

So, which is it? Should the remedy for the absence of the *feeling of knowing* be more conscious effort and hard thought, or less? Or are both of these common teachings at odds with more basic neurobiology? Consider the curious phenomenon of *blindsight*, perhaps the best-studied example of the lack of the *feeling of knowing* in the presence of a state of knowledge.

Out of Sight Is Not Out of Mind

A patient has a stroke that selectively destroys his occipital cortex—the portion of the brain that receives primary visual inputs. His retina still records incoming information, but his malfunctioning visual cortex does not register the images sent from the retina. The result is that the patient consciously sees nothing. Now flash a light in various quadrants of his visual field. The patient reports that he sees nothing, yet he can fairly accurately localize the flashing light to the appropriate quadrant. He feels that he is guessing and is unaware that he is performing any better than by chance.

How is this possible?

First, let's trace the pathway of the "unseen" light. Some fibers from the retina proceed directly to the primary visual cortex in the occipital lobe. But other fibers bypass the region responsible for conscious "seeing" and instead project to subcortical and upper brain stem regions that do not produce a visual image. These lower brain areas are primarily concerned with automatic, reflexive functions such as fight-or-flight. Quickly approaching or looming objects cause the body to swing the head into position so that the eyes can examine the threat. An immediate reflexive action has clear evolutionary benefits over more time-consuming conscious perception and deliberation. In the broadest sense, you could say that these subcortical regions "see" the threat without sending a visual image into awareness.

Blindsight is a primitive unconscious visual localization and navigation system uncovered by the patient's cortical blindness. The patient's subliminal knowledge of the location of the flashing light doesn't trigger the *feeling of knowing* because news of this knowledge can't reach the higher cortical regions that generate the feeling. As a result, the patient swears that he hasn't seen a flashing light, yet he clearly possesses a subliminal knowledge of the light's location. When he chooses the proper visual field for the flashing light, he has no feeling that this is a correct answer. *He does not know what he knows.*[1]

With blindsight, we see the disconnect between knowledge and awareness of this knowledge as being related to a fundamental flaw in our circuitry. This broken connection cannot be restored either through conscious effort or stilling of the mind—the problem is not within our control.

Though clinically apparent blindsight is a rare event usually caused by a stroke that interferes with the blood supply to the occipital cortex, faulty expressions of the *feeling of knowing* are everyday occurrences. Let's begin with our own memories.

The *Challenger* Study

Try to remember where you were when Kennedy was assassinated, the *Challenger* blew up, or the World Trade Center was attacked. Now ask yourself how certain you are of those memories. If you believe that you are quite sure of where you were when you heard the news, keep that feeling in mind as you read about the *Challenger* study in the next pages. If you don't remember where you were, ask yourself how you know that you don't remember. (Keep in mind the blindsight example when asking this question.) Either way, try to understand the feeling and your degree of certainty of this memory.

At my most recent med school reunion dinner, several former classmates were recalling where they were when Kennedy was assassinated. We had been in the second year of medical school, which meant that we all went to the same classes. Wherever one was, we probably all were. But the recollections were strikingly different; after dinner the discussion was becoming increasingly heated, as though each classmate's mind was on trial. A urologist thought we were at lunch, an internist said we were in the lab. A pathologist remembered being at a pub down the street from the med center. "That can't be true," the urologist said. "The assassination was at noon, Dallas time. You didn't go to the bars 'til after class."

I laughed and briefly described the *Challenger* study.[2]

Within one day of the space shuttle *Challenger* explosion, Ulric Neisser, a psychologist studying "flashbulb" memories (the recall of highly dramatic events), asked his class of 106 students to write down exactly how they'd heard about the explosion, where they were, what they'd been doing, and how they felt. Two and a half years later they were again interviewed. Twenty-five percent of the students' subsequent accounts were strikingly different than their original journal entries. More than half the people had lesser degrees of error, and less than ten percent had all the details correct. (Prior to seeing their original journals, most students presumed that their memories were correct.)

Most of us reluctantly admit that memory changes over time.

As kids, we saw how a story changed with retellings around a campfire. We have been at enough family reunions to hear once-familiar shared events morphed into unrecognizable and often contradictory descriptions. So, seeing that your journal entries were different than your recollection a couple of years later shouldn't be surprising. What startled me about the *Challenger* study were the students' responses when confronted with their conflicting accounts. Many expressed a high level of confidence that their false recollections were correct, despite being confronted with their own handwritten journals. The most unnerving was one student's comment, "That's my handwriting, but that's not what happened."

Why wouldn't the students consider their journal entries written shortly after the event to be more accurate than a recollection pulled up several years later? Pride, stubbornness, or fear of admitting an error? Not remembering the details of the *Challenger* explosion doesn't imply some massive personal failing that would make resistance to contrary evidence so overwhelming. Conversely, wouldn't pride in being logical and rational steer the students toward choosing their own handwriting over memories that they know might have been altered with time?

The inflamed urologist interrupted me, insisting the pathologist concede that he was wrong. The pathologist refused, turned to me, and said, "You tell them, Burton. You were there in the bar with me."

"Beats me. I just don't remember."

"That's not possible," the two warring doctors said simultaneously. "Everyone remembers the Kennedy assassination."

I shrugged and silently marveled at the vehemence of my classmates' convictions. Even telling them of the *Challenger* study persuaded no one, as though they were intent upon reproducing the very study that questioned their recollections. All felt that they were right, that they absolutely *knew* where they were and what they were doing when Kennedy was assassinated.

Cognitive Dissonance

In 1957, Stanford professor of social psychology Leon Festinger introduced the term *cognitive dissonance* to describe the distressing mental state in which people "find themselves doing things that don't fit with what they know, or having opinions that do not fit with other opinions they hold."[3] In a series of clever experiments, Festinger demonstrated that such tensions were more often minimized or resolved through changes in personal attitudes than by relinquishing the dissonant belief or opinion.

As an example, Festinger and his associates described a cult that believed that the earth was going to be destroyed by a flood. When the flood did not happen, those less involved with the cult were more inclined to recognize that they had been wrong. The more invested members who had given up their homes and jobs to work for the cult were more likely to reinterpret the evidence to show that they were right all along, but that the earth was not destroyed because of their faithfulness.[4]

Festinger's seminal observation: The more committed we are to a belief, the harder it is to relinquish, even in the face of overwhelming contradictory evidence. Instead of acknowledging an error in judgment and abandoning the opinion, we tend to develop a new attitude or belief that will justify retaining it. By giving us a model to consider how we deal with conflicting values, the theory of cognitive dissonance has become one of the most influential theories in social psychology. Yet it fails to convincingly answer why it is so difficult to relinquish unreasonable opinions, especially in light of seemingly convincing contrary evidence. It is easy to dismiss such behavior in cult members and others "on the fringe," but what about those of us who presume ourselves to be less flaky, those of us who pride ourselves on being levelheaded and reasonable?

WE MIGHT THINK of the *Challenger* study as an oddity, but here are additional examples of consciously choosing a false belief because it *feels* correct even when we know better. I have chosen the first example as a prelude to a later discussion in chapter 13 of the deeply rooted biological component of the science-versus-religion struggle. The second example, highlighting the cognitive dissonance of the placebo effect, introduces the idea that an unjustified *feeling of knowing* can have a clear adaptive benefit.

A Scientist Contemplates Creationism

Kurt Wise, with a B.A. in geophysics from the University of Chicago, a Ph.D. in geology from Harvard, where he studied under Steven Jay Gould, and a professorship at Bryan College in Dayton, Tennessee, writes of his personal conflict between science and religion.[5]

> I had to make a decision between evolution and Scripture. Either the Scripture was true and evolution was wrong or evolution was true and I must toss out the Bible. . . . It was there that night that I accepted the Word of God and rejected all that would ever

> counter it, including evolution. With that, in great sorrow, I tossed into the fire all my dreams and hopes in science. . . . *If all the evidence in the universe turns against creationism, I would be the first to admit it, but I would still be a creationist because that is what the Word of God seems to indicate.* (Italics mine.)

A Patient Confronts the Placebo Effect

In a study of 180 people with osteoarthritis of the knee, a team of Houston surgeons headed by Bruce Moseley, M.D., found that patients who had "sham" arthroscopic surgery reported as much pain relief and improved mobility as patients who actually underwent the procedure.[6]

Mr. A, a seventy-six-year-old retired World War II veteran with a five-year history of disabling knee pain from X-ray-documented degenerative osteoarthritis was assigned to the placebo group (sham surgery in which general anesthesia was given, superficial incisions were made in the skin over the knee, but no actual surgical repair was performed). After the procedure, Mr. A was informed that he had received sham surgery; the procedure was described in detail. Nevertheless, he dramatically improved; for the first time in years he was able to walk without a cane. When questioned, he both fully understood what sham surgery meant and fully believed that his knee had been fixed.

"The surgery was two years ago and the knee has never bothered me since. It's just like my other knee now. I give a whole lot of credit to Dr. Moseley. Whenever I see him on the TV, I call the wife in and say, 'Hey, there's the doctor that fixed my knee!' "[7]

Our creationist geologist cringes at his own irrationality and yet declares that he does not have a choice. A patient "knows" that he hasn't had any reparative surgery performed, yet insists that the doctor fixed his knee. What if we could find patients who developed similar difficulties with reason as the result of specific brain insults (lesions)? If brain malfunctions can produce a similar flawed logic, what might that tell us about the biological underpinnings of cognitive dissonances?

Cotard's Syndrome

Ms. B, a twenty-nine-year-old grad student hospitalized for an acute viral encephalitis (a viral inflammation of the brain) complained: "Nothing feels real. I am dead." The patient refused any medical care. "There is no point in treating a dead person," she insisted. Her internist tried to reason with her. He asked her to put her hand on her chest and feel her heart beating. She did, and agreed that her heart was beating. He suggested that the presence of a pulse must mean that she was not dead. The patient countered that, since she was dead, her beating heart could not be evidence for being alive. She said she recognized that there was a logical inconsistency between being dead and being able to feel her beating heart, but that being dead felt more "real" than any contrary evidence that she was alive.

Weeks later, Ms. B began to recover; eventually she no longer believed that she was dead. She was able to make a distinction between her recovered "reality" and her prior delusions, yet she continued to believe that it must be possible to feel one's heart beat after death. After all, it had happened to her.

Cotard's syndrome—*le délire de négation*—is attributed to a French psychiatrist, Jules Cotard, who in 1882 described several patients with delusions of self-negation. These ranged from the belief that parts of the body were missing, or had putrefied, to the complete denial of bodily existence. The syndrome has been described with a variety of brain injuries, strokes, and dementia, as well as severe psychiatric disorders. The most extraordinary element of the syndrome is the patient's unshakable belief in being dead that overpowers any logical counterconclusion. Feeling one's beating heart isn't sufficient evidence to overcome the more powerful sense of the reality of being dead.

Other delusional syndromes associated with acute brain lesions include believing that a friend or a relative has been replaced by an impostor, or a double, or has taken on different appearances or identities, or that an inanimate object has been replaced by an inferior copy. The clinical feature common to all of these syndromes is the inability of the patient to shake a belief that he logically knows is wrong.

Mr. C, an elegant retired art dealer, was hospitalized overnight with a small stroke. The next morning, he felt well and was discharged. Within moments of returning home, he phoned my office in a panic. He was certain that his favorite antique desk had been replaced by a cheap Levitz reproduction. "Hurry over and see for yourself." He lived near my office; I dropped by at lunchtime. The desk in question was a massive eighteenth-century Italian refectory table that took up most of his den. It could easily seat a dozen; just lifting it would require a minimum of several men. And it was far too wide to fit through the doorway without removing the French doors. I quickly pointed out the impossibility of someone sneaking in, moving out the desk, and substituting a fake. Mr. C shook his head. "Yes, I admit that it is physically impossible that the desk has been replaced. But it has. You have to take my word for it. I know real when I see real, and this desk isn't real." He ran his hand along the grain, repeatedly fingering a couple of prominent wormholes. "It's funny," he said with a puzzled expression. "These are exact replicas of the holes in my desk. But they don't feel the least bit familiar. No," he announced emphatically, "someone must have replaced it." He then delivered the cognitive checkmate: "After all, I know what I know."

Although not restricted to a single area of the brain or a single definitive physiology, the most striking shared characteristic of these *delusional misidentification syndromes* is that the conflict between logic and a contrary *feeling of knowing* tends to be resolved in favor of feeling. Rather than rejecting ideas and beliefs that defy common sense and overwhelming contrary evidence, such patients end up using tortured logic to justify the more powerful sense of *knowing what they know.*[8]

Mr. C's statements also point out that *knowing* may also involve additional hard-to-define mental states such as a sense of familiarity and feelings of "realness." Like the tip-of-the-tongue sensation or the feeling of déjà vu, a sense of being familiar suggests some prior experience or knowledge. When stumped on a multiple choice test question, we tend to choose the answer that feels most familiar. Though we have no justification, we presume that such answers are more likely to be correct than those that we don't recognize or seem unfamiliar. Mr. C's "I know real when I see real" points out how a sense of "realness" might also bias us toward believing that an idea is correct. Patients with delusional misidentification syndromes often use "correct" interchangeably with "real."

It is likely that Mr. C's stroke affected his ability to appropriately experience feelings of familiarity and "realness." When neither the sight nor the feel of the desk triggered these feelings, he was forced to conclude that this desk could not be the original. Such delusions might be seen as an attempt to resolve a cognitive dissonance between hard evidence (the table is too big to move) and the absence of any feeling of familiarity and realness when Mr. C examined his desk.

In chapter 3 we shall see that the mental states of familiarity, "realness," conviction, truth, déjà vu, and tip of the tongue share a similar physiology with the *feeling of knowing*, including the ability to be directly triggered with electrical stimulation of the brain's limbic system.

It May Be Right, But It's Not *Right*

The other day, at a downtown garage, I left my car with valet parking. I returned, started to drive away, but felt something was wrong. I questioned the attendant's gaze, wondering if I'd paid too much. I checked the gas and oil gauge, and whether one of the doors was ajar. Then I realized that the seat had been readjusted by the attendant. It was a nominal difference, the seat was at most a half inch higher than usual. My derriere knew immediately; it took me considerably longer.

I was reminded of a story attributed to Ludwig Wittgenstein.

A man walks into a tailor's shop. The sign over the front door reads: CUSTOMER SATISFACTION GUARANTEED. The man orders a custom-made suit that should fit exactly like the one he is wearing. The tailor painstakingly measures every detail and jots them down in a notebook. A week later the customer returns to try on the new suit.

"It's not right," the customer says with annoyance. . . .

"Of course it is," the tailor says. "Here, I'll show you." The tailor takes out his measuring tape, compares the suit's readings with those in his notebook. "See, they're identical."

The customer shifts in his new suit but is still uncomfortable and displeased. "It may be right, but it's not *right*." He refuses to pay for the suit and storms out.

In the case of my car seat, I was forced to think through all the possible reasons that I sensed something was wrong. Fortunately, there was something measurable (the new angle of the car seat) that *explained* what I was *feeling*. With the tailor example, the customer's sense of something amiss is a matter of taste, of inexpressible or subconscious aesthetics. No matter what the measurements, the suit does not *feel* right.

The tailor demands his money; the customer admits that the suit was to his specifications, but not to his liking, and therefore he is under no obligation to buy the suit. Each *feels* that he is right. Hence that irritating popular refrain—end of discussion. We often talk about gut feelings. There is now extensive literature on the neuroenteric brain, as though some form of thought might actually originate in the pit of your stomach. Maybe so. And maybe my body just *knew* that my car seat was out of whack. But whatever the origin of the sensation, the key feature is that there seems to be an underlying *sense* or *feeling* that something is either correct or incorrect.

Consider the similarity in tone between the *Challenger* study student who said, "That's my writing, but that's not what happened," and the suit customer's "It might be right, but it's not *right*." When such a sense of conviction overrides obvious logical inconsistencies or scientific evidence, what is happening? Is it possible that there is an underlying neurophysiological basis for the specific sensation of *feeling right* or of *being right* that is so powerful that ordinary rational thought *feels* either wrong or irrelevant? Conviction versus knowledge—is the jury rigged, the game fixed by a basic physiology hidden beneath awareness?

Creswell, J. W. 2009. Research Design: Qualitative, Quantitative, and Mixed Methods.Ch 1: The Selection of a Research Design

This page intentionally left blank

CHAPTER ONE

The Selection of a Research Design

Research designs are plans and the procedures for research that span the decisions from broad assumptions to detailed methods of data collection and analysis. This plan involves several decisions, and they need not be taken in the order in which they make sense to me and the order of their presentation here. The overall decision involves which design should be used to study a topic. Informing this decision should be the worldview assumptions the researcher brings to the study; procedures of inquiry (called strategies); and specific methods of data collection, analysis, and interpretation. The selection of a research design is also based on the nature of the research problem or issue being addressed, the researchers' personal experiences, and the audiences for the study.

THE THREE TYPES OF DESIGNS

In this book, three types of designs are advanced: qualitative, quantitative, and mixed methods. Unquestionably, the three approaches are not as discrete as they first appear. Qualitative and quantitative approaches should not be viewed as polar opposites or dichotomies; instead, they represent different ends on a continuum (Newman & Benz, 1998). A study *tends* to be more qualitative than quantitative or vice versa. Mixed methods research resides in the middle of this continuum because it incorporates elements of both qualitative and quantitative approaches.

Often the distinction between qualitative and quantitative research is framed in terms of using words (qualitative) rather than numbers (quantitative), or using closed-ended questions (quantitative hypotheses) rather than open-ended questions (qualitative interview questions). A more complete way to view the gradations of differences between them is in the basic philosophical assumptions researchers bring to the study, the types of research strategies used overall in the research (e.g., quantitative experiments or qualitative case studies), and the specific methods employed in conducting these strategies (e.g., collecting data quantitatively on instruments versus collecting qualitative data through observing a setting). Moreover, there is a historical evolution to both approaches, with the quantitative approaches dominating the forms of research in the social sciences from the late 19th century up until the mid-20th century. During the latter half of the 20th century, interest in qualitative research increased and along with it, the development of mixed methods research (see Creswell, 2008, for more of this history). With this background, it should prove helpful to view definitions of these three key terms as used in this book:

- **Qualitative research** is a means for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data. The final written report has a flexible structure. Those who engage in this form of inquiry support a way of looking at research that honors an inductive style, a focus on individual meaning, and the importance of rendering the complexity of a situation (adapted from Creswell, 2007).

- **Quantitative research** is a means for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures. The final written report has a set structure consisting of introduction, literature and theory, methods, results, and discussion (Creswell, 2008). Like qualitative researchers, those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings.

- **Mixed methods research** is an approach to inquiry that combines or associates both qualitative and quantitative forms. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study. Thus, it is more than simply collecting and analyzing both kinds of data; it also involves the use of both approaches in tandem so that the overall strength of a study is greater than either qualitative or quantitative research (Creswell & Plano Clark, 2007).

These definitions have considerable information in each one of them. Throughout this book, I discuss the parts of the definitions so that their meanings become clear to you.

THREE COMPONENTS INVOLVED IN A DESIGN

Two important components in each definition are that the approach to research involves philosophical assumptions as well as distinct methods or procedures. **Research design**, which I refer to as the *plan or proposal to conduct research*, involves the intersection of philosophy, strategies of inquiry, and specific methods. A framework that I use to explain the interaction of these three components is seen in Figure 1.1. To reiterate, in planning a study, researchers need to think through the philosophical worldview assumptions that they bring to the study, the strategy of inquiry that is related to this worldview, and the specific methods or procedures of research that translate the approach into practice.

Philosophical Worldviews

Although philosophical ideas remain largely hidden in research (Slife & Williams, 1995), they still influence the practice of research and need to be identified. I suggest that individuals preparing a research proposal or plan make explicit the larger philosophical ideas they espouse. This information will help explain why they chose qualitative, quantitative, or mixed methods

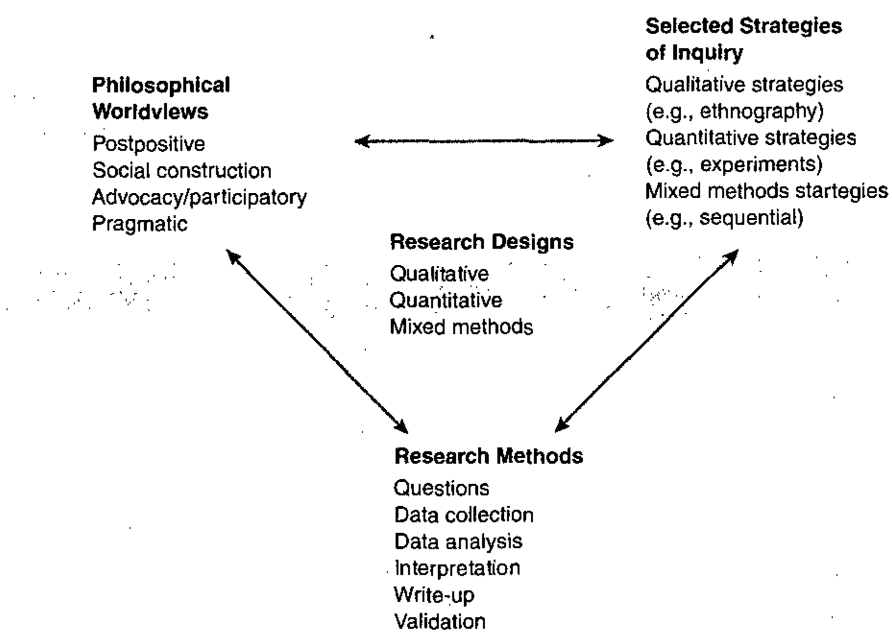


Figure 1.1 A Framework for Design—The Interconnection of Worldviews, Strategies of Inquiry, and Research Methods

approaches for their research. In writing about worldviews, a proposal might include a section that addresses the following:

- The philosophical worldview proposed in the study
- A definition of basic considerations of that worldview
- How the worldview shaped their approach to research

I have chosen to use the term **worldview** as meaning "a basic set of beliefs that guide action" (Guba, 1990, p. 17). Others have called them *paradigms* (Lincoln & Guba, 2000; Mertens, 1998); *epistemologies* and *ontologies* (Crotty, 1998), or *broadly conceived research methodologies* (Neuman, 2000). I see worldviews as a general orientation about the world and the nature of research that a researcher holds. These worldviews are shaped by the discipline area of the student, the beliefs of advisers and faculty in a student's area, and past research experiences. The types of beliefs held by individual researchers will often lead to embracing a qualitative, quantitative, or mixed methods approach in their research. Four different worldviews are discussed: postpositivism, constructivism, advocacy/participatory, and pragmatism. The major elements of each position are presented in Table 1.1.

The Postpositivist Worldview

The postpositivist assumptions have represented the traditional form of research, and these assumptions hold true more for quantitative research than qualitative research. This worldview is sometimes called the *scientific method* or doing *science research*. It is also called *positivist/postpositivist research, empirical science,* and *postpostivism.* This last term is called postpositivism because it represents the thinking after positivism, challenging

Table 1.1 Four Worldviews

Postpositivism	Constructivism
• Determination • Reductionism • Empirical observation and measurement • Theory verification	• Understanding • Multiple participant meanings • Social and historical construction • Theory generation
Advocacy/Participatory	**Pragmatism**
• Political • Empowerment Issue-oriented • Collaborative • Change-oriented	• Consequences of actions • Problem-centered • Pluralistic • Real-world practice oriented

the traditional notion of the absolute truth of knowledge (Phillips & Burbules, 2000) and recognizing that we cannot be "positive" about our claims of knowledge when studying the behavior and actions of humans. The postpositivist tradition comes from 19th-century writers, such as Comte, Mill, Durkheim, Newton, and Locke (Smith, 1983), and it has been most recently articulated by writers such as Phillips and Burbules (2000).

Postpositivists hold a deterministic philosophy in which causes probably determine effects or outcomes. Thus, the problems studied by postpositivists reflect the need to identify and assess the causes that influence outcomes, such as found in experiments. It is also reductionistic in that the intent is to reduce the ideas into a small, discrete set of ideas to test, such as the variables that comprise hypotheses and research questions. The knowledge that develops through a postpositivist lens is based on careful observation and measurement of the objective reality that exists "out there" in the world. Thus, developing numeric measures of observations and studying the behavior of individuals becomes paramount for a postpositivist. Finally, there are laws or theories that govern the world, and these need to be tested or verified and refined so that we can understand the world. Thus, in the scientific method, the accepted approach to research by postpostivists, an individual begins with a theory, collects data that either supports or refutes the theory, and then makes necessary revisions before additional tests are made.

In reading Phillips and Burbules (2000), you can gain a sense of the key assumptions of this position, such as,

1. Knowledge is conjectural (and antifoundational)—absolute truth can never be found. Thus, evidence established in research is always imperfect and fallible. It is for this reason that researchers state that they do not prove a hypothesis; instead, they indicate a failure to reject the hypothesis.

2. Research is the process of making claims and then refining or abandoning some of them for other claims more strongly warranted. Most quantitative research, for example, starts with the test of a theory.

3. Data, evidence, and rational considerations shape knowledge. In practice, the researcher collects information on instruments based on measures completed by the participants or by observations recorded by the researcher.

4. Research seeks to develop relevant, true statements, ones that can serve to explain the situation of concern or that describe the causal relationships of interest. In quantitative studies, researchers advance the relationship among variables and pose this in terms of questions or hypotheses.

5. Being objective is an essential aspect of competent inquiry; researchers must examine methods and conclusions for bias. For example, standard of validity and reliability are important in quantitative research.

The Social Constructivist Worldview

Others hold a different worldview. Social constructivism (often combined with interpretivism; see Mertens, 1998) is such a perspective, and it is typically seen as an approach to qualitative research. The ideas came from Mannheim and from works such as Berger and Luekmann's (1967) *The Social Construction of Reality* and Lincoln and Guba's (1985) *Naturalistic Inquiry*. More recent writers who have summarized this position are Lincoln and Guba (2000), Schwandt (2007), Neuman (2000), and Crotty (1998), among others. **Social constructivists** hold assumptions that individuals seek understanding of the world in which they live and work. Individuals develop subjective meanings of their experiences—meanings directed toward certain objects or things. These meanings are varied and multiple, leading the researcher to look for the complexity of views rather than narrowing meanings into a few categories or ideas. The goal of the research is to rely as much as possible on the participants' views of the situation being studied. The questions become broad and general so that the participants can construct the meaning of a situation, typically forged in discussions or interactions with other persons. The more open-ended the questioning, the better, as the researcher listens carefully to what people say or do in their life settings. Often these subjective meanings are negotiated socially and historically. They are not simply imprinted on individuals but are formed through interaction with others (hence social constructivism) and through historical and cultural norms that operate in individuals' lives. Thus, constructivist researchers often address the processes of interaction among individuals. They also focus on the specific contexts in which people live and work, in order to understand the historical and cultural settings of the participants. Researchers recognize that their own backgrounds shape their interpretation, and they position themselves in the research to acknowledge how their interpretation flows from their personal, cultural, and historical experiences. The researcher's intent is to make sense of (or interpret) the meanings others have about the world. Rather than starting with a theory (as in postpostivism), inquirers generate or inductively develop a theory or pattern of meaning.

For example, in discussing constructivism, Crotty (1998) identified several assumptions:

1. Meanings are constructed by human beings as they engage with the world they are interpreting. Qualitative researchers tend to use open-ended questions so that the participants can share their views.

2. Humans engage with their world and make sense of it based on their historical and social perspectives—we are all born into a world of meaning bestowed upon us by our culture. Thus, qualitative researchers seek to understand the context or setting of the participants through visiting this context and gathering information personally. They also interpret what

they find, an interpretation shaped by the researcher's own experiences and background.

3. The basic generation of meaning is always social, arising in and out of interaction with a human community. The process of qualitative research is largely inductive, with the inquirer generating meaning from the data collected in the field.

The Advocacy and Participatory Worldview

Another group of researchers holds to the philosophical assumptions of the advocacy/participatory approach. This position arose during the 1980s and 1990s from individuals who felt that the postpostivist assumptions imposed structural laws and theories that did not fit marginalized individuals in our society or issues of social justice that needed to be addressed. This worldview is typically seen with qualitative research, but it can be a foundation for quantitative research as well. Historically, some of the advocacy/participatory (or emancipatory) writers have drawn on the works of Marx, Adorno, Marcuse, Habermas, and Freire (Neuman, 2000). Fay (1987), Heron and Reason (1997), and Kemmis and Wilkinson (1998) are more recent writers to read for this perspective. In the main, these inquirers felt that the constructivist stance did not go far enough in advocating for an action agenda to help marginalized peoples. An **advocacy/participatory worldview** holds that research inquiry needs to be intertwined with politics and a political agenda. Thus, the research contains an action agenda for reform that may change the lives of the participants, the institutions in which individuals work or live, and the researcher's life. Moreover, specific issues need to be addressed that speak to important social issues of the day, issues such as empowerment, inequality, oppression, domination, suppression, and alienation. The researcher often begins with one of these issues as the focal point of the study. This research also assumes that the inquirer will proceed collaboratively so as to not further marginalize the participants as a result of the inquiry. In this sense, the participants may help design questions, collect data, analyze information, or reap the rewards of the research. Advocacy research provides a voice for these participants, raising their consciousness or advancing an agenda for change to improve their lives. It becomes a united voice for reform and change.

This philosophical worldview focuses on the needs of groups and individuals in our society that may be marginalized or disenfranchised. Therefore, theoretical perspectives may be integrated with the philosophical assumptions that construct a picture of the issues being examined, the people to be studied, and the changes that are needed, such as feminist perspectives, racialized discourses, critical theory, queer theory, and disability theory—theoretical lens to be discussed more in Chapter 3.

Although these are diverse groups and my explanations here are generalizations, it is helpful to view the summary by Kemmis and Wilkinson (1998) of key features of the advocacy or participatory forms of inquiry:

1. Participatory action is recursive or dialectical and focused on bringing about change in practices. Thus, at the end of advocacy/participatory studies, researchers advance an action agenda for change.

2. This form of inquiry is focused on helping individuals free themselves from constraints found in the media, in language, in work procedures, and in the relationships of power in educational settings. Advocacy/participatory studies often begin with an important issue or stance about the problems in society, such as the need for empowerment.

3. It is emancipatory in that it helps unshackle people from the constraints of irrational and unjust structures that limit self-development and self-determination. The advocacy/participatory studies aim to create a political debate and discussion so that change will occur.

4. It is practical and collaborative because it is inquiry completed *with* others rather than *on* or *to* others. In this spirit, advocacy/participatory authors engage the participants as active collaborators in their inquiries.

The Pragmatic Worldview

Another position about worldviews comes from the pragmatists. Pragmatism derives from the work of Peirce, James, Mead, and Dewey (Cherryholmes, 1992). Recent writers include Rorty (1990), Murphy (1990), Patton (1990), and Cherryholmes (1992). There are many forms of this philosophy, but for many, **pragmatism** as a worldview arises out of actions, situations, and consequences rather than antecedent conditions (as in postpositivism). There is a concern with applications—what works—and solutions to problems (Patton, 1990). Instead of focusing on methods, researchers emphasize the research problem and use all approaches available to understand the problem (see Rossman & Wilson, 1985). As a philosophical underpinning for mixed methods studies, Tashakkori and Teddlie (1998), Morgan (2007), and Patton (1990) convey its importance for focusing attention on the research problem in social science research and then using pluralistic approaches to derive knowledge about the problem. Using Cherryholmes (1992), Morgan (2007), and my own views, pragmatism provides a philosophical basis for research:

- Pragmatism is not committed to any one system of philosophy and reality. This applies to mixed methods research in that inquirers draw liberally from both quantitative and qualitative assumptions when they engage in their research.

- Individual researchers have a freedom of choice. In this way, researchers are free to choose the methods, techniques, and procedures of research that best meet their needs and purposes.

- Pragmatists do not see the world as an absolute unity. In a similar way, mixed methods researchers look to many approaches for collecting and analyzing data rather than subscribing to only one way (e.g., quantitative or qualitative).

- Truth is what works at the time. It is not based in a duality between reality independent of the mind or within the mind. Thus, in mixed methods research, investigators use both quantitative and qualitative data because they work to provide the best understanding of a research problem.

- The pragmatist researchers look to the *what* and *how* to research, based on the intended consequences—where they want to go with it. Mixed methods researchers need to establish a purpose for their mixing, a rationale for the reasons why quantitative and qualitative data need to be mixed in the first place.

- Pragmatists agree that research always occurs in social, historical, political, and other contexts. In this way, mixed methods studies may include a postmodern turn, a theoretical lens that is reflective of social justice and political aims.

- Pragmatists have believed in an external world independent of the mind as well as that lodged in the mind. But they believe that we need to stop asking questions about reality and the laws of nature (Cherryholmes, 1992). "They would simply like to change the subject" (Rorty, 1983, p. xiv).

- Thus, for the mixed methods researcher, pragmatism opens the door to multiple methods, different worldviews, and different assumptions, as well as different forms of data collection and analysis.

Strategies of Inquiry

The researcher not only selects a qualitative, quantitative, or mixed methods study to conduct, the inquirer also decides on a type of study within these three choices. **Strategies of inquiry** are types of qualitative, quantitative, and mixed methods designs or models that provide specific direction for procedures in a research design. Others have called them *approaches to inquiry* (Creswell, 2007) or *research methodologies* (Mertens, 1998). The strategies available to the researcher have grown over the years as computer technology has pushed forward our data analysis and ability to analyze complex models and as individuals have articulated new procedures for conducting social science research. Select types will be emphasized in Chapters 8, 9, and 10, strategies frequently used in the social sciences. Here I introduce those that are discussed later and that are cited in examples throughout the book. An overview of these strategies is shown in Table 1.2.

Table 1.2 Alternative Strategies of Inquiry

Quantitative	Qualitative	Mixed Methods
• Experimental designs • Non-experimental designs, such as surveys	• Narrative research • Phenomenology • Ethnographies • Grounded theory studies • Case study	• Sequential • Concurrent • Transformative

Quantitative Strategies

During the late 19th and throughout the 20th century, strategies of inquiry associated with quantitative research were those that invoked the postpositivist worldview. These include true experiments and the less rigorous experiments called *quasi-experiments* and *correlational studies* (Campbell & Stanley, 1963) and specific single-subject experiments (Cooper, Heron, & Heward, 1987; Neuman & McCormick, 1995). More recently, quantitative strategies have involved complex experiments with many variables and treatments (e.g., factorial designs and repeated measure designs). They have also included elaborate structural equation models that incorporate causal paths and the identification of the collective strength of multiple variables. In this book, I focus on two strategies of inquiry: surveys and experiments.

- **Survey research** provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population. It includes cross-sectional and longitudinal studies using questionnaires or structured interviews for data collection, with the intent of generalizing from a sample to a population (Babbie, 1990).

- **Experimental research** seeks to determine if a specific treatment influences an outcome. This impact is assessed by providing a specific treatment to one group and withholding it from another and then determining how both groups scored on an outcome. Experiments include true experiments, with the random assignment of subjects to treatment conditions, and quasi-experiments that use nonrandomized designs (Keppel, 1991). Included within quasi-experiments are single-subject designs.

Qualitative Strategies

In qualitative research, the numbers and types of approaches have also become more clearly visible during the 1990s and into the 21st century. Books have summarized the various types (such as the 19 strategies identified by Wolcott, 2001), and complete procedures are now available on specific qualitative inquiry approaches. For example, Clandinin and Connelly (2000) constructed a picture of what narrative researchers do.

Moustakas (1994) discussed the philosophical tenets and the procedures of the phenomenological method, and Strauss and Corbin (1990, 1998) identified the procedures of grounded theory. Wolcott (1999) summarized ethnographic procedures, and Stake (1995) suggested processes involved in case study research. In this book, illustrations are drawn from the following strategies, recognizing that approaches such as participatory action research (Kemmis & Wilkinson, 1998), discourse analysis (Cheek, 2004), and others not mentioned (see Creswell, 2007b) are also viable ways to conduct qualitative studies:

- **Ethnography** is a strategy of inquiry in which the researcher studies an intact cultural group in a natural setting over a prolonged period of time by collecting, primarily, observational and interview data (Creswell, 2007b). The research process is flexible and typically evolves contextually in response to the lived realities encountered in the field setting (LeCompte & Schensul, 1999).

- **Grounded theory** is a strategy of inquiry in which the researcher derives a general, abstract theory of a process, action, or interaction grounded in the views of participants. This process involves using multiple stages of data collection and the refinement and interrelationship of categories of information (Charmaz, 2006; Strauss and Corbin, 1990, 1998). Two primary characteristics of this design are the constant comparison of data with emerging categories and theoretical sampling of different groups to maximize the similarities and the differences of information.

- **Case studies** are a strategy of inquiry in which the researcher explores in depth a program, event, activity, process, or one or more individuals. Cases are bounded by time and activity, and researchers collect detailed information using a variety of data collection procedures over a sustained period of time (Stake, 1995).

- **Phenomenological research** is a strategy of inquiry in which the researcher identifies the essence of human experiences about a phenomenon as described by participants. Understanding the lived experiences marks phenomenology as a philosophy as well as a method, and the procedure involves studying a small number of subjects through extensive and prolonged engagement to develop patterns and relationships of meaning (Moustakas, 1994). In this process, the researcher brackets or sets aside his or her own experiences in order to understand those of the participants in the study (Nieswiadomy, 1993).

- **Narrative research** is a strategy of inquiry in which the researcher studies the lives of individuals and asks one or more individuals to provide stories about their lives. This information is then often retold or restoried by the researcher into a narrative chronology. In the end, the narrative combines views from the participant's life with those of the researcher's life in a collaborative narrative (Clandinin & Connelly, 2000).

Mixed Methods Strategies

Mixed methods strategies are less well known than either the quantitative or qualitative approaches. The concept of mixing different methods originated in 1959 when Campbell and Fisk used multimethods to study validity of psychological traits. They encouraged others to employ their multimethod matrix to examine multiple approaches to data collection. This prompted others to mix methods, and soon approaches associated with field methods, such as observations and interviews (qualitative data), were combined with traditional surveys (quantitative data; Sieber, 1973). Recognizing that all methods have limitations, researchers felt that biases inherent in any single method could neutralize or cancel the biases of other methods. Triangulating data sources—a means for seeking convergence across qualitative and quantitative methods—was born (Jick, 1979). By the early 1990s, the idea of mixing moved from seeking convergence to actually integrating or connecting the quantitative and qualitative data. For example, the results from one method can help identify participants to study or questions to ask for the other method (Tashakkori & Teddlie, 1998). Alternatively, the qualitative and quantitative data can be merged into one large database or the results used side by side to reinforce each other (e.g., qualitative quotes support statistical results; Creswell & Plano Clark, 2007). Or the methods can serve a larger, transformative purpose to advocate for marginalized groups, such as women, ethnic/racial minorities, members of gay and lesbian communities, people with disabilities, and those who are poor (Mertens, 2003).

These reasons for mixing methods have led writers from around the world to develop procedures for mixed methods strategies of inquiry, and these take the numerous terms found in the literature, such as *multimethod, convergence, integrated,* and *combined* (Creswell & Plano Clark, 2007), and shape procedures for research (Tashakkori & Teddlie, 2003).

In particular, three general strategies and several variations within them are illustrated in this book:

- **Sequential mixed methods** procedures are those in which the researcher seeks to elaborate on or expand on the findings of one method with another method. This may involve beginning with a qualitative interview for exploratory purposes and following up with a quantitative, survey method with a large sample so that the researcher can generalize results to a population. Alternatively, the study may begin with a quantitative method in which a theory or concept is tested, followed by a qualitative method involving detailed exploration with a few cases or individuals.

- **Concurrent mixed methods** procedures are those in which the researcher converges or merges quantitative and qualitative data in order to provide a comprehensive analysis of the research problem. In this design, the investigator collects both forms of data at the same time and then integrates the information in the interpretation of the overall results. Also, in this design, the researcher may embed one smaller form of data within another larger data collection in order to analyze different types of questions (the qualitative addresses the process while the quantitative, the outcomes).

- **Transformative mixed methods** procedures are those in which the researcher uses a theoretical lens (see Chapter 3) as an overarching perspective within a design that contains both quantitative and qualitative data. This lens provides a framework for topics of interest, methods for collecting data, and outcomes or changes anticipated by the study. Within this lens could be a data collection method that involves a sequential or a concurrent approach.

Research Methods

The third major element in the framework is the specific **research methods** that involve the forms of data collection, analysis, and interpretation that researchers propose for their studies. As shown in Table 1.3, it is useful to consider the full range of possibilities of data collection and to organize these methods, for example, by their degree of predetermined nature, their use of closed-ended versus open-ended questioning, and their focus on numeric versus nonnumeric data analysis. These methods will be developed further in Chapters 8 through 10.

Researchers collect data on an instrument or test (e.g., a set of questions about attitudes toward self-esteem) or gather information on a behavioral checklist (e.g., observation of a worker engaged in a complex skill). On the other end of the continuum, collecting data might involve visiting a research site and observing the behavior of individuals without predetermined questions or conducting an interview in which the individual is allowed to talk openly about a topic, largely without the use of specific

Table 1.3 *Quantitative, Mixed, and Qualitative Methods*

Quantitative Methods →	Mixed Methods	← Qualitative Methods
• Pre-determined • Instrument based questions • Performance data, attitude data, observational data, and census data • Statistical analysis • Statistical interpretation	• Both pre-determined and emerging methods • Both open- and closed-ended questions • Multiple forms of data drawing on all possibilities • Statistical and text analysis • Across databases interpretation	• Emerging methods • Open-ended questions • Interview data, observation data, document data, and audio-visual data • Text and image analysis • Themes, patterns interpretation

questions. The choice of methods turns on whether the intent is to specify the type of information to be collected in advance of the study or allow it to emerge from participants in the project. Also, the type of data analyzed may be numeric information gathered on scales of instruments or text information recording and reporting the voice of the participants. Researchers make interpretations of the statistical results, or they interpret the themes or patterns that emerge from the data. In some forms of research, both quantitative and qualitative data are collected, analyzed, and interpreted. Instrument data may be augmented with open-ended observations, or census data may be followed by in-depth exploratory interviews. In this case of mixing methods, the researcher makes inferences across both the quantitative and qualitative databases.

RESEARCH DESIGNS AS WORLDVIEWS, STRATEGIES, AND METHODS

The worldviews, the strategies, and the methods all contribute to a research design that *tends* to be quantitative, qualitative, or mixed. Table 1.4 creates distinctions that may be useful in choosing an approach. This table also includes practices of all three approaches that are emphasized in remaining chapters of this book.

Typical scenarios of research can illustrate how these three elements combine into a research design.

- *Quantitative* approach—Postpositivist worldview, experimental strategy of inquiry, and pre- and post-test measures of attitudes

In this scenario, the researcher tests a theory by specifying narrow hypotheses and the collection of data to support or refute the hypotheses. An experimental design is used in which attitudes are assessed both before and after an experimental treatment. The data are collected on an instrument that measures attitudes, and the information is analyzed using statistical procedures and hypothesis testing.

- *Qualitative* approach—Constructivist worldview, ethnographic design, and observation of behavior

In this situation, the researcher seeks to establish the meaning of a phenomenon from the views of participants. This means identifying a culture-sharing group and studying how it develops shared patterns of behavior over time (i.e., ethnography). One of the key elements of collecting data in this way is to observe participants' behaviors by engaging in their activities.

- *Qualitative* approach—Participatory worldview, narrative design, and open-ended interviewing

For this study, the inquirer seeks to examine an issue related to oppression of individuals. To study this, stories are collected of individual oppression

Table 1.4 Qualitative, Quantitative, and Mixed Methods Approaches

Tend to or Typically . . .	Qualitative Approaches	Quantitative Approaches	Mixed Methods Approaches
• Use these philosophical assumptions	• Constructivist/ advocacy/ participatory knowledge claims	• Post-positivist knowledge claims	• Pragmatic knowledge claims
• Employ these strategies of inquiry	• Phenomenology, grounded theory, ethnography, case study, and narrative	• Surveys and experiments	• Sequential, concurrent, and transformative
• Employ these methods	• Open-ended questions, emerging approaches, text or image data	• Closed-ended questions, predetermined approaches, numeric data	• Both open- and closed-ended questions, both emerging and predetermined approaches, and both quantitative and qualitative data and analysis
• Use these practices of research as the researcher	• Positions him- or herself • Collects participant meanings • Focuses on a single concept or phenomenon • Brings personal values into the study • Studies the context or setting of participants • Validates the accuracy of findings • Makes interpretations of the data • Creates an agenda for change or reform • Collaborates with the participants	• Tests or verifies theories or explanations • Identifies variables to study • Relates variables in questions or hypotheses • Uses standards of validity and reliability • Observes and measures information numerically • Uses unbiased approaches • Employs statistical procedures	• Collects both quantitative and qualitative data • Develops a rationale for mixing • Integrates the data at different stages of inquiry • Presents visual pictures of the procedures in the study • Employs the practices of both qualitative and quantitative research

using a narrative approach. Individuals are interviewed at some length to determine how they have personally experienced oppression.

- *Mixed methods* approach—Pragmatic worldview, collection of both quantitative and qualitative data sequentially

The researcher bases the inquiry on the assumption that collecting diverse types of data best provides an understanding of a research problem. The study begins with a broad survey in order to generalize results to a population and then, in a second phase, focuses on qualitative, open-ended interviews to collect detailed views from participants.

CRITERIA FOR SELECTING A RESEARCH DESIGN

Given the possibility of qualitative, quantitative, or mixed methods approaches, what factors affect a choice of one approach over another for the design of a proposal? Added to worldview, strategy, and methods would be the research problem, the personal experiences of the researcher, and the audience(s) for whom the report will be written.

The Research Problem

A research problem, more thoroughly discussed in Chapter 5, is an issue or concern that needs to be addressed (e.g., the issue of racial discrimination). Certain types of social research problems call for specific approaches. For example, if the problem calls for (a) the identification of factors that influence an outcome, (b) the utility of an intervention, or (c) understanding the best predictors of outcomes, then a quantitative approach is best. It is also the best approach to use to test a theory or explanation.

On the other hand, if a concept or phenomenon needs to be understood because little research has been done on it, then it merits a qualitative approach. Qualitative research is exploratory and is useful when the researcher does not know the important variables to examine. This type of approach may be needed because the topic is new, the topic has never been addressed with a certain sample or group of people, and existing theories do not apply with the particular sample or group under study (Morse, 1991).

A mixed methods design is useful when either the quantitative or qualitative approach by itself is inadequate to best understand a research problem or the strengths of both quantitative and qualitative research can provide the best understanding. For example, a researcher may want to both generalize the findings to a population as well as develop a detailed view of the meaning of a phenomenon or concept for individuals. In this research, the inquirer first explores generally to learn what variables to study and then studies those variables with a large sample of individuals.

Alternatively, researchers may first survey a large number of individuals and then follow up with a few participants to obtain their specific language and voices about the topic. In these situations, collecting both closed-ended quantitative data and open-ended qualitative data proves advantageous.

Personal Experiences

Researchers' own personal training and experiences also influence their choice of approach. An individual trained in technical, scientific writing, statistics, and computer statistical programs and familiar with quantitative journals in the library would most likely choose the quantitative design. On the other hand, individuals who enjoy writing in a literary way or conducting personal interviews or making up-close observations may gravitate to the qualitative approach. The mixed methods researcher is an individual familiar with both quantitative and qualitative research. This person also has the time and resources to collect both quantitative and qualitative data and has outlets for mixed methods studies, which tend to be large in scope.

Since quantitative studies are the traditional mode of research, carefully worked out procedures and rules exist for them. Researchers may be more comfortable with the highly systematic procedures of quantitative research. Also, for some individuals, it can be uncomfortable to challenge accepted approaches among some faculty by using qualitative and advocacy/participatory approaches to inquiry. On the other hand, qualitative approaches allow room to be innovative and to work more within researcher-designed frameworks. They allow more creative, literary-style writing, a form that individuals may like to use. For advocacy/participatory writers, there is undoubtedly a strong stimulus to pursue topics that are of personal interest—issues that relate to marginalized people and an interest in creating a better society for them and everyone.

For the mixed methods researcher, the project will take extra time because of the need to collect and analyze both quantitative and qualitative data. It fits a person who enjoys both the structure of quantitative research and the flexibility of qualitative inquiry.

Audience

Finally, researchers write for audiences that will accept their research. These audiences may be journal editors, journal readers, graduate committees, conference attendees, or colleagues in the field. Students should consider the approaches typically supported and used by their advisers. The experiences of these audiences with quantitative, qualitative, or mixed methods studies can shape the decision made about this choice.

SUMMARY

In planning a research project, researchers need to identify whether they will employ a qualitative, quantitative, or mixed methods design. This design is based on bringing together a worldview or assumptions about research, the specific strategies of inquiry, and research methods. Decisions about choice of a design are further influenced by the research problem or issue being studied, the personal experiences of the researcher, and the audience for whom the researcher writes.

WRITING EXERCISES

Writing Exercises

1. Identify a research question in a journal article and discuss what design would be best to study the question and why.

2. Take a topic that you would like to study, and using the four combinations of worldviews, strategies of inquiry, and research methods in Figure 1.1, discuss a project that brings together a worldview, strategies, and methods. Identify whether this would be quantitative, qualitative, or mixed methods research.

3. What distinguishes a quantitative study from a qualitative study? Mention three characteristics.

ADDITIONAL READINGS

Cherryholmes, C. H. (1992, August-September). Notes on pragmatism and scientific realism. *Educational Researcher, 14*, 13–17.

Cleo Cherryholmes discusses pragmatism as a contrasting perspective from scientific realism. The strength of this article lies in the numerous citations of writers about pragmatism and a clarification of one version of pragmatism. Cherryholmes's version points out that pragmatism is driven by anticipated consequences, reluctance to tell a true story, and the embrace of the idea that there is an external world independent of our minds. Also included in this article are numerous references to historical and recent writers about pragmatism as a philosophical position.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage.

Michael Crotty offers a useful framework for tying together the many epistemological issues, theoretical perspectives, methodology, and methods of social research. He interrelates the four components of the research process and shows in a table a representative sampling of topics of each component. He then goes on to discuss nine different theoretical orientations in social research, such as postmodernism, feminism, critical inquiry, interpretivism, constructionism, and positivism.

Kemmis, S., & Wilkinson, M. (1998). Participatory action research and the study of practice. In B. Atweh, S. Kemmis, & P. Weeks (Eds.), *Action research in practice: Partnerships for social justice in education* (pp. 21–36). New York: Routledge.

Stephen Kemmis and Mervyn Wilkinson provide an excellent overview of participatory research. In particular, they note the six major features of this inquiry approach and then discuss how action research is practiced at the individual level, the social level, or both levels.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln, *The Sage handbook of qualitative research* (3rd ed., pp. 191–215). Thousand Oaks, CA: Sage

Yvonna Lincoln and Egon Guba have provided the basic beliefs of five alternative inquiry paradigms in social science research: positivism, postpositivism, critical theory, constructivism, and participatory. These extend the earlier analysis provided in the first and second editions of the *Handbook*. Each is presented in terms of ontology (i.e., nature of reality), epistemology (i.e., how we know what we know), and methodology (i.e., the process of research). The participatory paradigm adds another alternative paradigm to those originally advanced in the first edition. After briefly presenting these five approaches, they contrast them in terms of seven issues, such as the nature of knowledge, how knowledge accumulates, and goodness or quality criteria.

Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches*. Boston: Allyn & Bacon.

Lawrence Neuman provides a comprehensive research method text as an introduction to social science research. Especially helpful in understanding the alternative meaning of methodology is Chapter 4, titled, "The Meanings of Methodology," in which he contrasts three methodologies—positivist social science, interpretive social science, and critical social science—in terms of eight questions (e.g., What constitutes an explanation or theory of social reality? What does good evidence or factual information look like?).

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield.

D. C. Phillips and Nicholas Burbules summarize the major ideas of postpostivist thinking. Through two chapters, "What is Postpositivism?" and "Philosophical Commitments of Postpositivist Researchers," the authors advance major ideas about postpositivism, especially those that differentiate it from positivism. These include knowing that human knowledge is conjectural rather than unchallengeable and that our warrants for knowledge can be withdrawn in light of further investigations.

Creswell, J. W. 2009. Research Design: Qualitative, Quantitative, and Mixed Methods. Ch 3: The Use of Theory

This page intentionally left blank

CHAPTER THREE

The Use of Theory

One component of reviewing the literature is to determine what theories might be used to explore the questions in a scholarly study. In *quantitative research*, researchers often test theories as an explanation for answers to their questions. In a quantitative dissertation, an entire section of a research proposal might be devoted to presenting the theory for the study. In *qualitative research*, the use of theory is much more varied. The inquirer may generate a theory as the final outcome of a study and place it at the end of a project, such as in grounded theory. In other qualitative studies, it comes at the beginning and provides a lens that shapes what is looked at and the questions asked, such as in ethnographies or in advocacy research. In mixed methods research, researchers may both test theories and generate them. Moreover, mixed methods research may contain a theoretical lens, such as a focus on feminist, racial, or class issues, that guides the entire study.

I begin this chapter by focusing on theory use in a quantitative study. It reviews a definition of a theory, the use of variables in a quantitative study, the placement of theory in a quantitative study, and the alternative forms it might assume in a written plan. Procedures in identifying a theory are next presented, followed by a script of a theoretical perspective section of a quantitative research proposal. Then the discussion moves to the use of theory in a qualitative study. Qualitative inquirers use different terms for theories, such as *patterns, theoretical lens,* or *naturalistic generalizations,* to describe the broader explanations used or developed in their studies. Examples in this chapter illustrate the alternatives available to qualitative researchers. Finally, the chapter turns to the use of theories in mixed methods research and the use of a transformative perspective that is popular in this approach.

QUANTITATIVE THEORY USE

Variables in Quantitative Research

Before discussing quantitative theories, it is important to understand variables and the types that are used in forming theories. A **variable** refers to a characteristic or attribute of an individual or an organization that can be measured or observed and that varies among the people or organization being studied (Creswell, 2007a). A variable typically will vary in two or more categories or on a continuum of scores, and it can be measured or assessed on a scale. Psychologists prefer to use the term *construct* (rather than *variable*), which carries the connotation more of an abstract idea than a specifically defined term. However, social scientists typically use the term *variable*, and it will be employed in this discussion. Variables often measured in studies include gender, age, socioeconomic status (SES), and attitudes or behaviors such as racism, social control, political power, or leadership. Several texts provide detailed discussions about the types of variables one can use and their scales of measurement (e.g., Isaac & Michael, 1981; Keppel, 1991; Kerlinger, 1979; Thorndike, 1997). Variables are distinguished by two characteristics: temporal order and their measurement (or observation).

Temporal order means that one variable precedes another in time. Because of this time ordering, it is said that one variable affects or causes another variable, though a more accurate statement would be that one variable *probably* causes another. When dealing with studies in the natural setting and with humans, researchers cannot absolutely prove cause and effect (Rosenthal & Rosnow, 1991), and social scientists now say that there is probable causation. Temporal order means that quantitative researchers think about variables in an order from "left to right" (Punch, 2005) and order the variables in purpose statements, research questions, and visual models into left-to-right, cause-and-effect presentations. Thus,

- *Independent variables* are those that (probably) cause, influence, or affect outcomes. They are also called *treatment, manipulated, antecedent*, or *predictor* variables.

- *Dependent variables* are those that depend on the independent variables; they are the outcomes or results of the influence of the independent variables. Other names for dependent variables are *criterion, outcome*, and *effect* variables.

- *Intervening* or *mediating variables* stand between the independent and dependent variables, and they mediate the effects of the independent variable on the dependent variable. For example, if students do well on a research methods test (dependent variable), that result may be due to (a) their study preparation (independent variable) and/or (b) their organization of study ideas into a framework (intervening variable) that influenced their performance on the test. The mediating variable, the organization of study, stands between the independent and dependent variables.

- *Moderating variables* are new variables constructed by a researcher by taking one variable and multiplying it by another to determine the joint impact of both (e.g., age X attitudes toward quality of life). These variables are typically found in experiments.

- Two other types of variables are *control variables* and *confounding variables*. Control variables play an active role in quantitative studies. These are a special type of independent variable that researchers measure because they potentially influence the dependent variable. Researchers use statistical procedures (e.g., analysis of covariance) to control for these variables. They may be demographic or personal variables (e.g., age or gender) that need to be "controlled" so that the true influence of the independent variable on the dependent can be determined. Another type of variable, a *confounding* (or *spurious*) *variable*, is not actually measured or observed in a study. It exists, but its influence cannot be directly detected. Researchers comment on the influence of confounding variables after the study has been completed, because these variables may have operated to explain the relationship between the independent variable and dependent variable, but they were not or could not be easily assessed (e.g., discriminatory attitudes).

In a quantitative research study, variables are related to answer a research question (e.g., "How does self-esteem influence the formation of friendships among adolescents?") or to make predictions about what the researcher expects the results to show. These predictions are called *hypotheses* (e.g., "Individual positive self-esteem expands the number of friends of adolescents.")

Definition of a Theory

With this background on variables, we can proceed to the use of quantitative theories. In *quantitative* research, some historical precedent exists for viewing a theory as a scientific prediction or explanation (see G. Thomas, 1997, for different ways of conceptualizing theories and how they might constrain thought). For example, Kerlinger's (1979) definition of a theory is still valid today. He said, a theory is "a set of interrelated constructs (variables), definitions, and propositions that presents a systematic view of phenomena by specifying relations among variables, with the purpose of explaining natural phenomena" (p. 64).

In this definition, a **theory** is an interrelated set of constructs (or variables) formed into propositions, or hypotheses, that specify the relationship among variables (typically in terms of magnitude or direction). A theory might appear in a research study as an argument, a discussion, or a rationale, and it helps to explain (or predict) phenomena that occur in the world. Labovitz and Hagedorn (1971) add to this definition the idea of a *theoretical rationale*, which they define as "specifying how and why the variables and relational statements are interrelated" (p. 17). Why would an independent variable, X, influence or affect a dependent variable, Y? The theory would provide the explanation for this expectation or prediction. A discussion about this theory would appear in a section of a proposal on the literature review or on the *theory base, the theoretical rationale,* or *the theoretical perspective.* I prefer the term *theoretical perspective* because it has been popularly used as a required section for proposals for research when one submits an application to present a paper at the American Educational Research Association conference.

The metaphor of a rainbow can help to visualize how a theory operates. Assume that the rainbow *bridges* the independent and dependent variables (or constructs) in a study. This rainbow ties together the variables and provides an overarching explanation for *how* and *why* one would expect the independent variable to explain or predict the dependent variable. Theories develop when researchers test a prediction over and over. For example, here is how the process of developing a theory works. Investigators combine independent, mediating, and dependent variables based on different forms of measures into questions. These questions provide information about the type of relationship (positive, negative, or unknown) and its magnitude (e.g., high or low). Forming this information into a predictive statement (hypothesis), a researcher might write, "The greater the centralization of power in leaders, the greater the disenfranchisement of the followers." When researchers test hypotheses such as this over and over in different settings and with different populations (e.g., the Boy Scouts, a Presbyterian church, the Rotary Club, and a group of high school students), a theory emerges, and someone gives it a name (e.g., a theory of attribution). Thus, theory develops as an explanation to advance knowledge in particular fields (Thomas, 1997).

Another aspect of theories is that they vary in their breadth of coverage. Neuman (2000) reviews theories at three levels: micro-level, meso-level, and macro-level. Micro-level theories provide explanations limited to small slices of time, space, or numbers of people, such as Goffman's theory of face work, which explains how people engage in rituals during face-to-face interactions. Meso-level theories link the micro and macro levels. These are theories of organizations, social movement, or communities, such as Collins's theory of control in organizations. Macro-level theories explain larger aggregates, such as social institutions, cultural systems, and whole societies. Lenski's macro-level theory of social stratification, for example, explains how the amount of surplus a society produces increases with the development of the society.

Theories are found in the social science disciplines of psychology, sociology, anthropology, education, and economics, as well as within many subfields. To locate and read about these theories requires searching literature databases (e.g., *Psychological Abstracts, Sociological Abstracts*) or reviewing guides to the literature about theories (e.g., see Webb, Beals, & White, 1986).

Forms of Theories

Researchers state their theories in research proposals in several ways, such as a series of hypotheses, if–then logic statements, or visual models. First, some researchers state theories in the form of interconnected hypotheses. For example, Hopkins (1964) conveyed his theory of influence processes as a series of 15 hypotheses. Some of the hypotheses are as follows (these have been slightly altered to remove the gender-specific pronouns):

1. The higher one's rank, the greater one's centrality.
2. The greater one's centrality, the greater one's observability.
3. The higher one's rank, the greater one's observability.
4. The greater one's centrality, the greater one's conformity.
5. The higher one's rank, the greater one's conformity.
6. The greater one's observability, the greater one's conformity.
7. The greater one's conformity, the greater one's observability. (p. 51)

A second way is to state a theory as a series of if–then statements that explain why one would expect the independent variables to influence or cause the dependent variables. For example, Homans (1950) explains a theory of interaction:

> If the frequency of interaction between two or more persons increases, the degree of their liking for one another will increase, and vice versa. . . . Persons who feel sentiments of liking for one another will express those sentiments in activities over and above the activities of the external system, and these activities may further strengthen the sentiments of liking. The more frequently persons interact with one another, the more alike in some respects both their activities and their sentiments tend to become. (pp. 112, 118, 120)

Third, an author may present a theory as a visual model. It is useful to translate variables into a visual picture. Blalock (1969, 1985, 1991) advocates causal modeling and recasts verbal theories into causal models so that a reader can visualize the interconnections of variables. Two simplified examples are presented here. As shown in Figure 3.1, three independent variables influence a single dependent variable, mediated by the influence of two intervening variables. A diagram such as this one shows the possible causal sequence among variables leading to modeling through path analysis and more advanced analyses using multiple measures of variables as found in structural equation modeling (see Kline, 1998). At an introductory level, Duncan (1985) provides useful suggestions about the notation for constructing these visual causal diagrams:

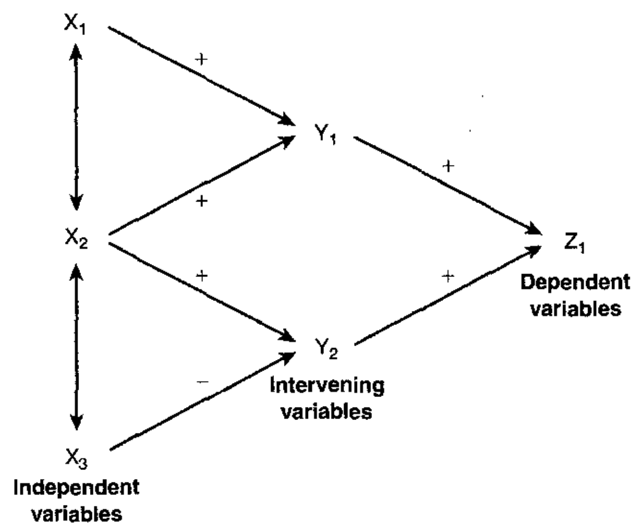


Figure 3.1 Three Independent Variables Influence a Single Dependent Variable Mediated by Two Intervening Variables

- Position the dependent variables on the right in the diagram and the independent variables on the left.
- Use one-way arrows leading from each determining variable to each variable dependent on it.
- Indicate the strength of the relationship among variables by inserting valence signs on the paths. Use positive or negative valences that postulate or infer relationships.
- Use two-headed arrows connected to show unanalyzed relationships between variables not dependent upon other relationships in the model.

More complicated causal diagrams can be constructed with additional notation. This one portrays a basic model of limited variables, such as typically found in a survey research study.

A variation on this theme is to have independent variables in which control and experimental groups are compared on one independent variable in terms of an outcome (dependent variable). As shown in Figure 3.2, two groups on variable X are compared in terms of their influence on Y, the dependent variable. This design is a between-groups experimental design (see Chapter 8). The same rules of notation previously discussed apply.

These two models are meant only to introduce possibilities for connecting independent and dependent variables to build theories. More complicated designs employ multiple independent and dependent variables in elaborate models of causation (Blalock, 1969, 1985). For example,

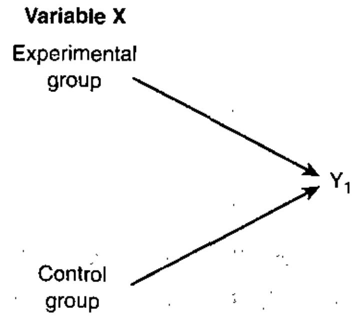


Figure 3.2 Two Groups With Different Treatments on X Are Compared in Terms of Y

Jungnickel (1990), in a doctoral dissertation proposal about research productivity among faculty in pharmacy schools, presented a complex visual model, as shown in Figure 3.3. Jungnickel asked what factors influence a faculty member's scholarly research performance. After identifying these factors in the literature, he adapted a theoretical framework found in nursing research (Megel, Langston, & Creswell, 1988) and developed a visual model portraying the relationship among these factors, following the rules for constructing a model introduced earlier. He listed the independent variables on the far left, the intervening variables in the middle, and the dependent variables on the right. The direction of influence flowed from the left to the right, and he used plus and minus signs to indicate the hypothesized direction.

Placement of Quantitative Theories

In *quantitative* studies, one uses theory deductively and places it toward the beginning of the proposal for a study. With the objective of testing or verifying a theory rather than developing it, the researcher advances a theory, collects data to test it, and reflects on its confirmation or disconfirmation by the results. The theory becomes a framework for the entire study, an organizing model for the research questions or hypotheses and for the data collection procedure. The deductive model of thinking used in a quantitative study is shown in Figure 3.4. The researcher tests or verifies a theory by examining hypotheses or questions derived from it. These hypotheses or questions contain variables (or constructs) that the researcher needs to define. Alternatively, an acceptable definition might be found in the literature. From here, the investigator locates an instrument to use in measuring or observing attitudes or behaviors of participants in a study. Then the investigator collects scores on these instruments to confirm or disconfirm the theory.

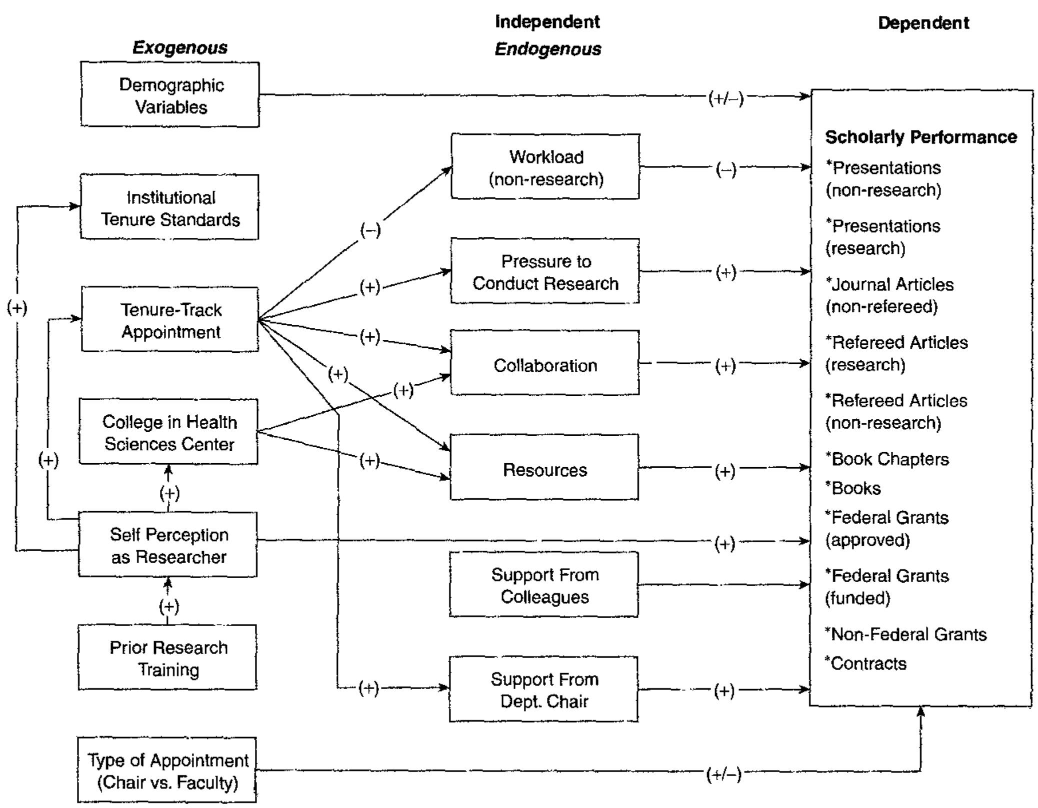


Figure 3.3 A Visual Model of a Theory of Faculty Scholarly Performance

SOURCE: Jungnickel (1990). Reprinted with permission.

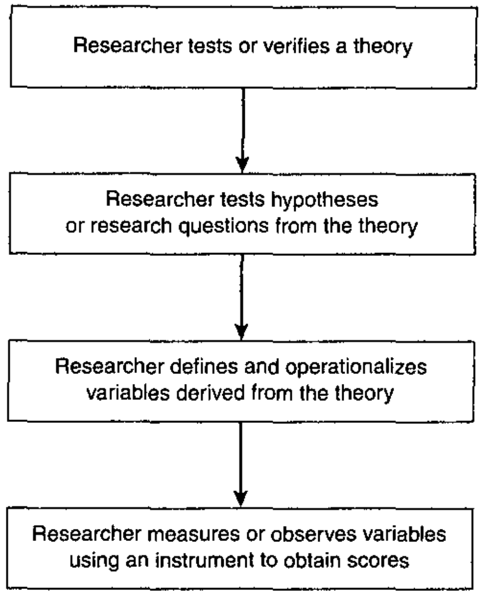


Figure 3.4 The Deductive Approach Typically Used in Quantitative Research

This deductive approach to research in the quantitative approach has implications for the *placement of a theory* in a quantitative research study (see Table 3.1).

A general guide is to introduce the theory early in a plan or study: in the introduction, in the literature review section, immediately after hypotheses or research questions (as a rationale for the connections among the variables), or in a separate section of the study. Each placement has its advantages and disadvantages.

A **research tip**: I write the theory into a separate section in a research proposal so that readers can clearly identify the theory from other components. Such a separate passage provides a complete explication of the theory section, its use, and how it relates to the study.

Writing a Quantitative Theoretical Perspective

Using these ideas, the following presents a model for writing a quantitative theoretical perspective section into a research plan. Assume that the task is to identify a theory that explains the relationship between independent and dependent variables.

1. Look in the discipline-based literature for a theory. If the unit of analysis for variables is an individual, look in the psychology literature; to study groups or organizations, look in the sociological literature. If the

Table 3.1 Options for Placing Theory in a Quantitative Study

Placement	Advantages	Disadvantages
In the introduction	An approach often found in journal articles, it will be familiar to readers. It conveys a deductive approach.	It is difficult for a reader to isolate and separate theory base from other components of the research process.
In the literature review	Theories are found in the literature and their inclusion in a literature review is a logical extension or part of the literature.	It is difficult for a reader to see the theory in isolation from the scholarly review of the literature.
After hypotheses or research questions	The theory discussion is a logical extension of hypotheses or research questions because it explains how and why variables are related.	A writer may include a theoretical rationale after hypotheses and questions and leave out an extended discussion about the origin and use of the theory.
In a separate section	This approach clearly separates the theory from other components of the research process, and it enables a reader to better identify and to understand the theory base for the study.	The theory discussion stands in isolation from other components of the research process and, as such, a reader may not easily connect it with other components of the research process.

project examines individuals and groups, consider the social psychology literature. Of course, theories from other disciplines may be useful, too (e.g., to study an economic issue, the theory may be found in economics).

2. Examine also prior studies that address the topic or a closely related topic. What theories were used by other authors? Limit the number of theories and try to identify *one overarching theory* that explains the central hypothesis or major research question.

3. As mentioned earlier, ask the *rainbow* question that bridges the independent and dependent variables: Why would the independent variable(s) influence the dependent variables?

4. Script out the theory section. Follow these lead sentences: "The theory that I will use is _____ (name the theory). It was developed by _____ (identify the origin, source, or developer of the theory), and it was

used to study _____ (identify the topics where one finds the theory being applied). This theory indicates that _____ (identify the propositions or hypotheses in the theory). As applied to my study, this theory holds that I would expect my independent variable(s) _____ (state independent variables) to influence or explain the dependent variable(s) _____ (state dependent variables) because _____ (provide a rationale based on the logic of the theory)."

Thus, the topics to include in a quantitative theory discussion are the theory to be used, its central hypotheses or propositions, information about past use of the theory and its application, and statements that reflect how it relates to a proposed study. This model is illustrated in the following example by Crutchfield (1986).

Example 3.1 *A Quantitative Theory Section*

Crutchfield (1986) wrote a doctoral dissertation titled *Locus of Control, Interpersonal Trust, and Scholarly Productivity*. Surveying nursing educators, her intent was to determine if locus of control and interpersonal trust affected the levels of publications of the faculty. Her dissertation included a separate section in the introductory chapter titled "Theoretical Perspective," which follows. It includes these points:

- The theory she planned to use
- The central hypotheses of the theory
- Information about who has used the theory and its applicability
- An adaptation of the theory to variables in her study using if–then logic

I have added annotations in italics to mark key passages.

Theoretical Perspective

In formulation of a theoretical perspective for studying the scholarly productivity of faculty, social learning theory provides a useful prototype. This conception of behavior attempts to achieve a balanced synthesis of cognitive psychology with the principles of behavior modification (Bower & Hilgard, 1981). Basically, this unified theoretical framework "approaches the explanation of human behavior in terms of a continuous (reciprocal) interaction between cognitive, behavioral, and environmental determinants" (Bandura, 1977, p. vii). *(Author identifies the theory for the study.)*

(Continued)

(Continued)

While social learning theory accepts the application of reinforcements such as shaping principles, it tends to see the role of rewards as both conveying information about the optimal response and providing incentive motivation for a given act because of the anticipated reward. In addition, the learning principles of this theory place special emphasis on the important roles played by vicarious, symbolic, and self-regulating processes (Bandura, 1971).

Social learning theory not only deals with learning, but seeks to describe how a group of social and personal competencies (so called personality) could evolve out of social conditions within which the learning occurs. It also addresses techniques of personality assessment (Mischel, 1968), and behavior modification in clinical and educational settings (Bandura, 1977; Bower & Hilgard, 1981; Rotter, 1954). *(Author describes social learning theory.)*

Further, the principles of social learning theory have been applied to a wide range of social behavior such as competitiveness, aggressiveness, sex roles, deviance, and pathological behavior (Bandura & Walters, 1963; Bandura, 1977; Mischel, 1968; Miller & Dollard, 1941; Rotter, 1954; Staats, 1975). *(Author describes the use of the theory.)*

Explaining social learning theory, Rotter (1954) indicated that four classes of variables must be considered: behavior, expectancies, reinforcement, and psychological situations. A general formula for behavior was proposed which states: "the potential for a behavior to occur in any specific psychological situation is the function of the expectancy that the behavior will lead to a particular reinforcement in that situation and the value of that reinforcement" (Rotter, 1975, p. 57).

Expectancy within the formula refers to the perceived degree of certainty (or probability) that a causal relationship generally exists between behavior and rewards. This construct of generalized expectancy has been defined as internal locus of control when an individual believes that reinforcements are a function of specific behavior, or as external locus of control when the effects are attributed to luck, fate, or powerful others. The perceptions of causal relationships need not be absolute positions, but rather tend to vary in degree along a continuum depending upon previous experiences and situational complexities (Rotter, 1966). *(Author explains variables in the theory.)*

In the application of social learning theory to this study of scholarly productivity, the four classes of variables identified by Rotter (1954) will be defined in the following manner.

1. Scholarly productivity is the desired behavior or activity.
2. Locus of control is the generalized expectancy that rewards are or are not dependent upon specific behaviors.
3. Reinforcements are the rewards from scholarly work and the value attached to these rewards.
4. The educational institution is the psychological situation which furnishes many of the rewards for scholarly productivity.

With these specific variables, the formula for behavior which was developed by Rotter (1975) would be adapted to read: The potential for scholarly behavior to occur within an educational institution is a function of the expectancy that this activity will lead to specific rewards and of the value that the faculty member places on these rewards. In addition, the interaction of interpersonal trust with locus of control must be considered in relation to the expectancy of attaining rewards through behaviors as recommended in subsequent statements by Rotter (1967). Finally, certain characteristics, such as educational preparation, chronological age, post-doctoral fellowships, tenure, or full-time versus part-time employment may be associated with the scholarly productivity of nurse faculty in a manner similar to that seen within other disciplines. *(Author applied the concepts to her study.)*

The following statement represents the underlying logic for designing and conducting this study. If faculty believe that: (a) their efforts and actions in producing scholarly works will lead to rewards (locus of control), (b) others can be relied upon to follow through on their promises (interpersonal trust), (c) the rewards for scholarly activity are worthwhile (reward values), and (d) the rewards are available within their discipline or institution (institutional setting), then they will attain high levels of scholarly productivity (pp. 12–16). *(Author concluded with the if-then logic to relate the independent variables to the dependent variables.)*

QUALITATIVE THEORY USE

Variation in Theory Use in Qualitative Research

Qualitative inquirers use theory in their studies in several ways. First, much like in quantitative research, it is used as a broad explanation for behavior and attitudes, and it may be complete with variables, constructs, and hypotheses. For example, ethnographers employ cultural themes or "aspects of culture" (Wolcott, 1999, p. 113) to study in their qualitative projects, such as social control, language, stability and change, or social organization, such as kinship or families (see Wolcott's 1999 discussion about texts that address cultural topics in anthropology). Themes in this context provide a ready-made series of hypotheses to be tested from the literature. Although researchers might not refer to them as theories, they provide broad explanations that anthropologists use to study the culture-sharing behavior and attitudes of people. This approach is popular in qualitative health science research in which investigators begin with a theoretical model, such as the adoption of health practices or a quality of life theoretical orientation.

Second, researchers increasingly use a **theoretical lens** or **perspective in qualitative research,** which provides an overall orienting lens for the study of questions of gender, class, and race (or other issues of marginalized groups). This lens becomes an advocacy perspective that shapes the types of questions asked, informs how data are collected and analyzed, and provides a call for action or change. Qualitative research of the 1980s underwent a transformation to broaden its scope of inquiry to include these theoretical lenses. They guide the researchers as to what issues are important to examine (e.g., marginalization, empowerment) and the people that need to be studied (e.g., women, homeless, minority groups). They also indicate how the researcher positions himself or herself in the qualitative study (e.g., up front or biased from personal, cultural, and historical contexts) and how the final written accounts need to be written (e.g., without further marginalizing individuals, by collaborating with participants). In critical ethnography studies, researchers begin with a theory that informs their studies. This causal theory might be one of emancipation or repression (Thomas, 1993).

Some of these qualitative theoretical perspectives available to the researcher are as follows (Creswell, 2007):

- *Feminist perspectives* view as problematic women's diverse situations and the institutions that frame those situations. Research topics may include policy issues related to realizing social justice for women in specific contexts or knowledge about oppressive situations for women (Olesen, 2000).

- *Racialized discourses* raise important questions about the control and production of knowledge, particularly about people and communities of color (Ladson-Billings, 2000).

- *Critical theory* perspectives are concerned with empowering human beings to transcend the constraints placed on them by race, class, and gender (Fay, 1987).

- *Queer theory*—a term used in this literature—focuses on individuals calling themselves lesbians, gays, bisexuals, or transgendered people. The research using this approach does not objectify individuals, is concerned with cultural and political means, and conveys the voices and experiences of individuals who have been suppressed (Gamson, 2000).

- *Disability inquiry* addresses the meaning of inclusion in schools and encompasses administrators, teachers, and parents who have children with disabilities (Mertens, 1998).

Rossman and Rallis (1998) capture the sense of theory as critical and postmodern perspectives in qualitative inquiry:

> As the 20th century draws to a close, traditional social science has come under increasing scrutiny and attack as those espousing critical and postmodern perspectives challenge objectivist assumptions and traditional norms for the conduct of research. Central to this attack are four interrelated notions: (a) Research fundamentally involves issues of power; (b) the research report is not transparent but rather it is authored by a raced, gendered, classed, and politically oriented individual; (c) race, class, and gender are crucial for understanding experience; and (d) historic, traditional research has silenced members of oppressed and marginalized groups. (p. 66)

Third, distinct from this theoretical orientation are qualitative studies in which theory (or some other broad explanation) becomes the *end point.* It is an inductive process of building from the data to broad themes to a generalized model or theory (see Punch, 2005). The logic of this inductive approach is shown in Figure 3.5.

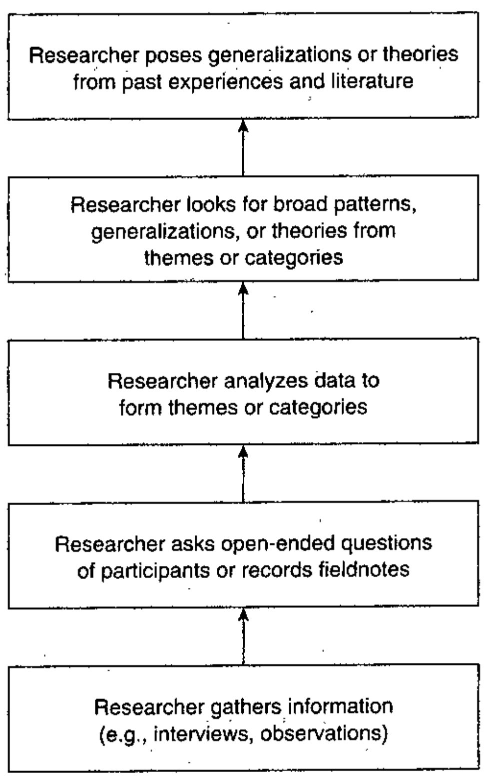


Figure 3.5 The Inductive Logic of Research in a Qualitative Study

The researcher begins by gathering detailed information from participants and then forms this information into categories or themes. These themes are developed into broad patterns, theories, or generalizations that are then compared with personal experiences or with existing literature on the topic.

The development of themes and categories into patterns, theories, or generalizations suggests varied end points for qualitative studies. For example, in case study research, Stake (1995) refers to an assertion as a *propositional generalization*—the researcher's summary of interpretations and claims—to which is added the researcher's own personal experiences, called "naturalistic generalizations" (p. 86). As another example, grounded theory provides a different end point. Inquirers hope to discover a theory that is grounded in information from participants (Strauss & Corbin, 1998). Lincoln and Guba (1985) refer to "pattern theories" as explanations that develop during naturalistic or qualitative research. Rather than the deductive form found in quantitative studies, these pattern theories or generalizations represent interconnected thoughts or parts linked to a whole.

Neuman (2000) provides additional information about pattern theories:

> Pattern theory does not emphasize logical deductive reasoning. Like causal theory, it contains an interconnected set of concepts and relationships, but it does not require causal statements. Instead, pattern theory uses metaphor or analogies so that relationship "makes sense." Pattern theories are systems of ideas that inform. The concepts and relations within them form a mutually reinforcing, closed system. They specify a sequence of phases or link parts to a whole. (p. 38)

Fourth and finally, some qualitative studies *do not employ any explicit theory*. However, the case can be made that no qualitative study begins from pure observation and that prior conceptual structure composed of theory and method provides the starting point for all observations (Schwandt, 1993). Still, one sees qualitative studies that contain no *explicit* theoretical orientation, such as in phenomenology, in which inquirers attempt to build the essence of experience from participants (e.g., see Riemen, 1986). In these studies, the inquirer constructs a rich, detailed description of a central phenomenon.

My **research tips** on theory use in a qualitative proposal are as follows:

- Decide if theory is to be used in the qualitative proposal.
- If it is used, then identify how the theory will be used in the study, such as an up-front explanation, as an end point, or as an advocacy lens.
- Locate the theory in the proposal in a manner consistent with its use.

Locating the Theory in Qualitative Research

How theory is used affects its placement in a qualitative study. In those studies with a cultural theme or a theoretical lens, the theory occurs in the opening passages of the study. Consistent with the emerging design of qualitative inquiry, the theory may appear at the beginning and be modified or adjusted based on participant views. Even in the most theory-oriented qualitative design, such as critical ethnography, Lather (1986) qualifies the use of theory:

> Building empirically grounded theory requires a reciprocal relationship between data and theory. Data must be allowed to generate propositions in a dialectical manner that permits use of *a priori* theoretical frameworks, but which keeps a particular framework from becoming the container into which the data must be poured. (p. 267)

Example 3.2 *A Theory Early in a Qualitative Study*

Murguia, Padilla, and Pavel (1991) studied the integration of 24 Hispanic and Native American students into the social system of a college campus. They were curious about how ethnicity influenced social integration, and they began by relating the participants' experiences to a theoretical model, the Tinto model of social integration. They felt that the model had been "incompletely conceptualized and, as a consequence, only imprecisely understood and measured" (p. 433).

Thus, the model was not being tested, as one would find in a quantitative project, but modified. At the end of the study, the authors refined Tinto's model and advanced their modification that described how ethnicity functions. In contrast to this approach, in qualitative studies with an end point of a theory (e.g., a grounded theory), a pattern, or a generalization, the theory emerges at the end of the study. This theory might be presented as a logic diagram, a visual representation of relationships among concepts.

Example 3.3 *A Theory at the End of a Qualitative Study*

Using a national database of 33 interviews with academic department chairpersons, we (Creswell & Brown, 1992) developed a grounded theory interrelating variables (or categories) of chair influence on scholarly performance of faculty. The theory section came into the article as the last section.

(Continued)

(Continued)

where we presented a visual model of the theory developed inductively from categories of information supplied by interviewees. In addition, we also advanced directional hypotheses that logically followed from the model. Moreover, in the section on the model and the hypotheses, we compared the results from participants with results from other studies and the theoretical speculations in the literature. For example, we stated,

> This proposition and its sub-propositions represent unusual, even contrary evidence, to our expectations. Contrary to proposition 2.1, we expected that the career stages would be similar not in type of issue but in the range of issues. Instead we found that the issues for post-tenure faculty covered almost all the possible problems on the list. Why were the tenured faculty's needs more extensive than non-tenured faculty? The research productivity literature suggests that one's research performance does not decline with the award of tenure (Holley 1977). Perhaps diffuse career goals of post-tenure faculty expand the possibilities for "types" of issues. In any case, this sub-proposition focuses attention on the understudied career group that Furniss (1981) reminds us needs to be examined in more detail.
>
> (Creswell & Brown, 1992, p. 58)

As this example shows, we developed a visual model that interrelated variables, derived this model inductively from informant comments, and placed the model at the end of the study, where the central propositions in it could be contrasted with the existing theories and literature.

MIXED METHODS THEORY USE

Theory use in mixed methods studies may include theory deductively, in quantitative theory testing and verification, or inductively as in an emerging qualitative theory or pattern. A social science or a health science theory may be used as a framework to be tested in either a quantitative or qualitative approach to inquiry. Another way to think about theory in mixed methods research is as a *theoretical lens* or *perspective* to guide the study. Studies are beginning to emerge that employ mixed methods designs using a lens to study gender, race or ethnicity, disability, sexual orientation, and other bases of diversity (Mertens, 2003).

Historically, the idea of using a theoretical lens in mixed methods research was mentioned by Greene and Caracelli in 1997. They identified the use of a *transformative design* as a distinct form of mixed methods research. This design gave primacy to value-based, action-oriented research, such as in participatory action research and empowerment approaches. In this design, they suggest mixing the value commitments of different traditions (e.g., bias-free from quantitative and bias-laden from qualitative), the use of diverse methods, and a focus on action solutions. The implementation of these ideas in the practice of mixed methods research has now been carried forward by other authors.

More information on procedures has appeared in a chapter written by Creswell, Plano Clark, Gutmann, and Hanson (2003). They identified the use of theoretical perspectives, such as gendered, feminist; cultural/ racial/ethnic; lifestyle; critical; and class and social status. These perspectives became an overlay over mixed methods designs (see Chapter 10). They further developed visual models to portray how these lenses might provide a guiding perspective for a mixed methods study. Mertens (2003) continued the discussion. As outlined in Box 3.1, she advocated for the importance of a theory lens in mixed methods research. In detailing a transformative–emancipatory paradigm and specific procedures, she emphasized the role that values played in studying feminist, ethnic/racial, and disability issues. Her transformative theory was an umbrella term for research that was emancipatory, antidiscriminatory, participative, Freirian, feminist, racial/ ethnic, for individuals with disabilities, and for all marginalized groups.

Mertens identifies the implications of these transformative theories for mixed methods research. These involve integration of the transformative–emancipatory methodology into all phases of the research process. Reading through the questions in Box 3.1, one gains a sense of the importance of studying issues of discrimination and oppression and of recognizing diversity among study participants. These questions also address treating individuals respectfully through gathering and communicating data collection and through reporting results that lead to changes in social processes and relationships.

Box 3.1 Transformative-Emancipatory Questions for Mixed Methods Researchers Throughout the Research Process

Defining the Problem and Searching the Literature

- Did you deliberately search the literature for concerns of diverse groups and issues of discrimination and oppression?
- Did the problem definition arise from the community of concern?
- Did your mixed methods approach arise from spending quality time with these communities? (i.e., building trust? using an appropriate theoretical framework other than a deficit model? developing balanced—positive and negative—questions? developing questions that lead to transformative answers, such as questions focused on authority and relations of power in institutions and communities?)

(Continued)

(Continued)

Identifying the Research Design

- Does your research design deny treatment to any groups and respect ethical considerations of participants?

Identifying Data Sources and Selecting Participants

- Are the participants of groups associated with discrimination and oppression?
- Are the participants appropriately labeled?
- Is there a recognition of diversity within the target population?
- What can be done to improve the inclusiveness of the sample to increase the probability that traditionally marginalized groups are adequately and accurately represented?

Identifying or Constructing Data Collection Instruments and Methods

- Will the data collection process and outcomes benefit the community being studied?
- Will the research findings be credible to that community?
- Will communication with that community be effective?
- Will the data collection open up avenues for participation in the social change process?

Analyzing, Interpreting, and Reporting and Using Results

- Will the results raise new hypotheses?
- Will the research examine subgroups (i.e., multilevel analyses) to analyze the differential impact on diverse groups?
- Will the results help understand and elucidate power relationships?
- Will the results facilitate social change?

SOURCE: Adapted from D. M. Mertens (2003), "Mixed Methods and the Politics of Human Research: The Transformative-Emancipatory Perspective," in A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in the Social & Behavioral Sciences*. Adapted with permission.

Example 3.4 *Theory in a Transformative-Emancipatory Mixed Methods Study*

Hopson, Lucas, and Peterson (2000) studied issues in an urban, predominantly African American HIV/AIDS community. Consistent with a transformative–emancipatory framework, they examined the language of participants with HIV/AIDS within the participants' social context. They first conducted 75 open-ended ethnographic interviews to identify "language themes" (p. 31), such as blame, ownership, and acceptance or nonacceptance. They also collected 40 semistructured interviews that addressed demographics, daily routine, drug use, knowledge of HIV/AIDS risks, and drug and sexual socio-behavioral characteristics. From this qualitative data, the authors used concepts and questions to refine follow-up questions, including the design of a quantitative postintervention instrument. The authors suggested that empowerment approaches in evaluation can be useful, with researchers listening to the voices of real people and acting on what program participants say.

The design in this study gave "primacy to the value-based and action-oriented dimensions of different inquiry traditions" (Greene & Caracelli, 1997, p. 24) in a mixed methods study. The authors used a theoretical lens for reconfiguring the language and dialogue of participants, and they advanced the importance of empowerment in research.

In using theory in a mixed methods proposal,

- Determine if theory is to be used.
- Identify its use in accord with quantitative or qualitative approaches.
- If theory is used as in a transformational strategy of inquiry, define this strategy and discuss the points in the proposed study in which the emancipatory ideas will be used.

SUMMARY

Theory has a place in quantitative, qualitative, and mixed methods research. Researchers use theory in a quantitative study to provide an explanation or prediction about the relationship among variables in the study. Thus, it is essential to have grounding in the nature and use of variables as they form research questions and hypotheses. A theory explains how and why the variables are related, acting as a bridge between or among the variables. Theory may be broad or narrow in scope, and researchers state their theories in several ways, such as a series of hypotheses, if–then logic statements, or visual models. Using theories deductively, investigators advance them at the beginning of the study in the literature review. They also include them with the hypotheses or research questions or place them in a separate section. A script can help design the theory section for a research proposal.

In qualitative research, inquirers employ theory as a broad explanation, much like in quantitative research, such as in ethnographies. It may also be a theoretical lens or perspective that raises questions related to gender, class, race, or some combination of these. Theory also appears as an end point of a qualitative study, a generated theory, a pattern, or a generalization that emerges inductively from data collection and analysis. Grounded theorists, for example, generate a theory grounded in the views of participants and place it as the conclusion of their studies. Some qualitative studies do not include an explicit theory and present descriptive research of the central phenomenon.

Mixed methods researchers use theory either deductively (as in quantitative research) or inductively (as in qualitative research). Writers also are beginning to identify the use of theoretical lenses or perspectives (e.g., related to gender, lifestyle, race/ethnicity, and class) in their mixed methods studies. A transformational–emancipatory design incorporates this perspective, and recent developments have identified procedures for incorporating this perspective into all phases of the research process.

WRITING EXERCISES

Writing Exercises

1. Write a theoretical perspective section for your research plan following the script for a quantitative theory discussion presented in this chapter.

2. For a quantitative proposal you are planning, draw a visual model of the variables in the theory using the procedures for causal model design advanced in this chapter.

3. Locate qualitative journal articles that (a) use an a priori theory that is modified during the process of research, (b) generate or develop a theory at the end of the study, and (c) represent descriptive research without the use of an explicit theoretical model.

4. Locate a mixed methods study that uses a theoretical lens, such as a feminist, ethnic/racial, or class perspective. Identify specifically how the lens shapes the steps taken in the research process, using Box 3.1 as a guide.

ADDITIONAL READINGS

Flinders, D. J., & Mills, G. E. (Eds.). (1993). *Theory and concepts in qualitative research: Perspectives from the field.* New York: Teachers College Press, Teachers College, Columbia University.

David Flinders and Geoffrey Mills have edited a book about perspectives from the field—"theory at work"—as described by different qualitative researchers. The chapters illustrate little consensus about defining theory and whether it is a vice or virtue. Further, theory operates at many levels in research, such as formal theories, epistemological theories, methodological theories, and meta-theories. Given this diversity, it is best to see actual theory at work in qualitative studies, and this volume illustrates practice from critical, personal, formal, and educational criticism.

Mertens, D. M. (2003). Mixed methods and the politics of human research: The transformative-emancipatory perspective. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 135–164). Thousand Oaks, CA: Sage.

Donna Mertens recognizes that historically, research methods have not concerned themselves with the issues of the politics of human research and social justice. Her chapter explores the transformative–emancipatory paradigm of research as a framework or lens for mixed methods research as it has emerged from scholars from diverse ethnic/racial groups, people with disabilities, and feminists. A unique aspect of her chapter is how she weaves together this paradigm of thinking and the steps in the process of conducting mixed methods research.

Thomas, G. (1997). What's the use of theory? *Harvard Educational Review, 67*(1), 75–104.

Gary Thomas presents a reasoned critique of the use of theory in educational inquiry. He notes the various definitions of theory and maps out four broad uses of theory: (a) as thinking and reflection, (b) as tighter or looser hypotheses, (c) as explanations for adding to knowledge in different fields, and (d) as formally expressed statements in science. Having noted these uses, he then embraces the thesis that theory unnecessarily structures and constrains thought. Instead, ideas should be in a constant flux and should be "ad hocery," as characterized by Toffler.

Creswell, J. W. 2009. Research Design: Qualitative, Quantitative, and Mixed Methods. Ch 7: Research Questions & Hypotheses

This page intentionally left blank

CHAPTER SEVEN

Research Questions and Hypotheses

Investigators place signposts to carry the reader through a plan for a study. The first signpost is the purpose statement, which establishes the central direction for the study. From the broad, general purpose statement, the researcher narrows the focus to specific questions to be answered or predictions based on hypotheses to be tested. This chapter begins by advancing several principles in designing and scripts for writing qualitative research questions; quantitative research questions, objectives, and hypotheses; and mixed methods research questions.

QUALITATIVE RESEARCH QUESTIONS

In a qualitative study, inquirers state research questions, not objectives (i.e., specific goals for the research) or hypotheses (i.e., predictions that involve variables and statistical tests). These research questions assume two forms: a central question and associated subquestions.

The **central question** is a broad question that asks for an exploration of the central phenomenon or concept in a study. The inquirer poses this question, consistent with the emerging methodology of qualitative research, as a general issue so as to not limit the inquiry. To arrive at this question, *ask*, "What is the broadest question that I can ask in the study?" Beginning researchers trained in *quantitative* research might struggle with this approach because they are accustomed to the reverse approach: identifying specific, narrow questions or hypotheses based on a few variables. In qualitative research, the intent is to explore the complex set of factors surrounding the central phenomenon and present the varied perspectives or meanings that participants hold. The following are guidelines for writing broad, qualitative research questions:

- *Ask one or two central questions followed by no more than five to seven subquestions.* Several subquestions follow each general central question; the subquestions narrow the focus of the study but leave open the questioning. This approach is well within the limits set by Miles and Huberman (1994), who recommended that researchers write no more than a dozen qualitative research questions in all (central and subquestions). The subquestions, in turn, can become specific questions used during interviews (or in observing or when looking at documents). In developing an interview protocol or guide, the researcher might ask an ice breaker question at the beginning, for example, followed by five or so subquestions in the study (see Chapter 9). The interview would then end with an additional wrap-up or summary question or ask, as I did in one of my qualitative case studies, "Who should I turn to, to learn more about this topic?" (Asmussen & Creswell, 1995).

- *Relate the central question to the specific qualitative strategy of inquiry.* For example, the specificity of the questions in ethnography at this stage of the design differs from that in other qualitative strategies. In ethnographic research, Spradley (1980) advanced a taxonomy of ethnographic questions that included a mini-tour of the culture-sharing group, their experiences, use of native language, contrasts with other cultural groups, and questions to verify the accuracy of the data. In critical ethnography, the research questions may build on a body of existing literature. These questions become working guidelines rather than truths to be proven (Thomas, 1993, p. 35). Alternatively, in phenomenology, the questions might be broadly stated without specific reference to the existing literature or a typology of questions. Moustakas (1994) talks about asking what the participants experienced and the contexts or situations in which they experienced it. A phenomenological example is, "What is it like for a mother to live with a teenage child who is dying of cancer?" (Nieswiadomy, 1993, p. 151). In grounded theory, the questions may be directed toward generating a theory of some process, such as the exploration of a process as to how caregivers and patients interact in a hospital setting. In a qualitative case study, the questions may address a description of the case and the themes that emerge from studying it.

- *Begin the research questions with the words* what *or* how *to convey an open and emerging design.* The word *why* often implies that the researcher is trying to explain why something occurs, and this suggests to me a cause-and-effect type of thinking that I associate with *quantitative* research instead of the more open and emerging stance of qualitative research.

- Focus on a single phenomenon or concept. As a study develops over time, factors will emerge that may influence this single phenomenon, but begin a study with a single focus to explore in great detail.

- Use exploratory verbs that convey the language of emerging design. These verbs tell the reader that the study will
    - Discover (e.g., grounded theory)
    - Seek to understand (e.g., ethnography)

- Explore a process (e.g., case study)
- Describe the experiences (e.g., phenomenology)
- Report the stories (e.g., narrative research)

- Use these more exploratory verbs that are nondirectional rather than directional words that suggest *quantitative* research, such as "affect," "influence," "impact," "determine," "cause," and "relate."

- Expect the research questions to evolve and change during the study in a manner consistent with the assumptions of an emerging design. Often in *qualitative* studies, the questions are under continual review and reformulation (as in a grounded theory study). This approach may be problematic for individuals accustomed to *quantitative* designs, in which the research questions remain fixed throughout the study.

- *Use open-ended questions* without reference to the literature or theory unless otherwise indicated by a qualitative strategy of inquiry.

- *Specify the participants and the research site* for the study, if the information has not yet been given.

Here is a script for a qualitative central question:

_________ (How or what) is the _________ ("story for" for narrative research; "meaning of" the phenomenon for phenomenology; "theory that explains the process of" for grounded theory; "culture-sharing pattern" for ethnography; "issue" in the "case" for case study) of _________ (central phenomenon) for _________ (participants) at _________ (research site).

The following are examples of qualitative research questions drawn from several types of strategies.

Example 7.1 *A Qualitative Central Question From an Ethnography*

Finders (1996) used ethnographic procedures to document the reading of teen magazines by middle-class European American seventh-grade girls. By examining the reading of teen zines (magazines), the researcher explored how the girls perceive and construct their social roles and relationships as they enter junior high school. She asked one guiding central question in her study:

> How do early adolescent females read literature that falls outside the realm of fiction?
>
> (Finders, 1996, p. 72)

Finders's (1996) central question begins with *how*; it uses an open-ended verb, *read*; it focuses on a single concept, the literature or teen magazines; and it mentions the participants, adolescent females, as the culture-sharing group. Notice how the author crafted a concise, single question that needed to be answered in the study. It is a broad question stated to permit participants to share diverse perspectives about reading the literature.

Example 7.2 *Qualitative Central Questions From a Case Study*

Padula and Miller (1999) conducted a multiple case study that described the experiences of women who went back to school, after a time away, in a psychology doctoral program at a major Midwestern research university. The intent was to document the women's experiences, providing a gendered and feminist perspective for women in the literature. The authors asked three central questions that guided the inquiry:

> (a) How do women in a psychology doctoral program describe their decision to return to school? (b) How do women in a psychology doctoral program describe their reentry experiences? And (c) How does returning to graduate school change these women's lives?
>
> (Padula & Miller, 1999, p. 328)

These three central questions all begin with the word *how*; they include open-ended verbs, such as "describe," and they focus on three aspects of the doctoral experience—returning to school, reentering, and changing. They also mention the participants as women in a doctoral program at a Midwestern research university.

QUANTITATIVE RESEARCH QUESTIONS AND HYPOTHESES

In quantitative studies, investigators use quantitative research questions and hypotheses, and sometimes objectives, to shape and specifically focus the purpose of the study. **Quantitative research questions** inquire about the relationships among variables that the investigator seeks to know. They are used frequently in social science research and especially in survey studies. **Quantitative hypotheses**, on the other hand, are predictions the researcher makes about the expected relationships among variables. They are numeric estimates of population values based on data collected from samples. Testing of hypotheses employs statistical procedures in which the investigator draws inferences about the population from a study sample. Hypotheses are used often in experiments in which investigators compare groups. Advisers often recommend their use in a formal research project, such as a dissertation or thesis, as a means of stating the direction a study will take. Objectives, on the other hand, indicate the goals or objectives for a study. They often appear in proposals for funding, but tend to be used with less frequency in social and health science research today. Because of this, the focus here will be on research questions and hypotheses. Here is an example of a script for a quantitative research question:

> Does ________ (name the theory) explain the relationship between ________ (independent variable) and ________ (dependent variable), controlling for the effects of ________ (control variable)?

Alternatively, a script for a quantitative null hypothesis might be as follows:

> There is no significant difference between ________ (the control and experimental groups on the independent variable) on ________ (dependent variable).

Guidelines for writing good quantitative research questions and hypotheses include the following.

- The use of variables in research questions or hypotheses is typically limited to three basic approaches. The researcher may *compare* groups on an independent variable to see its impact on a dependent variable. Alternatively, the investigator may *relate* one or more independent variables to one or more dependent variables. Third, the researcher may *describe* responses to the independent, mediating, or dependent variables. Most quantitative research falls into one or more of these three categories.

- The most rigorous form of quantitative research follows from a test of a theory (see Chapter 3) and the specification of research questions or hypotheses that are included in the theory.

- The independent and dependent variables must be measured separately. This procedure reinforces the cause-and-effect logic of quantitative research.

- To eliminate redundancy, write only research questions or hypotheses, not both, unless the hypotheses build on the research questions (discussion follows). Choose the form based on tradition, recommendations from an adviser or faculty committee, or whether past research indicates a prediction about outcomes.

- If hypotheses are used, there are two forms: null and alternative. A **null hypothesis** represents the traditional approach: it makes a prediction that in the general population, no relationship or no significant difference exists between groups on a variable. The wording is, "There is no difference (or relationship)" between the groups. The following example illustrates a null hypothesis.

Example 7.3 *A Null Hypothesis*

An investigator might examine three types of reinforcement for children with autism: verbal cues, a reward, and no reinforcement. The investigator collects behavioral measures assessing social interaction of the children with their siblings. A null hypothesis might read,

> There is no significant difference between the effects of verbal cues, rewards, and no reinforcement in terms of social interaction for children with autism and their siblings.

- The second form, popular in journal articles, is the alternative or **directional hypothesis**. The investigator makes a prediction about the expected outcome, basing this prediction on prior literature and studies on the topic that suggest a potential outcome. For example, the researcher may predict that "Scores will be higher for Group A than for Group B" on the dependent variable or that "Group A will change more than Group B" on the outcome. These examples illustrate a directional hypothesis because an expected prediction (e.g., higher, more change) is made. The following illustrates a directional hypothesis.

Example 7.4 *Directional Hypotheses*

Mascarenhas (1989) studied the differences between types of ownership (state-owned, publicly traded, and private) of firms in the offshore drilling industry. Specifically, the study explored such differences as domestic market dominance, international presence, and customer orientation. The study was a controlled field study using quasi-experimental procedures.

> Hypothesis 1: Publicly traded firms will have higher growth rates than privately held firms.

> Hypothesis 2: Publicly traded enterprises will have a larger international scope than state-owned and privately held firms.

Hypothesis 3: State-owned firms will have a greater share of the domestic market than publicly traded or privately held firms.

Hypothesis 4: Publicly traded firms will have broader product lines than state-owned and privately held firms.

Hypothesis 5: State-owned firms are more likely to have state-owned enterprises as customers overseas.

Hypothesis 6: State-owned firms will have a higher customer-base stability than privately held firms.

Hypothesis 7: In less visible contexts, publicly traded firms will employ more advanced technology than state-owned and privately held firms.

(Mascarenhas, 1989, pp. 585–588)

- Another type of alternative hypothesis is **nondirectional**—a prediction is made, but the exact form of differences (e.g., higher, lower, more, less) is not specified because the researcher does not know what can be predicted from past literature. Thus, the investigator might write, "There is a difference" between the two groups. An example follows which incorporates both types of hypotheses:

Example 7.5 *Nondirectional and Directional Hypotheses*

Sometimes directional hypotheses are created to examine the relationship among variables rather than to compare groups. For example, Moore (2000) studied the meaning of gender identity for religious and secular Jewish and Arab women in Israeli society. In a national probability sample of Jewish and Arab women, the author identified three hypotheses for study. The first is nondirectional and the last two are directional.

H_1: Gender identity of religious and secular Arab and Jewish women are related to different sociopolitical social orders that reflect the different value systems they embrace.

H_2: Religious women with salient gender identity are less socio-politically active than secular women with salient gender identities.

H_3: The relationships among gender identity, religiosity, and social actions are weaker among Arab women than among Jewish women.

- Unless the study intentionally employs demographic variables as predictors, use nondemographic variables (i.e., attitudes or behaviors) as independent and dependent variables. Because quantitative studies attempt to verify theories, demographic variables (e.g., age, income level, educational level, and so forth) typically enter these models as intervening (or mediating or moderating) variables instead of major independent variables.

- Use the same pattern of word order in the questions or hypotheses to enable a reader to easily identify the major variables. This calls for repeating key phrases and positioning the variables with the independent first and concluding with the dependent in left-to-right order (as discussed in Chapter 6 on good purpose statements). An example of word order with independent variables stated first in the phrase follows.

Example 7.6 *Standard Use of Language in Hypotheses*

1. There is no relationship between utilization of ancillary support services and academic persistence for non-traditional-aged women college students.
2. There is no relationship between family support systems and academic persistence for non-traditional-aged college women.
3. There is no relationship between ancillary support services and family support systems for non-traditional-aged college women.

A Model for Descriptive Questions and Hypotheses

Consider a model for writing questions or hypotheses based on writing descriptive questions (describing something) followed by inferential questions or hypotheses (drawing inferences from a sample to a population). These questions or hypotheses include both independent and dependent variables. In this model, the writer specifies descriptive questions for *each* independent and dependent variable and important intervening or moderating variables. Inferential questions (or hypotheses) that relate variables or compare groups follow these descriptive questions. A final set of questions may add inferential questions or hypotheses in which variables are controlled.

Example 7.7 *Descriptive and Inferential Questions*

To illustrate this approach, a researcher wants to examine the relationship of critical thinking skills (an independent variable measured on an instrument) to student achievement (a dependent variable measured by grades) in science classes for eighth-grade students in a large metropolitan school district. The researcher controls for the intervening effects of prior grades in science classes and parents' educational attainment. Following the proposed model, the research questions might be written as follows:

Descriptive Questions

1. How do the students rate on critical thinking skills? (A descriptive question focused on the independent variable)
2. What are the student's achievement levels (or grades) in science classes? (A descriptive question focused on the dependent variable)
3. What are the student's prior grades in science classes? (A descriptive question focused on the control variable of prior grades)
4. What is the educational attainment of the parents of the eighth-graders? (A descriptive question focused on another control variable, educational attainment of parents)

Inferential Questions

1. Does critical thinking ability relate to student achievement? (An inferential question relating the independent and the dependent variables)
2. Does critical thinking ability relate to student achievement, controlling for the effects of prior grades in science and the educational attainment of the eighth-graders' parents? (An inferential question relating the independent and the dependent variables, controlling for the effects of the two controlled variables)

This example illustrates how to organize all the research questions into descriptive and inferential questions. In another example, a researcher may want to compare groups, and the language may change to reflect this comparison in the inferential questions. In other studies, many more independent and dependent variables may be present in the model being tested, and a longer list of descriptive and inferential questions would result. I recommend this descriptive–inferential model.

This example also illustrates the use of variables to describe as well as relate. It specifies the independent variables in the first position in the questions, the dependent in the second, and the control variables in the third. It employs demographics as controls rather than central variables in the questions, and a reader needs to assume that the questions flow from a theoretical model.

MIXED METHODS RESEARCH QUESTIONS AND HYPOTHESES

In discussions about methods, researchers typically do not see specific questions or hypotheses especially tailored to mixed methods research. However, discussion has begun concerning the use of mixed methods questions in studies and also how to design them (see Creswell & Plano Clark, 2007; Tashakkori & Creswell, 2007). A strong mixed methods study should start with a mixed methods research question, to shape the methods and the overall design of a study. Because a mixed methods study relies on neither quantitative or qualitative research alone, some combination of the two provides the best information for the research questions and hypotheses. To be considered are what types of questions should be presented and when and what information is most needed to convey the nature of the study:

- Both qualitative and quantitative research questions (or hypotheses) need to be advanced in a mixed methods study in order to narrow and focus the purpose statement. These questions or hypotheses can be advanced at the beginning or when they emerge during a later phase of the research. For example, if the study begins with a quantitative phase, the investigator might introduce hypotheses. Later in the study, when the qualitative phase is addressed, the qualitative research questions appear.

- When writing these questions or hypotheses, follow the guidelines in this chapter for scripting good questions or hypotheses.

- Some attention should be given to the order of the research questions and hypotheses. In a two-phase project, the first-phase questions would come first, followed by the second-phase questions so that readers see them in the order in which they will be addressed in the proposed study. In a single-phase strategy of inquiry, the questions might be ordered according to the method that is given the most weight in the design.

- Include a **mixed methods research question** that directly addresses the mixing of the quantitative and qualitative strands of the research. This is the question that will be answered in the study based on the mixing (see Creswell & Plano Clark, 2007). This is a new form of question in research methods, and Tashakkori and Creswell (2007, p. 208) call it a "hybrid" or "integrated" question. This question could either be written at the beginning or when it emerges; for instance, in a two-phase study in which one phase builds on the other, the mixed methods questions might be placed in a discussion between the two phases. This can assume one of two forms. The first is to write it in a way that conveys the *methods* or *procedures* in a study (e.g., Does the qualitative data help explain the results from the initial quantitative phase of the study? See

Creswell & Plano Clark, 2007). The second form is to write it in a way that conveys the *content* of the study (e.g., Does the theme of social support help to explain why some students become bullies in schools? (see Tashakkori & Creswell, 2007.)

● Consider several different ways that all types of research questions (i.e., quantitative, qualitative, and mixed) can be written into a mixed methods study:

- Write separate quantitative questions or hypotheses and qualitative questions. These could be written at the beginning of a study or when they appear in the project if the study unfolds in stages or phases. With this approach, the emphasis is placed on the two approaches and not on the mixed methods or integrative component of the study.
- Write separate quantitative questions or hypotheses and qualitative questions and follow them with a mixed methods question. This highlights the importance of both the qualitative and quantitative phases of the study as well as their combined strength, and thus is probably the ideal approach.
- Write only a mixed methods question that reflects the *procedures* or the *content* (or write the mixed methods question in both a procedural and a content approach), and do not include separate quantitative and qualitative questions. This approach would enhance the viewpoint that the study intends to lead to some integration or connection between the quantitative and qualitative phases of the study (i.e., the sum of both parts is greater than each part).

Example 7.8 *Hypotheses and Research Questions in a Mixed Methods Study*

Houtz (1995) provides an example of a two-phase study with the separate quantitative and qualitative research hypotheses and questions stated in sections introducing each phase. She did not use a separate, distinct mixed methods research question. Her study investigated the differences between middle-school (nontraditional) and junior high (traditional) instructional strategies for seventh-grade and eighth-grade students and their attitudes toward science and their science achievement. Her study was conducted at a point when many schools were moving away from the two-year junior high concept to the three-year middle school (including sixth grade) approach to education. In this two-phase study, the first phase involved assessing pre-test

(Continued)

(Continued)

and post-test attitudes and achievement using scales and examination scores. Houtz then followed the quantitative results with qualitative interviews with science teachers, the school principal, and consultants. This second phase helped to explain differences and similarities in the two instructional approaches obtained in the first phase.

With a first-phase quantitative study, Houtz (1995, p. 630) mentioned the hypotheses guiding her research:

> It was hypothesized that there would be no significant difference between students in the middle school and those in the junior high in attitude toward science as a school subject. It was also hypothesized that there would be no significant difference between students in the middle school and those in the junior high in achievement in science.

These hypotheses appeared at the beginning of the study as an introduction to the quantitative phase. Prior to the qualitative phase, Houtz raised questions to explore the quantitative results in more depth. Focusing in on the achievement test results, she interviewed science teachers, the principal, and the university consultants and asked three questions:

> What differences currently exist between the middle school instructional strategy and the junior high instructional strategy at this school in transition? How has this transition period impacted science attitude and achievement of your students? How do teachers feel about this change process?
>
> (Houtz, 1995, p. 649)

Examining this mixed methods study shows that the author included both quantitative and qualitative questions, specified them at the beginning of each phase of her study, and used good elements for writing both quantitative hypotheses and qualitative research questions. Had Houtz (1995) developed a mixed methods question, it might have been stated from a *procedural* perspective:

> How do the interviews with teachers, the principal, and university consultants help to explain any quantitative differences in achievement for middle-school and junior high students?

Alternatively, the mixed methods question might have been written from a *content* orientation, such as:

> How do the themes mentioned by the teachers help to explain why middle-school children score lower than the junior high students?

Example 7.9 *A Mixed Methods Question Written in Terms of Mixing Procedures*

To what extent and in what ways do qualitative interviews with students and faculty members serve to contribute to a more comprehensive and nuanced understanding of this predicting relationship between CEEPT scores and student academic performance, via integrative mixed methods analysis?

(Lee & Greene, 2007)

This is a good example of a mixed methods question focused on the intent of mixing, to integrate the qualitative interviews and the quantitative data, the relationship of scores and student performance. This question emphasized what the integration was attempting to accomplish—a comprehensive and nuanced understanding—and at the end of the article, the authors presented evidence answering this question.

SUMMARY

Research questions and hypotheses narrow the purpose statement and become major signposts for readers. Qualitative researchers ask at least one central question and several subquestions. They begin the questions with words such as *how* or *what* and use exploratory verbs, such as *explore* or *describe*. They pose broad, general questions to allow the participants to explain their ideas. They also focus initially on one central phenomenon of interest. The questions may also mention the participants and the site for the research.

Quantitative researchers write either research questions or hypotheses. Both forms include variables that are described, related, categorized into groups for comparison, and the independent and dependent variables are measured separately. In many quantitative proposals, writers use research questions; however, a more formal statement of research employs hypotheses. These hypotheses are predictions about the outcomes of the results, and they may be written as alternative hypotheses specifying the exact results to be expected (more or less, higher or lower of something). They also may be stated in the null form, indicating no expected difference or no relationship between groups on a dependent variable. Typically, the researcher writes the independent variable(s) first, followed by the dependent variable(s). One model for ordering the questions in a quantitative proposal is to begin with descriptive questions followed by the inferential questions that relate variables or compare groups.

I encourage mixed methods researchers to construct separate mixed methods questions in their studies. This question might be written to emphasize the procedures or the content of the study, and it might be placed at different points. By writing this question, the researcher conveys the importance of integrating or combining the quantitative and qualitative elements. Several models exist for writing mixed methods questions into studies: writing only quantitative questions or hypotheses and qualitative questions, or writing both quantitative questions or hypotheses and qualitative questions followed by a mixed methods question, or writing only a mixed methods question.

WRITING EXERCISES

Writing Exercises

1. For a qualitative study, write one or two central questions followed by five to seven subquestions.

2. For a quantitative study, write two sets of questions. The first set should be descriptive questions about the independent and dependent variables in the study. The second set should pose questions that relate (or compare) the independent variable(s) with the dependent variable(s). This follows the model presented in this chapter for combining descriptive and inferential questions.

3. Write a mixed methods research question. Write it first as a question incorporating the procedures of your mixed methods study and then rewrite it to incorporate the content. Comment on which approach works best for you.

ADDITIONAL READINGS

Creswell, J. W. (1999). Mixed-method research: Introduction and application. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 455–472). San Diego: Academic Press.

In this chapter, I discuss the nine steps in conducting a mixed methods study. These are as follows:

1. Determine if a mixed methods study is needed to study the problem;
2. Consider whether a mixed methods study is feasible;
3. Write both qualitative and quantitative research questions;

4. Review and decide on the types of data collection;
5. Assess the relative weight and implementation strategy for each method;
6. Present a visual model;
7. Determine how the data will be analyzed;
8. Assess the criteria for evaluating the study; and
9. Develop a plan for the study.

In writing the research questions, I recommend developing both qualitative and quantitative types and stating within them the type of qualitative strategy of inquiry being used.

Tashakkori, A., & Creswell, J. W. (2007). Exploring the nature of research questions in mixed methods research. Editorial. *Journal of Mixed Methods Research, 1*(3), 207–211.

This editorial addresses the use and nature of research questions in mixed methods research. It highlights the importance of research questions in the process of research and discusses the need for a better understanding of the use of mixed methods questions. It asks, "How does one frame a research question in a mixed methods study?" (p. 207). Three models are presented: writing separate quantitative and qualitative questions, writing an overarching mixed methods question, or writing research questions for each phase of a study as the research evolves.

Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220–235). Thousand Oaks, CA: Sage.

Janice Morse, a nursing researcher, identifies and describes the major design issues involved in planning a qualitative project. She compares several strategies of inquiry and maps the type of research questions used in each strategy. For phenomenology and ethnography, the research calls for meaning and descriptive questions. For grounded theory, the questions need to address process, whereas in ethnomethodology and discourse analysis, the questions relate to verbal interaction and dialogue. She indicates that the wording of the research question determines the focus and scope of the study.

Tuckman, B. W. (1999). *Conducting educational research* (5th ed.). Fort Worth, TX: Harcourt Brace.

Bruce Tuckman provides an entire chapter on constructing hypotheses. He identifies the origin of hypotheses in deductive theoretical positions and in inductive observations. He further defines and illustrates both alternative and null hypotheses and takes the reader through the hypothesis testing procedure.

Creswell, J. W. 2009. Research Design: Qualitative, Quantitative, and Mixed Methods. Ch 8: Quantitative Methods

This page intentionally left blank

CHAPTER EIGHT

Quantitative Methods

For many proposal writers, the method section is the most concrete, specific part of a proposal. This chapter presents essential steps in designing quantitative methods for a research proposal or study, with specific focus on survey and experimental designs. These designs reflect postpositivist philosophical assumptions, as discussed in Chapter 1. For example, determinism suggests that examining the relationships between and among variables is central to answering questions and hypotheses through surveys and experiments. The reduction to a parsimonious set of variables, tightly controlled through design or statistical analysis, provides measures or observations for testing a theory. Objective data result from empirical observations and measures. Validity and reliability of scores on instruments lead to meaningful interpretations of data.

In relating these assumptions and the procedures that implement them, this discussion does not exhaustively treat quantitative research methods. Excellent, detailed texts provide information about survey research (e.g., see Babbie, 1990, 2007; Fink, 2002; Salant & Dillman, 1994). For experimental procedures, some traditional books (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979), as well as some newer texts, extend the ideas presented here (e.g., Bausell, 1994; Boruch, 1998; Field & Hole, 2003; Keppel, 1991; Lipsey, 1990; Reichardt & Mark, 1998). In this chapter, the focus is on the essential components of a method section in proposals for a survey and an experiment.

DEFINING SURVEYS AND EXPERIMENTS

A **survey design** provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population. From sample results, the researcher generalizes or makes claims about the population. In an *experiment*, investigators may also identify a sample and generalize to a population; however, the basic intent of an **experimental design** is to test the impact of a treatment (or an intervention) on an outcome, controlling for all other factors that might influence that outcome. As one form of control, researchers randomly assign individuals to groups. When one group receives a treatment and the other group does not, the experimenter can isolate whether it is the treatment and not other factors that influence the outcome.

COMPONENTS OF A SURVEY METHOD PLAN

The design of a survey method section follows a standard format. Numerous examples of this format appear in scholarly journals, and these examples provide useful models. The following sections detail typical components. In preparing to design these components into a proposal, consider the questions on the checklist shown in Table 8.1 as a general guide.

The Survey Design

In a proposal or plan, one of the first parts of the method section can introduce readers to the basic purpose and rationale for survey research. Begin the discussion by reviewing the purpose of a survey and the rationale for its selection for the proposed study. This discussion can

- Identify the purpose of survey research. This purpose is to generalize from a sample to a population so that inferences can be made about some characteristic, attitude, or behavior of this population (Babbie, 1990). Provide a reference to this purpose from one of the survey method texts (several are identified in this chapter).

- Indicate why a survey is the preferred type of data collection procedure for the study. In this rationale, consider the advantages of survey designs, such as the economy of the design and the rapid turnaround in data collection. Discuss the advantage of identifying attributes of a large population from a small group of individuals (Babbie, 1990; Fowler, 2002).

- Indicate whether the survey will be cross-sectional, with the data collected at one point in time, or whether it will be longitudinal, with data collected over time.

- Specify the form of data collection. Fink (2002) identifies four types: self-administered questionnaires; interviews; structured record reviews to collect financial, medical, or school information; and structured observations. The data collection may also involve creating a Web-based or Internet survey and administering it online (Nesbary, 2000; Sue & Ritter, 2007). Regardless of the form of data collection, provide a rationale for the

Table 8.1 A Checklist of Questions for Designing a Survey Method	
________	Is the purpose of a survey design stated?
________	Are the reasons for choosing the design mentioned?
________	Is the nature of the survey (cross-sectional vs. longitudinal) identified?
________	Are the population and its size mentioned?
________	Will the population be stratified? If so, how?
________	How many people will be in the sample? On what basis was this size chosen?
________	What will be the procedure for sampling these individuals (e.g., random, nonrandom)?
________	What instrument will be used in the survey? Who developed the instrument?
________	What are the content areas addressed in the survey? The scales?
________	What procedure will be used to pilot or field test the survey?
________	What is the timeline for administering the survey?
________	What are the variables in the study?
________	How do these variables cross-reference with the research questions and items on the survey?
	What specific steps will be taken in data analysis to
(a)______	analyze returns?
(b)______	check for response bias?
(c)______	conduct a descriptive analysis?
(d)______	collapse items into scales?
(e)______	check for reliability of scales?
(f)______	run inferential statistics to answer the research questions?
________	How will the results be interpreted?

procedure, using arguments based on its strengths and weaknesses, costs, data availability, and convenience.

The Population and Sample

Specify the characteristics of the population and the sampling procedure. Methodologists have written excellent discussions about the underlying logic of sampling theory (e.g., Babbie, 1990, 2007). Here are essential aspects of the population and sample to describe in a research plan:

- Identify the population in the study. Also state the size of this population, if size can be determined, and the means of identifying individuals in the population. Questions of access arise here, and the researcher might refer to availability of sampling frames—mail or published lists—of potential respondents in the population.

- Identify whether the sampling design for this population is single stage or multistage (called clustering). Cluster sampling is ideal when it is impossible or impractical to compile a list of the elements composing the population (Babbie, 2007). A single-stage sampling procedure is one in which the researcher has access to names in the population and can sample the people (or other elements) directly. In a multistage or clustering procedure, the researcher first identifies clusters (groups or organizations), obtains names of individuals within those clusters, and then samples within them.

- Identify the selection process for individuals. I recommend selecting a *random* sample, in which each individual in the population has an equal probability of being selected (a systematic or probabilistic sample). Less desirable is a nonprobability sample (or convenience sample), in which respondents are chosen based on their convenience and availability (Babbie, 1990). With randomization, a representative sample from a population provides the ability to generalize to a population.

- Identify whether the study will involve stratification of the population before selecting the sample. *Stratification* means that specific characteristics of individuals (e.g., both females and males) are represented in the sample and the sample reflects the true proportion in the population of individuals with certain characteristics (Fowler, 2002). When randomly selecting people from a population, these characteristics may or may not be present in the sample in the same proportions as in the population; stratification ensures their representation. Also identify the characteristics used in stratifying the population (e.g., gender, income levels, education). Within each stratum, identify whether the sample contains individuals with the characteristic in the same proportion as the characteristic appears in the entire population (Babbie, 1990; Miller, 1991).

- Discuss the procedures for selecting the sample from available lists. The most rigorous method for selecting the sample is to choose individuals using a random numbers table, a table available in many introductory statistics texts (e.g., Gravetter & Wallnau, 2000).

- Indicate the number of people in the sample and the procedures used to compute this number. In survey research, I recommend that one use a sample size formula available in many survey texts (e.g., see Babbie, 1990; Fowler, 2002).

Instrumentation

As part of rigorous data collection, the proposal developer also provides detailed information about the actual survey instrument to be used in the proposed study. Consider the following:

- Name the survey instrument used to collect data. Discuss whether it is an instrument designed for this research, a modified instrument, or an intact instrument developed by someone else. If it is a modified instrument, indicate whether the developer has provided appropriate permission to use it. In some survey projects, the researcher assembles an instrument from components of several instruments. Again, permission to use any part of other instruments needs to be obtained. In addition, instruments are being increasingly designed for online surveys (see Sue & Ritter, 2007). An online survey tool is SurveyMonkey (SurveyMonkey.com), a commercial product available since 1999. Using this service, researchers can create their own surveys quickly using custom templates and post them on Web sites or e-mail them for participants to complete. SurveyMonkey then can generate results and report them back to the researcher as descriptive statistics or as graphed information. The results can be downloaded into a spreadsheet or a database for further analysis. The basic program is free for 100 responses per survey and no more than 10 questions per survey. For additional responses, more questions, and several custom features, SurveyMonkey charges a monthly or annual fee.

- To use an existing instrument, describe the established validity and reliability of scores obtained from past use of the instrument. This means reporting efforts by authors to establish **validity**—whether one can draw meaningful and useful inferences from scores on the instruments. The three traditional forms of validity to look for are content validity (do the items measure the content they were intended to measure?), predictive or concurrent validity (do scores predict a criterion measure? Do results correlate with other results?), and construct validity (do items measure hypothetical constructs or concepts?). In more recent studies, construct validity has also included whether the scores serve a useful purpose and have positive consequences when they are used in practice (Humbley & Zumbo, 1996). Establishing the validity of the scores in a survey helps to identify whether an instrument might be a good one to use in survey research. This form of validity is different than identifying the threats to validity in experimental research, as discussed later in this chapter.

Also discuss whether scores resulting from past use of the instrument demonstrate **reliability**. Look for whether authors report measures of internal consistency (are the items' responses consistent across constructs?)

and test–retest correlations (are scores stable over time when the instrument is administered a second time?). Also determine whether there was consistency in test administration and scoring (were errors caused by carelessness in administration or scoring?; Borg, Gall, & Gall, 1993).

- When one modifies an instrument or combines instruments in a study, the original validity and reliability may not hold for the new instrument, and it becomes important to reestablish validity and reliability during data analysis.

- Include sample items from the instrument so that readers can see the actual items used. In an appendix to the proposal, attach sample items or the entire instrument.

- Indicate the major content sections in the instrument, such as the cover letter (Dillman, 1978, provides a useful list of items to include in cover letters), the items (e.g., demographics, attitudinal items, behavioral items, factual items), and the closing instructions. Also mention the type of scales used to measure the items on the instrument, such as continuous scales (e.g., *strongly agree* to *strongly disagree*) and categorical scales (e.g., yes/no, rank from highest to lowest importance).

- Discuss plans for pilot testing or field testing the survey and provide a rationale for these plans. This testing is important to establish the content validity of an instrument and to improve questions, format, and scales. Indicate the number of people who will test the instrument and the plans to incorporate their comments into final instrument revisions.

- For a mailed survey, identify steps for administering the survey and for following up to ensure a high response rate. Salant and Dillman (1994) suggest a four-phase administration process. The first mail-out is a short advance-notice letter to all members of the sample, and the second mail-out is the actual mail survey, distributed about 1 week after the advance-notice letter. The third mail-out consists of a postcard follow-up sent to all members of the sample 4 to 8 days after the initial questionnaire. The fourth mail-out, sent to all nonrespondents, consists of a personalized cover letter with a handwritten signature, the questionnaire, and a preaddressed return envelope with postage. Researchers send this fourth mail-out 3 weeks after the second mail-out. Thus, in total, the researcher concludes the administration period 4 weeks after its start, providing the returns meet project objectives.

Variables in the Study

Although readers of a proposal learn about the variables in purpose statements and research questions/hypotheses sections, it is useful in the method section to relate the variables to the specific questions or hypotheses on the instrument. One technique is to relate the variables, the research questions or hypotheses, and items on the survey instrument so that a reader can easily determine how the researcher will use the questionnaire items. Plan to include a table and a discussion that cross-reference the variables, the questions or hypotheses, and specific survey items. This procedure is especially helpful in dissertations in which investigators test large-scale models. Table 8.2 illustrates such a table using hypothetical data.

Table 8.2 Variables, Research Questions, and Items on a Survey

Variable Name	Research Question	Item on Survey
Independent Variable 1: Prior publications	Descriptive research Question 1: How many publications did the faculty member produce prior to receipt of the doctorate?	See Questions 11, 12, 13, 14, and 15: publication counts for journal articles, books, conference papers, book chapters published before receiving the doctorate
Dependent Variable 1: Grants funded	Descriptive research Question 3: How many grants has the faculty member received in the past 3 years?	See Questions 16, 17, and 18: grants from foundations, federal grants, state grants
Control Variable 1: Tenure status	Descriptive research Question 5: Is the faculty member tenured?	See Question 19: tenured (yes/no)

Data Analysis and Interpretation

In the proposal, present information about the steps involved in analyzing the data. I recommend the following **research tips**, presenting them as a series of steps so that a reader can see how one step leads to another for a complete discussion of the data analysis procedures.

Step 1. Report information about the number of members of the sample who did and did not return the survey. A table with numbers and percentages describing respondents and nonrespondents is a useful tool to present this information.

Step 2. Discuss the method by which response bias will be determined. **Response bias** is the effect of nonresponses on survey estimates (Fowler, 2002). *Bias* means that if nonrespondents had responded, their responses would have substantially changed the overall results. Mention the procedures used to check for response bias, such as wave analysis or a respondent/nonrespondent analysis. In wave analysis, the researcher examines returns on select items week by week to determine if average responses change (Leslie, 1972). Based on the assumption that those who return surveys in the final weeks of the response period are nearly all nonrespondents, if the responses begin to change, a potential exists for response bias. An alternative check for response bias is to contact a few nonrespondents by phone and determine if their responses differ substantially from respondents. This constitutes a respondent-nonrespondent check for response bias.

Step 3. Discuss a plan to provide a **descriptive analysis** of data for all independent and dependent variables in the study. This analysis should indicate the means, standard deviations, and range of scores for these variables.

Step 4. If the proposal contains an instrument with scales or a plan to develop scales (combining items into scales), identify the statistical procedure (i.e., factor analysis) for accomplishing this. Also mention reliability checks for the internal consistency of the scales (i.e., the Cronbach alpha statistic).

Step 5. Identify the statistics and the statistical computer program for testing the major inferential research questions or hypotheses in the proposed study. The **inferential questions** or **hypotheses** relate variables or compare groups in terms of variables so that inferences can be drawn from the sample to a population. Provide a rationale for the choice of statistical test and mention the assumptions associated with the statistic. As shown in Table 8.3, base this choice on the nature of the research question (e.g., relating variables or comparing groups as the most popular), the number of independent and dependent variables, and the number of variables controlled (e.g., see Rudestam & Newton, 2007). Further, consider whether the variables will be measured on an instrument as a continuous score (e.g., age, from 18 to 36) or as a categorical score (e.g., women = 1, men = 2). Finally, consider whether the scores from the sample might be normally distributed in a bell-shaped curve if plotted out on a graph or non-normally distributed. There are additional ways to determine if the scores are normally distributed (see Creswell, 2008). These factors, in combination, enable a researcher to determine what statistical test will be suited for answering the research question or hypothesis. In Table 8.3, I show how the factors, in combination, lead to the selection of a number of common statistical tests. For further types of statistical tests, readers are referred to statistics methods books, such as Gravetter and Wallnau (2000).

Step 6. A final step in the data analysis is to present the results in tables or figures and interpret the results from the statistical test. An **interpretation of the results** means that the researcher draws conclusions from the results for the research questions, hypotheses, and the larger meaning of the results. This interpretation involves several steps.

Table 8.3 Criteria for Choosing Select Statistical Tests

Nature of Question	Number of Independent Variables	Number of Dependent Variables	Number of Control Variables (covariates)	Type of Score Independent/ Dependent Variables	Distribution of Scores	Statistical Test
Group comparison	1	1	0	Categorical/ continuous	Normal	*t*-test
Group comparison	1 or more	1	0	Categorical/ continuous	Normal	Analysis of variance
Group comparison	1 or more	1	1	Categorical/ continuous	Normal	Analysis of covariance
Group comparison	1	1	0	Categorical/ continuous	Non-normal	Mann-Whitney U test
Association between groups	1	1	0	Categorical/ categorical	Non-normal	Chi-square
Relate variables	1	1	0	Continuous/ continuous	Normal	Pearson product moment correlation
Relate variables	2 or more	1	0	Continuous/ continuous	Normal	Multiple regression
Relate variables	1	1 or more	0	Categorical/ categorical	Non-normal	Spearman rank-order correlation

- Report whether the results of the statistical test were statistically significant or not, such as "the analysis of variance revealed a statistically significant difference between men and women in terms of attitudes toward banning smoking in restaurants $F(2; 6) = 8.55, p = .001$."
- Report how these results answered the research question or hypothesis. Did the results support the hypothesis or did they contradict what was expected?
- Indicate what might explain why the results occurred. This explanation might refer back to the theory advanced in the proposed study (see Chapter 3), past literature as reviewed in the literature review (see Chapter 2), or logical reasoning.
- Discuss the implications of the results for practice or for future research on the topic.

Example 8.1 *A Survey Method Section*

An example follows of a survey method section that illustrates many of the steps just mentioned. This excerpt (used with permission) comes from a journal article reporting a study of factors affecting student attrition in one small liberal arts college (Bean & Creswell, 1980, pp. 321–322).

Methodology

The site of this study was a small (enrollment 1,000), religious, coeducational, liberal arts college in a Midwestern city with a population of 175,000 people. *(Authors identified the research site and population.)*

The dropout rate the previous year was 25%. Dropout rates tend to be highest among freshmen and sophomores, so an attempt was made to reach as many freshmen and sophomores as possible by distribution of the questionnaire through classes. Research on attrition indicates that males and females drop out of college for different reasons (Bean, 1978, in press; Spady, 1971). Therefore, only women were analyzed in this study.

During April 1979, 169 women returned questionnaires. A homogeneous sample of 135 women who were 25 years old or younger, unmarried, full-time U.S. citizens, and Caucasian was selected for this analysis to exclude some possible confounding variables (Kerlinger, 1973).

Of these women, 71 were freshmen, 55 were sophomores, and 9 were juniors. Of the students, 95% were between the ages of 18 and 21. This sample is biased toward higher-ability students as indicated by scores on the ACT test. *(Authors presented descriptive information about the sample.)*

Data were collected by means of a questionnaire containing 116 items. The majority of these were Likert-like items based on a scale from "a very small extent" to "a very great extent." Other questions asked for factual information, such as ACT scores, high school grades, and parents' educational level. All information used in this analysis was derived from questionnaire data. This questionnaire had been developed and tested at three other institutions before its use at this college. *(Authors discussed the instrument.)*

Concurrent and convergent validity (Campbell & Fiske, 1959) of these measures was established through factor analysis, and was found to be at an adequate level. Reliability of the factors was established through the coefficient alpha. The constructs were represented by 25 measures—multiple items combined on the basis of factor analysis to make indices—and 27 measures were single item indicators. *(Validity and reliability were addressed.)*

Multiple regression and path analysis (Heise, 1969; Kerlinger & Pedhazur, 1973) were used to analyze the data.

In the causal model . . . , intent to leave was regressed on all variables which preceded it in the causal sequence. Intervening variables significantly related to intent to leave were then regressed on organizational variables, personal variables, environmental variables, and background variables. *(Data analysis steps were presented.)*

COMPONENTS OF AN EXPERIMENTAL METHOD PLAN

An experimental method discussion follows a standard form: participants, materials, procedures, and measures. These four topics generally are sufficient. In this section of the chapter, I review these components as well as information about the experimental design and statistical analysis. As with the section on surveys, the intent here is to highlight key topics to be addressed in an experimental method proposal. An overall guide to these topics is found by answering the questions on the checklist shown in Table 8.4.

Participants

Readers need to know about the selection, assignment, and number of participants who will take part in the experiment. Consider the following suggestions when writing the method section for an experiment:

- Describe the selection process for participants as either random or nonrandom (e.g., conveniently selected). The participants might be selected by *random selection* or *random sampling.* With random selection or **random sampling**, each individual has an equal probability of being selected from the population, ensuring that the sample will be representative of the population (Keppel, 1991). In many experiments, however, only a *convenience* sample is possible because the investigator must use naturally formed groups (e.g., a classroom, an organization, a family unit) or volunteers. When individuals are not randomly assigned, the procedure is called a **quasi-experiment.**

- When individuals can be randomly assigned to groups, the procedure is called a **true experiment.** If random assignment is made, discuss how the project will *randomly assign* individuals to the treatment groups. This means that of the pool of participants, Individual 1 goes to Group 1, Individual 2 to Group 2, and so forth, so that there is no systematic bias in assigning the individuals. This procedure eliminates the possibility of systematic differences among characteristics of the participants that could affect the outcomes, so that any differences in outcomes can be attributed to the experimental treatment (Keppel, 1991).

- Identify other features in the experimental design that will systematically control the variables that might influence the outcome. One approach is **matching participants** in terms of a certain trait or characteristic and then assigning one individual from each matched set to each group. For example, scores on a pre-test might be obtained. Individuals might then be assigned to groups, with each group having the same numbers of high, medium, and low scorers on the pre-test. Alternatively, the criteria for matching might be ability levels or demographic variables.

Table 8.4 A Checklist of Questions for Designing an Experimental Procedure

_______	Who are the participants in the study?
_______	What is the population to which the results of the participants will be generalized?
_______	How were the participants selected? Was a random selection method used?
_______	How will the participants be randomly assigned? Will they be matched? How?
_______	How many participants will be in the experimental and control group(s)?
_______	What is the dependent variable or variables (i.e., outcome variable) in the study? How will it be measured? Will it be measured before and after the experiment?
_______	What is the treatment condition(s)? How was it operationalized?
_______	Will variables be covaried in the experiment? How will they be measured?
_______	What experimental research design will be used? What would a visual model of this design look like?
_______	What instrument(s) will be used to measure the outcome in the study? Why was it chosen? Who developed it? Does it have established validity and reliability? Has permission been sought to use it?
_______	What are the steps in the procedure (e.g., random assignment of participants to groups, collection of demographic information, administration of pretest, administration of treatment(s), administration of posttest)?
_______	What are potential threats to internal and external validity for the experimental design and procedure? How will they be addressed?
_______	Will a pilot test of the experiment be conducted?
_______	What statistics will be used to analyze the data (e.g., descriptive and inferential)?
_______	How will the results be interpreted?

A researcher may decide not to match, however, because it is expensive, takes time (Salkind, 1990), and leads to incomparable groups if participants leave the experiment (Rosenthal & Rosnow, 1991). Other procedures to place control into experiments involve using covariates (e.g., pre-test scores) as moderating variables and controlling for their effects statistically, selecting homogeneous samples, or blocking the participants into subgroups or categories and analyzing the impact of each subgroup on the outcome (Creswell, 2008).

- Tell the reader about the number of participants in each group and the systematic procedures for determining the size of each group. For experimental research, investigators use a power analysis (Lipsey, 1990) to identify the appropriate sample size for groups. This calculation involves

  - A consideration of the level of statistical significance for the experiment, or alpha
  - The amount of power desired in a study—typically presented as high, medium, or low—for the statistical test of the null hypothesis with sample data when the null hypothesis is, in fact, false
  - The effect size, the expected differences in the means between the control and experimental groups expressed in standard deviation units

- Researchers set values for these three factors (e.g., alpha = .05, power = .80, and effect size = .50) and can look up in a table the size needed for each group (see Cohen, 1977; Lipsey, 1990). In this way, the experiment is planned so that the size of each treatment group provides the greatest sensitivity that the effect on the outcome actually is due to the experimental manipulation in the study.

Variables

The variables need to be specified in an experiment so that it is clear to readers what groups are receiving the experimental treatment and what outcomes are being measured. Here are some suggestions for developing ideas about variables in a proposal:

- Clearly identify the *independent variables* in the experiment (recall the discussion of variables in Chapter 3). One independent variable must be the *treatment variable.* One or more groups receive the experimental manipulation, or treatment, from the researcher. Other independent variables may simply be measured variables in which no manipulation occurs (e.g., attitudes or personal characteristics of participants). Still other independent variables can be statistically controlled, such as demographics (e.g., gender or age). The method section must list and clearly identify all the independent variables in an experiment.

- Identify the *dependent variable or variables* (i.e., the outcomes) in the experiment. The dependent variable is the response or the criterion variable that is presumed to be caused by or influenced by the independent treatment conditions and any other independent variables). Rosenthal and

Rosnow (1991) advanced three prototypic outcomes measures: the direction of observed change, the amount of this change, and the ease with which the participant changes (e.g., the participant reacquires the correct response as in a single-subject design).

Instrumentation and Materials

During an experiment, one makes observations or obtains measures using instruments at a pre-test or post-test (or both) stage of the procedures. A sound research plan calls for a thorough discussion about the instrument or instruments—their development, their items, their scales, and reports of reliability and validity of scores on past uses. The researcher also should report on the materials used for the experimental treatment (e.g., the special program or specific activities given to the experimental group).

- Describe the instrument or instruments participants complete in the experiment, typically completed before the experiment begins and at its end. Indicate the established validity and reliability of the scores on instruments, the individuals who developed them, and any permissions needed to use them.

- Thoroughly discuss the materials used for the experimental treatment. One group, for example, may participate in a special computer-assisted learning plan used by a teacher in a classroom. This plan might involve handouts, lessons, and special written instructions to help students in this experimental group learn how to study a subject using computers. A pilot test of these materials may also be discussed, as well as any training required to administer the materials in a standard way. The intent of this pilot test is to ensure that materials can be administered without variability to the experimental group.

Experimental Procedures

The specific experimental design procedures also need to be identified. This discussion involves indicating the overall experiment type, citing reasons for the design, and advancing a visual model to help the reader understand the procedures.

- Identify the type of experimental design to be used in the proposed study. The types available in experiments are pre-experimental designs, true experiments, quasi-experiments, and single-subject designs. With *pre-experimental* designs, the researcher studies a single group and provides an intervention during the experiment. This design does not have a control group to compare with the experimental group. In *quasi-experiments*, the investigator uses control and experimental groups but does not randomly assign participants to groups (e.g., they may be intact groups available to the researcher). In a *true experiment,* the investigator randomly assigns the participants to treatment groups. A **single-subject design** or *N* of 1 design involves observing the behavior of a single individual (or a small number of individuals) over time.

● Identify what is being compared in the experiment. In many experiments, those of a type called *between-subject* designs, the investigator compares two or more groups (Keppel, 1991; Rosenthal & Rosnow, 1991). For example, a *factorial design* experiment, a variation on the between-group design, involves using two or more treatment variables to examine the independent and simultaneous effects of these treatment variables on an outcome (Vogt, 1999). This widely used behavioral research design explores the effects of each treatment separately and also the effects of variables used in combination, thereby providing a rich and revealing multidimensional view (Keppel, 1991). In other experiments, the researcher studies only one group in what is called a *within-group* design. For example, in a *repeated measures* design, participants are assigned to different treatments at different times during the experiment. Another example of a within-group design would be a study of the behavior of a single individual over time in which the experimenter provides and withholds a treatment at different times in the experiment, to determine its impact.

● Provide a diagram or a figure to illustrate the specific research design to be used. A standard notation system needs to be used in this figure. A **research tip** I recommend is to use a classic notation system provided by Campbell and Stanley (1963, p. 6):

- X represents an exposure of a group to an experimental variable or event, the effects of which are to be measured.
- O represents an observation or measurement recorded on an instrument.
- X's and O's in a given row are applied to the same specific persons. X's and O's in the same column, or placed vertically relative to each other, are simultaneous.
- The left-to-right dimension indicates the temporal order of procedures in the experiment (sometimes indicated with an arrow).
- The symbol R indicates random assignment.
- Separation of parallel rows by a horizontal line indicates that comparison groups are not equal (or equated) by random assignment. No horizontal line between the groups displays random assignment of individuals to treatment groups.

In the following examples, this notation is used to illustrate pre-experimental, quasi-experimental, true experimental, and single-subject designs.

Example 8.2 *Pre-Experimental Designs*

One-Shot Case Study

This design involves an exposure of a group to a treatment followed by a measure.

Group A X——————————O

One-Group Pre-Test–Post-Test Design

This design includes a pre-test measure followed by a treatment and a post-test for a single group.

Group A 01————X————02

Static Group Comparison or Post-Test-Only With Nonequivalent Groups

Experimenters use this design after implementing a treatment. After the treatment, the researcher selects a comparison group and provides a post-test to both the experimental group(s) and the comparison group(s).

Group A X——————————O

Group B ——————————O

Alternative Treatment Post-Test-Only With Nonequivalent Groups Design

This design uses the same procedure as the Static Group Comparison, with the exception that the nonequivalent comparison group received a different treatment.

Group A X1——————————O

Group B X2——————————O

Example 8.3 *Quasi-Experimental Designs*

Nonequivalent (Pre-Test and Post-Test) Control-Group Design

In this design, a popular approach to quasi-experiments, the experimental group A and the control group B are selected without random assignment. Both groups take a pre-test and post-test. Only the experimental group receives the treatment.

Group A O————X————O

Group B O————————O

Single-Group Interrupted Time-Series Design

In this design, the researcher records measures for a single group both before and after a treatment.

Group A O—O—O—O—X—O—O—O—O

Control-Group Interrupted Time-Series Design

A modification of the Single-Group Interrupted Time-Series design in which two groups of participants, not randomly assigned, are observed over time. A treatment is administered to only one of the groups (i.e., Group A).

Group A O—O—O—O—X—O—O—O—O

Group B O—O—O—O—O—O—O—O—O

Example 8.4 *True Experimental Designs*

Pre-Test–Post-Test Control-Group Design

A traditional, classical design, this procedure involves random assignment of participants to two groups. Both groups are administered both a pre-test and a post-test, but the treatment is provided only to experimental Group A.

Group A R———O———X———O
Group B R———O——————O

Post-Test-Only Control-Group Design

This design controls for any confounding effects of a pre-test and is a popular experimental design. The participants are randomly assigned to groups, a treatment is given only to the experimental group, and both groups are measured on the post-test.

Group A R————X————O
Group B R————————O

Solomon Four-Group Design

A special case of a 2 X 2 factorial design, this procedure involves the random assignment of participants to four groups. Pre-tests and treatments are varied for the four groups. All groups receive a post-test.

Group A R———O———X———O
Group B R———O——————O
Group C R——————X———O
Group D R—————————O

Example 8.5 *Single-Subject Designs*

A-B-A Single-Subject Design

This design involves multiple observations of a single individual. The target behavior of a single individual is established over time and is referred to as a baseline behavior. The baseline behavior is assessed, the treatment provided, and then the treatment is withdrawn.

Baseline A Treatment B Baseline A

O–O–O–O–O–X–X–X–X–X–O–O–O–O–O–O

Threats to Validity

There are several threats to validity that will raise questions about an experimenter's ability to conclude that the intervention affects an outcome and not some other factor. Experimental researchers need to identify potential threats to the internal validity of their experiments and design them so that these threats will not likely arise or are minimized. There are two types of threats to validity: internal threats and external threats. **Internal validity threats** are experimental procedures, treatments, or experiences of the participants that threaten the researcher's ability to draw correct inferences from the data about the population in an experiment. Table 8.5 displays these threats, provides a description of each one of them, and suggests potential responses by the researcher so that the threat may not occur. There are those involving participants (i.e., history, maturation, regression, selection, and mortality), those related to the use of an experimental treatment that the researcher manipulates (i.e., diffusion, compensatory and resentful demoralization, and compensatory rivalry), and those involving procedures used in the experiment (i.e., testing and instruments).

Potential threats to external validity also must be identified and designs created to minimize these threats. **External validity threats** arise when experimenters draw incorrect inferences from the sample data to other persons, other settings, and past or future situations. As shown in Table 8.6, these threats arise because of the characteristics of individuals selected for the sample, the uniqueness of the setting, and the timing of the experiment. For example, threats to external validity arise when the researcher generalizes beyond the groups in the experiment to other racial or social groups not under study, to settings not studied, or to past or future situations. Steps for addressing these potential issues are also presented in Table 8.6.

Other threats that might be mentioned in the method section are the threats to **statistical conclusion validity** that arise when experimenters draw inaccurate inferences from the data because of inadequate statistical

Table 8.5 Types of Threats to Internal Validity

Type of Threat to Internal Validity	Description of Threat	In Response, Actions the Researcher Can Take
History	Because time passes during an experiment, events can occur that unduly influence the outcome beyond the experimental treatment.	The researcher can have both the experimental and control groups experience the same external events.
Maturation	Participants in an experiment may mature or change during the experiment, thus influencing the results.	The researcher can select participants who mature or change at the same rate (e.g., same age) during the experiment.
Regression	Participants with extreme scores are selected for the experiment. Naturally, their scores will probably change during the experiment. Scores, over time, regress toward the mean.	A researcher can select participants who do not have extreme scores as entering characteristics for the experiment.
Selection	Participants can be selected who have certain characteristics that predispose them to have certain outcomes (e.g., they are brighter).	The researcher can select participants randomly so that characteristics have the probability of being equally distributed among the experimental groups.
Mortality	Participants drop out during an experiment due to many possible reasons. The outcomes are thus unknown for these individuals.	A researcher can recruit a large sample to account for dropouts or compare those who drop out with those who continue, in terms of the outcome.
Diffusion of treatment	Participants in the control and experimental groups communicate with each other. This communication can influence how both groups score on the outcomes.	The researcher can keep the two groups as separate as possible during the experiment.

(Continued)

Table 8.5 (Continued)

Type of Threat to Internal Validity	Description of Threat	In Response, Actions the Researcher Can Take
Compensatory/resentful demoralization	The benefits of an experiment may be unequal or resented when only the experimental group receives the treatment (e.g., experimental group receives therapy and the control group receives nothing).	The researcher can provide benefits to both groups, such as giving the control group the treatment *after* the experiment ends or giving the control group some different type of treatment *during* the experiment.
Compensatory rivalry	Participants in the control group feel that they are being devalued, as compared to the experimental group, because they do not experience the treatment.	The researcher can take steps to create equality between the two groups, such as reducing the expectations of the control group.
Testing	Participants become familiar with the outcome measure and remember responses for later testing.	The researcher can have a longer time interval between administrations of the outcome or use different items on a later test than were used in an earlier test.
Instrumentation	The instrument changes between a pre-test and post-test, thus impacting the scores on the outcome.	The researcher can use the same instrument for the pre-test and post-test measures.

SOURCE: Adapted from Creswell (2008).

power or the violation of statistical assumptions. Threats to **construct validity** occur when investigators use inadequate definitions and measures of variables.

Practical **research tips** for proposal writers to address validity issues are as follows:

- Identify the potential threats to validity that may arise in your study. A separate section in a proposal may be composed to advance this threat.

- Define the exact type of threat and what potential issue it presents to your study.

Table 8.6 Types of Threats to External Validity

Types of Threats to External Validity	Description of Threat	In Response, Actions the Researcher Can Take
Interaction of selection and treatment	Because of the narrow characteristics of participants in the experiment, the researcher cannot generalize to individuals who do not have the characteristics of participants.	The researcher restricts claims about groups to which the results cannot be generalized. The researcher conducts additional experiments with groups with different characteristics.
Interaction of setting and treatment	Because of the characteristics of the setting of participants in an experiment, a researcher cannot generalize to individuals in other settings.	The researcher needs to conduct additional experiments in new settings to see if the same results occur as in the initial setting.
Interaction of history and treatment	Because results of an experiment are time-bound, a researcher cannot generalize the results to past or future situations.	The researcher needs to replicate the study at later times to determine if the same results occur as in the earlier time.

SOURCE: Adapted from Creswell (2008).

- Discuss how you plan to address the threat in the design of your experiment.

- Cite references to books that discuss the issue of threats to validity, such as Cook and Campbell (1979); Creswell (2008); Reichardt and Mark (1998); Shadish, Cook, & Campbell (2001); and Tuckman (1999).

The Procedure

A proposal developer needs to describe in detail the procedure for conducting the experiment. A reader should be able to understand the design being used, the observations, the treatment, and the timeline of activities.

- Discuss a step-by-step approach for the procedure in the experiment. For example, Borg and Gall (1989, p. 679) outlined six steps typically used in the procedure for a pre-test–post-test control group design with matching participants in the experimental and control groups:

1. Administer measures of the dependent variable or a variable closely correlated with the dependent variable to the research participants.

2. Assign participants to matched pairs on the basis of their scores on the measures described in Step 1.
3. Randomly assign one member of each pair to the experimental group and the other member to the control group.
4. Expose the experimental group to the experimental treatment and administer no treatment or an alternative treatment to the control group.
5. Administer measures of the dependent variables to the experimental and control groups.
6. Compare the performance of the experimental and control groups on the post-test(s) using tests of statistical significance.

Data Analysis

Tell the reader about the types of statistical analysis that will be used during the experiment.

- Report the descriptive statistics calculated for observations and measures at the pre-test or post-test stage of experimental designs. These statistics are means, standard deviations, and ranges.

- Indicate the inferential statistical tests used to examine the hypotheses in the study. For experimental designs with categorical information (groups) on the independent variable and continuous information on the dependent variable, researchers use *t* tests or univariate analysis of variance (ANOVA), analysis of covariance (ANCOVA), or multivariate analysis of variance (MANOVA—multiple dependent measures). (Several of these tests are mentioned in Table 8.3, presented earlier.) In factorial designs, both interaction and main effects of ANOVA are used. When data on a pretest or post-test show marked deviation from a normal distribution, use nonparametric statistical tests.

- For single-subject research designs, use line graphs for baseline and treatment observations for abscissa (horizontal axis) units of time and the ordinate (vertical axis) target behavior. Each data point is plotted separately on the graph, and the data points are connected by lines (e.g., see Neuman & McCormick, 1995). Occasionally, tests of statistical significance, such as *t* tests, are used to compare the pooled mean of the baseline and the treatment phases, although such procedures may violate the assumption of independent measures (Borg & Gall, 1989).

- With increasing frequency, experimental researchers report both statistical results of hypothesis testing and confidence intervals and effect size as indicators of practical significance of the findings. A **confidence interval** is an interval estimate of the range of upper and lower statistical values that are consistent with the observed data and are likely to contain the actual population mean. An **effect size** identifies the strength of the conclusions about group differences or the relationships among variables in quantitative studies. The calculation of effect size varies for different statistical tests.

Interpreting Results

The final step in an experiment is to interpret the findings in light of the hypotheses or research questions set forth in the beginning. In this interpretation, address whether the hypotheses or questions were supported or whether they were refuted. Consider whether the treatment that was implemented actually made a difference for the participants who experienced them. Suggest why or why not the results were significant, drawing on past literature that you reviewed (Chapter 2), the theory used in the study (Chapter 3), or persuasive logic that might explain the results. Address whether the results might have occurred because of inadequate experimental procedures, such as threats to internal validity, and indicate how the results might be generalized to certain people, settings, and times. Finally, indicate the implications of the results for the population studied or for future research.

Example 8.6 *An Experimental Method Section*

The following is a selected passage from a quasi-experimental study by Enns and Hackett (1990) that demonstrates many of the components in an experimental design. Their study addressed the general issue of matching client and counselor interests along the dimensions of attitudes toward feminism. They hypothesized that feminist participants would be more receptive to a radical feminist counselor than would nonfeminist participants and that nonfeminist participants would be more receptive to a nonsexist and liberal feminist counselor. Except for a limited discussion about data analysis and an interpretation section found in the discussion of their article, their approach contains the elements of a good method section for an experimental study.

Method

Participants

The participants were 150 undergraduate women enrolled in both lower- and upper-division courses in sociology, psychology, and communications at a midsized university and a community college, both on the west coast. (*The authors described the participants in this study.*)

(Continued)

(Continued)

Design and Experimental Manipulation

This study used a 3 X 2 X 2 factorial design: Orientation of Counselor (nonsexist-humanistic, liberal feminist, or radical feminist) X Statement of Values (implicit or explicit) X Participants' Identification with Feminism (feminist or nonfeminist). Occasional missing data on particular items were handled by a pairwise deletion procedure. *(Authors identified the overall design.)*

The three counseling conditions, nonsexist-humanistic, liberal, and radical feminist, were depicted by 10 min videotape vignettes of a second counseling session between a female counselor and a female client. . . . The implicit statement of values condition used the sample interview only; the counselor's values were therefore implicit in her responses. The explicit statement of values condition was created by adding to each of the three counseling conditions a 2-min leader that portrayed the counselor describing to the client her counseling approach and associated values including for the two feminist conditions a description of her feminist philosophical orientation, liberal or radical. . . . Three counseling scripts were initially developed on the basis of distinctions between nonsexist-humanistic, liberal, and radical feminist philosophies and attendant counseling implications. Client statements and the outcome of each interview were held constant, whereas counselor responses differed by approach. *(Authors described the three treatment conditions variables manipulated in the study.)*

Instruments

Manipulation checks. As a check on participants' perception of the experimental manipulation and as an assessment of participants' perceived similarity to the three counselors, two subscales of Berryman-Fink and Verderber's (1985) Attributions of the Term Feminist Scale were revised and used in this study as the Counselor Description Questionnaire (CDQ) and the Personal Description Questionnaire (PDQ). . . . Berryman-Fink and Verderber (1985) reported internal consistency reliabilities of .86 and .89 for the original versions of these two subscales. *(Authors discussed the instruments and the reliability of the scales for the dependent variable in the study.)*

Procedure

All experimental sessions were conducted individually. The experimenter, an advanced doctoral student in counseling psychology, greeted each subject, explained the purpose of the study as assessing students' reactions to counseling, and administered the ATF. The ATF was then collected and scored while each subject completed a demographic data form and reviewed a set of instructions for viewing the videotape. The first half of the sample was randomly assigned to one of the twelve videotapes (3 Approaches X 2 Statements X 2 Counselors), and a median was obtained on the ATF. The median for the first half of the sample was then used to categorize the second half of the group as feminist or nonfeminist, and the remainder of the participants was randomly assigned to conditions separately from each feminist orientation group to ensure nearly equal cell sizes. The median on the final sample was checked and a few participants recategorized by the final median split, which resulted in 12 or 13 participants per cell.

After viewing the videotape that corresponded to their experimental assignment, participants completed the dependent measures and were debriefed. (pp. 35–36; *Authors described the procedure used in the experiment.)*

SOURCE: Enns and Hackett (1990). © 1990 by the American Psychological Association. Reprinted with permission.

SUMMARY

This chapter identified essential components in designing a method procedure for a survey or experimental study. The outline of steps for a survey study began with a discussion about the purpose, the identification of the population and sample, the survey instruments to be used, the relationship between the variables, the research questions, specific items on the survey, and steps to be taken in the analysis and the interpretation of the data from the survey. In the design of an experiment, the researcher identifies participants in the study, the variables—the treatment conditions and the outcome variables—and the instruments used for pre-tests and post-tests and the materials to be used in the treatments. The design also includes the specific type of experiment, such as a pre-experimental, quasi-experimental, true experiment, or single-subject design. Then the researcher draws a figure to illustrate the design, using appropriate notation. This is followed by comments about potential threats to internal and external validity (and possibly statistical and construct validity) that relate to the experiment, the statistical analysis used to test the hypotheses or research questions, and the interpretation of the results.

Writing Exercises

1. Design a plan for the procedures to be used in a survey study. Review the checklist in Table 8.1 after you write the section to determine if all components have been addressed.

2. Design a plan for procedures for an experimental study. Refer to Table 8.4 after you complete your plan to determine if all questions have been addressed adequately.

WRITING EXERCISES

ADDITIONAL READINGS

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.

Earl Babbie provides a thorough, detailed text about all aspects of survey design. He reviews the types of designs, the logic of sampling, and examples of designs. He also discusses the conceptualization of a survey instrument and its scales. He then provides useful ideas about administering a questionnaire and processing the results. Also included is a discussion about data analysis with attention to constructing and understanding tables and writing a survey report. This book is detailed, informative, and technically oriented toward students at the intermediate or advanced level of survey research.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 1–76). Chicago: Rand-McNally.

This chapter in the Gage *Handbook* is the classical statement about experimental designs. Campbell and Stanley designed a notation system for experiments that is still used today; they also advanced the types of experimental designs, beginning with factors that jeopardize internal and external validity, the pre-experimental design types, true experiments, quasi-experimental designs, and correlational and ex post facto designs. The chapter presents an excellent summary of types of designs, their threats to validity, and statistical procedures to test the designs. This is an essential chapter for students beginning their study of experimental studies.

Fink, A. (2002). *The survey kit* (2nd ed.). Thousand Oaks, CA: Sage.

"The Survey Kit," is composed of multiple books and edited by Arlene Fink. An overview of the books in this series is provided in the first volume. As an introduction to the volumes, Fink discusses all aspects of survey research, including how to ask questions, how to conduct surveys, how to engage in telephone interviews, how to sample, and how to measure validity and reliability. Much of the discussion is oriented toward the beginning survey researcher, and the numerous examples and excellent illustrations make it a useful tool to learn the basics of survey research.

Fowler, F. J. (2002). *Survey research methods.* (3rd ed.). Thousand Oaks, CA: Sage.

Floyd Fowler provides a useful text about the decisions that go into the design of a survey research project. He addresses use of alternative sampling procedures, ways of reducing nonresponse rates, data collection, design of good questions, employing sound interviewing techniques, preparation of surveys for analysis, and ethical issues in survey designs.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Geoffrey Keppel provides a detailed, thorough treatment of the design of experiments from the principles of design to the statistical analysis of experimental data. Overall, this book is for the mid-level to advanced statistics student who seeks to understand the design and statistical analysis of experiments. The introductory chapter presents an informative overview of the components of experimental designs.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Mark Lipsey has authored a major book on the topics of experimental designs and statistical power of those designs. Its basic premise is that an experiment needs to have sufficient sensitivity to detect those effects it purports to investigate. The book explores statistical power and includes a table to help researchers identify the appropriate size of groups in an experiment.

Neuman, S. B., & McCormick, S. (Eds.). (1995). *Single-subject experimental research: Applications for literacy.* Newark, DE: International Reading Association.

Susan Neuman and Sandra McCormick have edited a useful, practical guide to the design of single-subject research. They present many examples of different types of designs, such as reversal designs and multiple-baseline designs, and they enumerate the statistical procedures that might be involved in analyzing the single-subject data. One chapter, for example, illustrates the conventions for displaying data on line graphs. Although this book cites many applications in literacy, it has broad application in the social and human sciences.

Competing on Analytics

This page intentionally left blank

Harvard Business Review

www.hbrreprints.org

Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.

Competing on Analytics

by Thomas H. Davenport

Included with this full-text *Harvard Business Review* article:

Reprint R0601H

Competing on Analytics

The Idea in Brief

It's virtually impossible to differentiate yourself from competitors based on products alone. Your rivals sell offerings similar to yours. And thanks to cheap offshore labor, you're hard-pressed to beat overseas competitors on product cost.

How to pull ahead of the pack? Become an **analytics competitor:** Use sophisticated data-collection technology and analysis to wring every last drop of value from all your business processes. With analytics, you discern not only what your customers want but also how much they're willing to pay and what keeps them loyal. You look beyond compensation costs to calculate your workforce's exact contribution to your bottom line. And you don't just track existing inventories; you also predict and prevent future inventory problems.

Analytics competitors seize the lead in their fields. Capital One's analytics initiative, for example, has spurred at least 20% growth in earnings per share every year since the company went public.

Make analytics part of *your* overarching competitive strategy, and push it down to decision makers at every level. You'll arm your employees with the best evidence and quantitative tools for making the best decisions—big and small, every day.

The Idea in Practice

To become an analytics competitor:

Champion Analytics from the Top

Acknowledge and endorse the changes in culture, processes, and skills that analytics competition will mean for much of your workforce. And prepare yourself to lead an analytics-focused organization: You will have to understand the theory behind various quantitative methods so you can recognize their limitations. If you lack background in statistical methods, consult experts who understand your business and know how analytics can be applied to it.

Create a Single Analytics Initiative

Place all data-collection and analysis activities under a common leadership, with common technology and tools. You'll facilitate data sharing and avoid the impediments of inconsistent reporting formats, data definitions, and standards.

▶ Example:
Procter & Gamble created a centrally managed "überanalytics" group of 100 analysts drawn from many different functions. It applies this critical mass of expertise to pressing cross-functional issues. For instance, sales and marketing analysts supply data on growth opportunities in existing markets to supply-chain analysts, who can then design more responsive supply networks.

Focus Your Analytics Effort

Channel your resources into analytics initiatives that most directly serve your overarching competitive strategy. Harrah's, for instance, aims much of its analytical activity at improving customer loyalty, customer service, and related areas such as pricing and promotions.

Establish an Analytics Culture

Instill a companywide respect for measuring, testing, and evaluating quantitative evidence. Urge employees to base decisions on hard facts. Gauge and reward performance the same way—applying metrics to compensation and rewards.

Hire the Right People

Pursue and hire analysts who possess top-notch quantitative-analysis skills, can express complex ideas in simple terms, and can interact productively with decision makers. This combination may be difficult to find, so start recruiting well before you need to fill analyst positions.

Use the Right Technology

Prepare to spend significant resources on technology such as customer relationship management (CRM) or enterprise resource planning (ERP) systems. Present data in standard formats, integrate it, store it in a data warehouse, and make it easily accessible to everyone. And expect to spend years gathering enough data to conduct meaningful analyses.

▶ Example:
It took Dell Computer seven years to create a database that includes 1.5 million records of all its print, radio, broadcast TV, and cable ads. Dell couples the database with data on sales for each region in which the ads appeared (before and after their appearance). The information enables Dell to fine-tune its promotions for every medium—in every region.

Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.

Competing on Analytics

by Thomas H. Davenport

We all know the power of the killer app. Over the years, groundbreaking systems from companies such as American Airlines (electronic reservations), Otis Elevator (predictive maintenance), and American Hospital Supply (online ordering) have dramatically boosted their creators' revenues and reputations. These heralded—and coveted—applications amassed and applied data in ways that upended customer expectations and optimized operations to unprecedented degrees. They transformed technology from a supporting tool into a strategic weapon.

Companies questing for killer apps generally focus all their firepower on the one area that promises to create the greatest competitive advantage. But a new breed of company is upping the stakes. Organizations such as Amazon, Harrah's, Capital One, and the Boston Red Sox have dominated their fields by deploying industrial-strength analytics across a wide variety of activities. In essence, they are transforming their organizations into armies of killer apps and crunching their way to victory.

Organizations are competing on analytics not just because they can—business today is awash in data and data crunchers—but also because they should. At a time when firms in many industries offer similar products and use comparable technologies, business processes are among the last remaining points of differentiation. And analytics competitors wring every last drop of value from those processes. So, like other companies, they know what products their customers want, but they also know what prices those customers will pay, how many items each will buy in a lifetime, and what triggers will make people buy more. Like other companies, they know compensation costs and turnover rates, but they can also calculate how much personnel contribute to or detract from the bottom line and how salary levels relate to individuals' performance. Like other companies, they know when inventories are running low, but they can also predict problems with demand and supply chains, to achieve low rates of inventory and high rates of perfect orders.

And analytics competitors do all those things in a coordinated way, as part of an overarching strategy championed by top leadership and pushed down to decision makers at every level. Employees hired for their expertise with numbers or trained to recognize their importance are armed with the best evidence and the best quantitative tools. As a result, they make the best decisions: big and small, every day, over and over and over.

Although numerous organizations are embracing analytics, only a handful have achieved this level of proficiency. But analytics competitors are the leaders in their varied fields—consumer products, finance, retail, and travel and entertainment among them. Analytics has been instrumental to Capital One, which has exceeded 20% growth in earnings per share every year since it became a public company. It has allowed Amazon to dominate online retailing and turn a profit despite enormous investments in growth and infrastructure. In sports, the real secret weapon isn't steroids, but stats, as dramatic victories by the Boston Red Sox, the New England Patriots, and the Oakland A's attest.

At such organizations, virtuosity with data is often part of the brand. Progressive makes advertising hay from its detailed parsing of individual insurance rates. Amazon customers can watch the company learning about them as its service grows more targeted with frequent purchases. Thanks to Michael Lewis's best-selling book *Moneyball,* which demonstrated the power of statistics in professional baseball, the Oakland A's are almost as famous for their geeky number crunching as they are for their athletic prowess.

To identify characteristics shared by analytics competitors, I and two of my colleagues at Babson College's Working Knowledge Research Center studied 32 organizations that have made a commitment to quantitative, fact-based analysis. Eleven of those organizations we classified as full-bore analytics competitors, meaning top management had announced that analytics was key to their strategies; they had multiple initiatives under way involving complex data and statistical analysis, and they managed analytical activity at the enterprise (not departmental) level.

This article lays out the characteristics and practices of these statistical masters and describes some of the very substantial changes other companies must undergo in order to compete on quantitative turf. As one would expect, the transformation requires a significant investment in technology, the accumulation of massive stores of data, and the formulation of companywide strategies for managing the data. But at least as important, it requires executives' vocal, unswerving commitment and willingness to change the way employees think, work, and are treated. As Gary Loveman, CEO of analytics competitor Harrah's, frequently puts it, "Do we think this is true? Or do we know?"

Anatomy of an Analytics Competitor

One analytics competitor that's at the top of its game is Marriott International. Over the past 20 years, the corporation has honed to a science its system for establishing the optimal price for guest rooms (the key analytics process in hotels, known as revenue management). Today, its ambitions are far grander. Through its Total Hotel Optimization program, Marriott has expanded its quantitative expertise to areas such as conference facilities and catering, and made related tools available over the Internet to property revenue managers and hotel owners. It has developed systems to optimize offerings to frequent customers and assess the likelihood of those customers' defecting to competitors. It has given local revenue managers the power to override the system's recommendations when certain local factors can't be predicted (like the large number of Hurricane Katrina evacuees arriving in Houston). The company has even created a revenue opportunity model, which computes actual revenues as a percentage of the optimal rates that could have been charged. That figure has grown from 83% to 91% as Marriott's revenue-management analytics has taken root throughout the enterprise. The word is out among property owners and franchisees: If you want to squeeze the most revenue from your inventory, Marriott's approach is the ticket.

Clearly, organizations such as Marriott don't behave like traditional companies. Customers notice the difference in every interaction; employees and vendors live the difference every day. Our study found three key attributes among analytics competitors:

Widespread use of modeling and optimization. Any company can generate simple descriptive statistics about aspects of its busi-

Thomas H. Davenport (tdavenport@babson.edu) is the President's Distinguished Professor of Information Technology and Management at Babson College in Babson Park, Massachusetts, the director of research at Babson Executive Education, and a fellow at Accenture. He is the author of *Thinking for a Living* (Harvard Business School Press, 2005).

ness—average revenue per employee, for example, or average order size. But analytics competitors look well beyond basic statistics. These companies use predictive modeling to identify the most profitable customers—plus those with the greatest profit potential and the ones most likely to cancel their accounts. They pool data generated in-house and data acquired from outside sources (which they analyze more deeply than do their less statistically savvy competitors) for a comprehensive understanding of their customers. They optimize their supply chains and can thus determine the impact of an unexpected constraint, simulate alternatives, and route shipments around problems. They establish prices in real time to get the highest yield possible from each of their customer transactions. They create complex models of how their operational costs relate to their financial performance.

Leaders in analytics also use sophisticated experiments to measure the overall impact or "lift" of intervention strategies and then apply the results to continuously improve subsequent analyses. Capital One, for example, conducts more than 30,000 experiments a year, with different interest rates, incentives, direct-mail packaging, and other variables. Its goal is to maximize the likelihood both that potential customers will sign up for credit cards and that they will pay back Capital One.

Progressive employs similar experiments using widely available insurance industry data. The company defines narrow groups, or cells, of customers: for example, motorcycle riders ages 30 and above, with college educations, credit scores over a certain level, and no accidents. For each cell, the company performs a regression analysis to identify factors that most closely correlate with the losses that group engenders. It then sets prices for the cells, which should enable the company to earn a profit across a portfolio of customer groups, and uses simulation software to test the financial implications of those hypotheses. With this approach, Progressive can profitably insure customers in traditionally high-risk categories. Other insurers reject high-risk customers out of hand, without bothering to delve more deeply into the data (although even traditional competitors, such as Allstate, are starting to embrace analytics as a strategy).

An enterprise approach. Analytics competitors understand that most business functions—even those, like marketing, that have historically depended on art rather than science—can be improved with sophisticated quantitative techniques. These organizations don't gain advantage from one killer app, but rather from multiple applications supporting many parts of the business—and, in a few cases, being rolled out for use by customers and suppliers.

UPS embodies the evolution from targeted analytics user to comprehensive analytics competitor. Although the company is among the world's most rigorous practitioners of operations research and industrial engineering, its capabilities were, until fairly recently, narrowly focused. Today, UPS is wielding its statistical skill to track the movement of packages and to anticipate and influence the actions of people—assessing the likelihood of customer attrition and identifying sources of problems. The UPS Customer Intelligence Group, for example, is able to accurately predict customer defections by examining usage patterns and complaints. When the data point to a potential defector, a salesperson contacts that customer to review and resolve the problem, dramatically reducing the loss of accounts. UPS still lacks the breadth of initiatives of a full-bore analytics competitor, but it is heading in that direction.

Analytics competitors treat all such activities from all provenances as a single, coherent initiative, often massed under one rubric, such as "information-based strategy" at Capital One or "information-based customer management" at Barclays Bank. These programs operate not just under a common label but also under common leadership and with common technology and tools. In traditional companies, "business intelligence" (the term IT people use for analytics and reporting processes and software) is generally managed by departments; number-crunching functions select their own tools, control their own data warehouses, and train their own people. But that way, chaos lies. For one thing, the proliferation of user-developed spreadsheets and databases inevitably leads to multiple versions of key indicators within an organization. Furthermore, research has shown that between 20% and 40% of spreadsheets contain errors; the more spreadsheets floating around a company, therefore, the more fecund the breeding ground for mistakes. Analytics competitors, by contrast, field centralized groups to ensure that critical data

Employees hired for their expertise with numbers or trained to recognize their importance are armed with the best evidence and the best quantitative tools. As a result, they make the best decisions.

and other resources are well managed and that different parts of the organization can share data easily, without the impediments of inconsistent formats, definitions, and standards.

Some analytics competitors apply the same enterprise approach to people as to technology. Procter & Gamble, for example, recently created a kind of überanalytics group consisting of more than 100 analysts from such functions as operations, supply chain, sales, consumer research, and marketing. Although most of the analysts are embedded in business operating units, the group is centrally managed. As a result of this consolidation, P&G can apply a critical mass of expertise to its most pressing issues. So, for example, sales and marketing analysts supply data on opportunities for growth in existing markets to analysts who design corporate supply networks. The supply chain analysts, in turn, apply their expertise in certain decision-analysis techniques to such new areas as competitive intelligence.

The group at P&G also raises the visibility of analytical and data-based decision making within the company. Previously, P&G's crack analysts had improved business processes and saved the firm money; but because they were squirreled away in dispersed domains, many executives didn't know what services they offered or how effective they could be. Now those executives are more likely to tap the company's deep pool of expertise for their projects. Meanwhile, masterful number crunching has become part of the story P&G tells to investors, the press, and the public.

Senior executive advocates. A companywide embrace of analytics impels changes in culture, processes, behavior, and skills for many employees. And so, like any major transition, it requires leadership from executives at the very top who have a passion for the quantitative approach. Ideally, the principal advocate is the CEO. Indeed, we found several chief executives who have driven the shift to analytics at their companies over the past few years, including Loveman of Harrah's, Jeff Bezos of Amazon, and Rich Fairbank of Capital One. Before he retired from the Sara Lee Bakery Group, former CEO Barry Beracha kept a sign on his desk that summed up his personal and organizational philosophy: "In God we trust. All others bring data." We did come across some companies in which a single functional or business unit leader was trying to push analytics throughout the organization, and a few were making some progress. But we found that these lower-level people lacked the clout, the perspective, and the cross-functional scope to change the culture in any meaningful way.

CEOs leading the analytics charge require both an appreciation of and a familiarity with the subject. A background in statistics isn't necessary, but those leaders must understand the theory behind various quantitative methods so

Going to Bat for Stats

The analysis-versus-instinct debate, a favorite of political commentators during the last two U.S. presidential elections, is raging in professional sports, thanks to several popular books and high-profile victories. For now, analysis seems to hold the lead.

Most notably, statistics are a major part of the selection and deployment of players. *Moneyball,* by Michael Lewis, focuses on the use of analytics in player selection for the Oakland A's—a team that wins on a shoestring. The New England Patriots, a team that devotes an enormous amount of attention to statistics, won three of the last four Super Bowls, and their payroll is currently ranked 24th in the league. The Boston Red Sox have embraced "sabermetrics" (the application of analysis to baseball), even going so far as to hire Bill James, the famous baseball statistician who popularized that term. Analytic HR strategies are taking hold in European soccer as well. One leading team, Italy's A.C. Milan, uses predictive models from its Milan Lab research center to prevent injuries by analyzing physiological, orthopedic, and psychological data from a variety of sources. A fast-rising English soccer team, the Bolton Wanderers, is known for its manager's use of extensive data to evaluate players' performance.

Still, sports managers—like business leaders—are rarely fact-or-feeling purists. St. Louis Cardinals manager Tony La Russa, for example, brilliantly combines analytics with intuition to decide when to substitute a charged-up player in the batting lineup or whether to hire a spark-plug personality to improve morale. In his recent book, *Three Nights in August,* Buzz Bissinger describes that balance: "La Russa appreciated the information generated by computers. He studied the rows and the columns. But he also knew they could take you only so far in baseball, maybe even confuse you with a fog of overanalysis. As far as he knew, there was no way to quantify desire. And those numbers told him exactly what he needed to know when added to twenty-four years of managing experience."

That final sentence is the key. Whether scrutinizing someone's performance record or observing the expression flitting across an employee's face, leaders consult their own experience to understand the "evidence" in all its forms.

that they recognize those methods' limitations—which factors are being weighed and which ones aren't. When the CEOs need help grasping quantitative techniques, they turn to experts who understand the business and how analytics can be applied to it. We interviewed several leaders who had retained such advisers, and these executives stressed the need to find someone who can explain things in plain language and be trusted not to spin the numbers. A few CEOs we spoke with had surrounded themselves with very analytical people—professors, consultants, MIT graduates, and the like. But that was a personal preference rather than a necessary practice.

Of course, not all decisions should be grounded in analytics—at least not wholly so. Personnel matters, in particular, are often well and appropriately informed by instinct and anecdote. More organizations are subjecting recruiting and hiring decisions to statistical analysis (see the sidebar "Going to Bat for Stats"). But research shows that human beings can make quick, surprisingly accurate assessments of personality and character based on simple observations. For analytics-minded leaders, then, the challenge boils down to knowing when to run with the numbers and when to run with their guts.

Their Sources of Strength

Analytics competitors are more than simple number-crunching factories. Certainly, they apply technology—with a mixture of brute force and finesse—to multiple business problems. But they also direct their energies toward finding the right focus, building the right culture, and hiring the right people to make optimal use of the data they constantly churn. In the end, people and strategy, as much as information technology, give such organizations strength.

The right focus. Although analytics competitors encourage universal fact-based decisions, they must choose where to direct resource-intensive efforts. Generally, they pick several functions or initiatives that together serve an overarching strategy. Harrah's, for example, has aimed much of its analytical activity at increasing customer loyalty, customer service, and related areas like pricing and promotions. UPS has broadened its focus from logistics to customers, in the interest of providing superior service. While such multipronged strate-

THINGS YOU CAN COUNT ON

Analytics competitors make expert use of statistics and modeling to improve a wide variety of functions. Here are some common applications:

FUNCTION	DESCRIPTION	EXEMPLARS
Supply chain	Simulate and optimize supply chain flows; reduce inventory and stock-outs.	Dell, Wal-Mart, Amazon
Customer selection, loyalty, and service	Identify customers with the greatest profit potential; increase likelihood that they will want the product or service offering; retain their loyalty.	Harrah's, Capital One, Barclays
Pricing	Identify the price that will maximize yield, or profit.	Progressive, Marriott
Human capital	Select the best employees for particular tasks or jobs, at particular compensation levels.	New England Patriots, Oakland A's, Boston Red Sox
Product and service quality	Detect quality problems early and minimize them.	Honda, Intel
Financial performance	Better understand the drivers of financial performance and the effects of nonfinancial factors.	MCI, Verizon
Research and development	Improve quality, efficacy, and, where applicable, safety of products and services.	Novartis, Amazon, Yahoo

In traditional companies, departments manage analytics —number-crunching functions select their own tools and train their own people. But that way, chaos lies.

gies define analytics competitors, executives we interviewed warned companies against becoming too diffuse in their initiatives or losing clear sight of the business purpose behind each.

Another consideration when allocating resources is how amenable certain functions are to deep analysis. There are at least seven common targets for analytical activity, and specific industries may present their own (see "Things You Can Count On"). Statistical models and algorithms that dangle the possibility of performance breakthroughs make some prospects especially tempting. Marketing, for example, has always been tough to quantify because it is rooted in psychology. But now consumer products companies can hone their market research using multiattribute utility theory—a tool for understanding and predicting consumer behaviors and decisions. Similarly, the advertising industry is adopting econometrics—statistical techniques for measuring the lift provided by different ads and promotions over time.

The most proficient analytics practitioners don't just measure their own navels—they also help customers and vendors measure theirs. Wal-Mart, for example, insists that suppliers use its Retail Link system to monitor product movement by store, to plan promotions and layouts within stores, and to reduce stock-outs. E.&J. Gallo provides distributors with data and analysis on retailers' costs and pricing so they can calculate the per-bottle profitability for each of Gallo's 95 wines. The distributors, in turn, use that information to help retailers optimize their mixes while persuading them to add shelf space for Gallo products. Procter & Gamble offers data and analysis to its retail customers, as part of a program called Joint Value Creation, and to its suppliers to help improve responsiveness and reduce costs. Hospital supplier Owens & Minor furnishes similar services, enabling customers and suppliers to access and analyze their buying and selling data, track ordering patterns in search of consolidation opportunities, and move off-contract purchases to group contracts that include products distributed by Owens & Minor and its competitors. For example, Owens & Minor might show a hospital chain's executives how much money they could save by consolidating purchases across multiple locations or help them see the trade-offs between increasing delivery frequency and carrying inventory.

The right culture. Culture is a soft concept; analytics is a hard discipline. Nonetheless, analytics competitors must instill a companywide respect for measuring, testing, and evaluating quantitative evidence. Employees are urged to base decisions on hard facts. And they know that their performance is gauged the same way. Human resource organizations within analytics competitors are rigorous about applying metrics to compensation and rewards. Harrah's, for example, has made a dramatic change from a rewards culture based on paternalism and tenure to one based on such meticulously collected performance measurements as financial and customer service results. Senior executives also set a consistent example with their own behavior, exhibiting a hunger for and confidence in fact and analysis. One exemplar of such leadership was Beracha of the Sara Lee Bakery Group, known to his employees as a "data dog" because he hounded them for data to support any assertion or hypothesis.

Not surprisingly, in an analytics culture, there's sometimes tension between innovative or entrepreneurial impulses and the requirement for evidence. Some companies place less emphasis on blue-sky development, in which designers or engineers chase after a gleam in someone's eye. In these organizations, R&D, like other functions, is rigorously metric-driven. At Yahoo, Progressive, and Capital One, process and product changes are tested on a small scale and implemented as they are validated. That approach, well established within various academic and business disciplines (including engineering, quality management, and psychology), can be applied to most corporate processes—even to not-so-obvious candidates, like human resources and customer service. HR, for example, might create profiles of managers' personality traits and leadership styles and then test those managers in different situations. It could then compare data on individuals' performance with data about personalities to determine what traits are most important to managing a project that is behind schedule, say, or helping a new group to assimilate.

There are, however, instances when a decision to change something or try something new must be made too quickly for extensive analysis, or when it's not possible to gather data beforehand. For example, even though Amazon's Jeff Bezos greatly prefers to rigorously quantify users' reactions before rolling out new features, he couldn't test the company's search-inside-the-book offering without applying it to a critical mass of books (120,000, to begin with). It was also expensive to develop, and that increased the risk. In this case, Bezos trusted his instincts and took a flier. And the feature did prove popular when introduced.

The right people. Analytical firms hire analytical people—and like all companies that compete on talent, they pursue the best. When Amazon needed a new head for its global supply chain, for example, it recruited Gang Yu, a professor of management science and software entrepreneur who is one of the world's leading authorities on optimization analytics. Amazon's business model requires the company to manage a constant flow of new products, suppliers, customers, and promotions, as well as deliver orders by promised dates. Since his arrival, Yu and his team have been designing and building sophisticated supply chain systems to optimize those processes. And while he tosses around phrases like "nonstationary stochastic processes," he's also good at explaining the new approaches to Amazon's executives in clear business terms.

Established analytics competitors such as Capital One employ squadrons of analysts to conduct quantitative experiments and, with the results in hand, design credit card and other financial offers. These efforts call for a specialized skill set, as you can see from this job description (typical for a Capital One analyst):

> High conceptual problem-solving and quantitative analytical aptitudes…Engineering, financial, consulting, and/or other analytical quantitative educational/work background. Ability to quickly learn how to use software applications. Experience with Excel models. Some graduate work preferred but not required (e.g., MBA). Some experience with project management methodology, process improvement tools (Lean, Six Sigma), or statistics preferred.

Other firms hire similar kinds of people, but analytics competitors have them in much greater numbers. Capital One is currently seeking three times as many analysts as operations people—hardly the common practice for a bank. "We are really a company of analysts," one executive there noted. "It's the primary job in this place."

Good analysts must also have the ability to express complex ideas in simple terms and have the relationship skills to interact well with decision makers. One consumer products company with a 30-person analytics group looks for what it calls "PhDs with personality"—people with expertise in math, statistics, and data analysis who can also speak the language of business and help market their work internally and sometimes externally. The head of a customer analytics group at Wachovia Bank describes the rapport with others his group seeks: "We are trying to build our people as part of the business team," he explains. "We want them sitting at the business table, participating in a discussion of what the key issues are, determining what information needs the businesspeople have, and recommending actions to the business partners. We want this [analytics group] to be not just a general utility, but rather an active and critical part of the business unit's success."

Of course, a combination of analytical, business, and relationship skills may be difficult to find. When the software company SAS (a sponsor of this research, along with Intel) knows it will need an expert in state-of-the-art business applications such as predictive modeling or recursive partitioning (a form of decision tree analysis applied to very complex data sets), it begins recruiting up to 18 months before it expects to fill the position.

In fact, analytical talent may be to the early 2000s what programming talent was to the late 1990s. Unfortunately, the U.S. and European labor markets aren't exactly teeming with analytically sophisticated job candidates. Some organizations cope by contracting work to countries such as India, home to many statistical experts. That strategy may succeed when offshore analysts work on stand-alone problems. But if an iterative discussion with business decision makers is required, the distance can become a major barrier.

The right technology. Competing on analytics means competing on technology. And while the most serious competitors investigate the latest statistical algorithms and decision science approaches, they also constantly monitor and push the IT frontier. The analytics group at one consumer products company went so far as to build its own supercomputer because it felt that commercially available models were inadequate for its demands. Such heroic feats usually aren't necessary, but seri-

The most proficient analytics practitioners don't just measure their own navels—they also help customers and vendors measure theirs.

ous analytics does require the following:

A data strategy. Companies have invested many millions of dollars in systems that snatch data from every conceivable source. Enterprise resource planning, customer relationship management, point-of-sale, and other systems ensure that no transaction or other significant exchange occurs without leaving a mark. But to compete on that information, companies must present it in standard formats, integrate it, store it in a data warehouse, and make it easily accessible to anyone and everyone. And they will need *a lot* of it. For example, a company may spend several years accumulating data on different marketing approaches before it has gathered enough to reliably analyze the effectiveness of an advertising campaign. Dell employed DDB Matrix, a unit of the advertising agency DDB Worldwide, to create (over a period of seven years) a database that includes 1.5 million records on all the computer maker's print, radio, network TV, and cable ads, coupled with data on Dell sales for each region in which the ads appeared (before and after their appearance). That information allows Dell to fine-tune its promotions for every medium in every region.

Business intelligence software. The term "business intelligence," which first popped up in the late 1980s, encompasses a wide array of processes and software used to collect, analyze, and disseminate data, all in the interests of better decision making. Business intelligence tools allow employees to extract, transform, and load (or ETL, as people in the industry would say) data for analysis and then make those analyses available in reports, alerts, and scorecards. The popularity of analytics competition is partly a response to the emergence of integrated packages of these tools.

Computing hardware. The volumes of data required for analytics applications may strain the capacity of low-end computers and servers. Many analytics competitors are converting their hardware to 64-bit processors that churn large amounts of data quickly.

You Know You Compete on Analytics When…

1. You apply sophisticated information systems and rigorous analysis not only to your core capability but also to a range of functions as varied as marketing and human resources.

2. Your senior executive team not only recognizes the importance of analytics capabilities but also makes their development and maintenance a primary focus.

3. You treat fact-based decision making not only as a best practice but also as a part of the culture that's constantly emphasized and communicated by senior executives.

4. You hire not only people with analytical skills but a lot of people with *the very best* analytical skills—and consider them a key to your success.

5. You not only employ analytics in almost every function and department but also consider it so strategically important that you manage it at the enterprise level.

6. You not only are expert at number crunching but also invent proprietary metrics for use in key business processes.

7. You not only use copious data and in-house analysis but also share them with customers and suppliers.

8. You not only avidly consume data but also seize every opportunity to generate information, creating a "test and learn" culture based on numerous small experiments.

9. You not only have committed to competing on analytics but also have been building your capabilities for several years.

10. You not only emphasize the importance of analytics internally but also make quantitative capabilities part of your company's story, to be shared in the annual report and in discussions with financial analysts.

The Long Road Ahead

Most companies in most industries have excellent reasons to pursue strategies shaped by analytics. Virtually all the organizations we identified as aggressive analytics competitors are clear leaders in their fields, and they attribute much of their success to the masterful exploitation of data. Rising global competition intensifies the need for this sort of proficiency. Western companies unable to beat their Indian or Chinese competitors on product cost, for example, can seek the upper hand through optimized business processes.

Companies just now embracing such strategies, however, will find that they take several years to come to fruition. The organizations in our study described a long, sometimes arduous journey. The UK Consumer Cards and Loans business within Barclays Bank, for example, spent five years executing its plan to apply analytics to the marketing of credit cards and other financial products. The company had to make process changes in virtually every aspect of its consumer business: underwriting risk, setting credit limits, servicing accounts, controlling fraud, cross selling, and so on. On the technical side, it had to integrate data on 10 million Barclaycard customers, improve the quality of the data, and build systems to step up data collection and analysis. In addition, the company embarked on a long series of small tests to begin learning how to attract and retain the best customers at the lowest price. And it had to hire new people with top-drawer quantitative skills.

Much of the time—and corresponding expense—that any company takes to become an analytics competitor will be devoted to technological tasks: refining the systems that produce transaction data, making data available in warehouses, selecting and implementing analytic software, and assembling the hardware and communications environment. And because those who don't record history are doomed not to learn from it, companies that have collected little information—or the wrong kind—will need to amass a sufficient body of data to support reliable forecasting. "We've been collecting data for six or seven years, but it's only become usable in the last two or three, because we needed time and experience to validate conclusions based on the data," remarked a manager of customer data analytics at UPS.

And, of course, new analytics competitors will have to stock their personnel larders with fresh people. (When Gary Loveman became COO, and then CEO, of Harrah's, he brought in a group of statistical experts who could design and implement quantitatively based marketing campaigns and loyalty programs.) Existing employees, meanwhile, will require extensive training. They need to know what data are available and all the ways the information can be analyzed; and they must learn to recognize such peculiarities and shortcomings as missing data, duplication, and quality problems. An analytics-minded executive at Procter & Gamble suggested to me that firms should begin to keep managers in their jobs for longer periods because of the time required to master quantitative approaches to their businesses.

The German pathologist Rudolph Virchow famously called the task of science "to stake out the limits of the knowable." Analytics competitors pursue a similar goal, although the universe they seek to know is a more circumscribed one of customer behavior, product movement, employee performance, and financial reactions. Every day, advances in technology and techniques give companies a better and better handle on the critical minutiae of their operations.

The Oakland A's aren't the only ones playing moneyball. Companies of every stripe want to be part of the game.

Reprint R0601H
To order, see the next page

Further Reading

ARTICLES

Diamonds in the Data Mine
by Gary Loveman
Harvard Business Review
May 2003
Product no. 3647

Gaming giant Harrah's CEO Loveman describes how his company uses analytics to win its clientele's devotion and supercharge revenues. Harrah's acquires extensive customer information through a transactional database that records each customer's activity at various points of sale, then slices and dices the data finely to develop strategies for encouraging customers to visit Harrah's casinos regularly. It identifies core customers by calculating their lifetime value, and rewards them for spending more. Thanks to analytics, Harrah's scored 16 straight quarters of same-store revenue growth.

The Surprising Economics of a "People Business"
by Felix Barber and Rainer Strack
Harvard Business Review
June 2005
Product no. R0506D

The authors explain how to use analytics to manage your company's human resources more effectively. In "people businesses"—companies with high employee costs, low capital investment, and limited spending on activities intended to generate future revenue—you need to use the right metrics to assess performance. Avoid relying on capital-oriented metrics (such as return on assets or return on equity); they mask weak performance or indicate market volatility where it may not exist. Instead, use financially rigorous, people-oriented metrics—such as a reformulation of a conventional calculation of economic profit—so you're gauging people's productivity. Reward excellent performance through variable compensation schemes, and price products and services in ways that capture a share of the additional value your people create for customers.

Countering the Biggest Risk of All
by Adrian J. Slywotzky and John Drzik
Harvard Business Review
April 2005
Product no. 977X

This article presents ideas for using analytics to understand and mitigate a particularly grave strategic risk—sudden shifts in customer tastes that redefine your industry. Mitigate this risk by gathering and analyzing proprietary information to detect potential shifts. And conduct fast, cheap experiments to identify attractive offerings for different customer microsegments. For example, Coach wondered whether its customers would remain loyal if it offered trendier styles. It conducted in-store product tests and market experiments to gauge the impact of new pricing, features, and offers by competitive brands. It used the information to quickly alter product designs, drop unappealing items, and create new lines featuring different fabrics and colors. The upshot? Coach retained its traditional fans *and* attracted new customers.

Harvard Business Review
www.hbrreprints.org

U.S. and Canada
800-988-0886
617-783-7500
617-783-7555 fax

In Vivo to in Vitro to in Silico: Coping with Tidal Waves of Data at Biogen

This page intentionally left blank

9-602-122
APRIL 09, 2002

JUAN ENRIQUEZ
GARY PISANO
GAYE L. BOK

In vivo to *in vitro* to *in silico*:[1] Coping with Tidal Waves of Data at Biogen

By early 2002, Biogen was one of the few success stories in biotech. Founded in 1978, it had survived in a tough industry, one in which early starts often lead to mergers, no profits, and/or bankruptcies. By FY2001, Biogen sales had reached $1.0 billion dollars with a healthy $273 million in net income. Its market capitalization was approximately $8.2 billion.[2] R&D spending, at 30% of revenue, was among the highest rates in the industry. (See **Exhibits 1** and **2**.)

Yet Jim Mullen, Biogen's CEO, knew the company now faced another critical juncture. The world in which biotechnology companies operated had changed dramatically over the past few years. Various governments and companies had launched massive genomics projects to sequence, annotate and use gene data from viruses, bacteria, plants, animals, and humans.[3] Parts of biology were rapidly evolving from being an individualistic, wet lab, bench-science driven field towards one where scientists manipulated huge amounts of data and divided up research steps into a factory-like production process. At the same time, the cost of developing a drug and bringing it to market had ballooned, from an estimated $231 million in 1991 to $802 million in 2000.[4] This had led some to believe that scale economies in research and development would give large, well-funded competitors an advantage.

---

[1] Biological research can be conducted *in vivo* that is on live animals or *in vitro* on tissue and cell cultures held in petri dishes. Now more and more preliminary research is carried out within computers that are *in silico* rather than in wet labs. "Wet" lab space refers to laboratory space equipped with benches, ventilation hoods, ventilated storage areas, sinks, lab equipment, etc. where "wet" experiments can be conducted safely. Such laboratory space is relatively expensive to build and maintain.

[2] "Biogen Delivers on Operating Earnings Guidance of $1.90 Per Share After One—Time Charges," *PR Newswire*, January 24, 2002.

[3] The first full complex genome sequenced was haemophilus influenzae (R.D. Fleisehmann, et al., "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd," *Science*, vol. 269 (July 28, 1995): 496-512), this process culminated six years later with the simultaneous publication of two draft versions of the human genome. (J. Craig Venter, et al., "The Sequence of the Human Genome," *Science*, vol. 269: 1304-1351 (February 16, 2001), and The Genome International Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409: 860-921 (February 15, 2001)).

[4] Tufts Center for the Study of Drug Development, Press Release, Boston, Mass., November 30, 2001. The 1991 figure is expressed in 1987 dollars.

Research Associate Gaye L. Bok prepared this case under the supervision of Juan Enriquez, Director of the Harvard Business School's Life Science Project, and Professor Gary P. Pisano. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Major pharmaceutical companies had been consolidating for years, with the top ten firms accounting for 64% of pharmaceutical revenues in 2000, up from 58% in 1995.[5] Most, if not all, had acquired biotechnology firms to add to their quiver of drug development tools. They were also deploying ever more sophisticated IT modules and accessing databases much larger than any databases ever used previously in the industry. (See **Exhibit 3**.)

Potentially, pharmaceutical and biotech companies now faced increased competition from what had begun as gene sequencing-data companies. Start-ups like Millennium, HGS, and Celera had achieved billions in market capitalization and were attempting to integrate forward into the therapeutics market by leveraging their massive genomic databases. Even information-computer companies, such as IBM and Compaq, were targeting life sciences as a key area for future growth.[6]

For a while, Biogen ignored many of these trends. After all it had been successful; it had more than enough to focus its energy and budget on, and it was a leader in Multiple Sclerosis (MS) research. But by the late nineties, it was clear that the world of biology had changed radically and that Biogen needed to catch-up in utilizing bioinformatic tools and understanding rapidly accumulating gene databases. This realization led to a restructuring of its computer systems and lab research. Suddenly massive amounts of data began to flow into the company.

"You must understand," said Michael Rosenberg, Biogen's Associate Director, Bioinformatics, "a revolution (as to how we do research) occurred here at Biogen eleven or twelve months ago. It was generally well accepted, and long awaited, and we expect to see the impact on candidates in the Discovery, Validation and Pre-clinical. It is harder to foresee the impact on Phases one, two, and three."

Restructuring the first part of a drug development pipeline, that is target identification and development prior to pre-clinical evaluation, might also imply significant changes down the road for other parts of Biogen. As Jim Mullen thought about what his company might look like in five years, he asked himself what changes might be required within various parts of his organization including finance, human resources, lab architecture, and clinical trials. Did Biogen have the scale to implement these changes by itself? And, as Mullen surveyed the external competitive landscape, he also wondered who might become a critical ally and who a foe?

Great Lab Science Begets Biogen

Biogen was founded and grounded in a tradition of great lab science. The company was one of the first to develop recombinant proteins using genetic engineering technology.[7] Two of its co-founders, Phillip Sharp and Walter Gilbert, won Nobel Prizes. Inventors like Charles Weissmann and Sir Kenneth Murray built great labs and dreamed up new products. The overall ethos was to find superstars and build labs around their talents.

But great science does not automatically imply great business. Biogen circa 1985 resembled a big post-doc lab. There was a lot of interesting work on protein and molecule development. This work was then published, but there was little focus or follow up on developing commercially attractive

[5] Derived from Pharmaceutical Companies sales data, Compustat Global Vantage.

[6] For a description of these trends see Juan Enriquez and Ray A. Goldberg, "Transforming Life, Transforming Business: The Life-Science Revolution," *Harvard Business Review* (March-April 2000), pp. 94-104.

[7] Walter Gilbert et al. Protein Secretion. July 6, 1982. USPTO Patent # 4,338,397.

products. For example, although Biogen filed a U.S. patent application for Recombinant DNA molecules and their use in producing human interferon-like peptides in 1980, the patent (# 4,530,901) was not issued until July 23, 1985. Even then, profits still seemed a distant promise. Beta Interferon was initially launched as a cancer drug, and failed. It was then pursued as a treatment for hepatitis with limited success, and only later was tried successfully against multiple sclerosis.

Wall Street became restless. Growing R&D expenditures on a broad range of projects led investors to question Biogen's commitment to turning a profit. In a bid to transform a research organization into a vertically integrated development, manufacturing, and marketing organization, the board recruited Jim Vincent to serve as CEO in 1985. He found that the company was close to a financial meltdown.[8] Vincent sold off Biogen's European operations, built up the management team, and renegotiated important royalties for the interferon products. He also refocused research on four areas: inflammation, thrombosis, virology and selected cancers. A number of products already in development were slashed and headcount reduced from 500 to 225 employees.[9] Over time, Vincent focused the company's product development efforts on two drug candidates, Hirulog (a blood thinning drug) and Avonex (then targeted at Hepatitis). Biogen became profitable in 1989.

As the two drugs progressed to Phase III clinical trials, Jim Vincent realized that Biogen had to add to its core competence and focused on bringing in a CEO with operating experience. He brought in James Tobin, from Baxter International, as President and COO in 1994. At that time, Hirulog was the favored project, but efficacy trials showed it was not that much better than the then market leader. At the same time a Biogen competitor, Chiron Corporation, was bringing a similar beta interferon to market for treating multiple sclerosis. Biogen's scientists recognized that their variant of beta interferon would outperform Chiron's. Management suspended trials of Avonex in hepatitis, brought forward their trials of Avonex in multiple sclerosis, and cut the Hirulog project in order to take Avonex to market quickly.[10] The gamble paid off. The product was a success and provided a rapidly rising stream of product revenues, enabling Biogen to remain profitable and grow.

Biogen's management had effectively bet the company on bringing one product, Avonex, an Interferon beta-1a to treat relapsing forms of multiple sclerosis, to market. Through innovative partnerships, Tobin was able to outsource drug formulation, packaging, warehousing and distribution, while retaining the critical bulk manufacturing operation in-house. This helped ensure a large supply and rapid distribution of Avonex.[11] In 1997, after the successful commercialization of the company's sole proprietary drug, Tobin was named CEO and asked to focus on increasing Biogen's drug pipeline.[12] During this period, Biogen initiated numerous partnerships to pursue new disease indications and tried to acquire early stage products.

However, in December 1998, Jim Tobin abruptly resigned citing personal reasons. Vincent once again stepped in as CEO and promoted Jim Mullen, the head of international operations, to President. Mullen had joined Biogen in 1989 after nine years at SmithKline-Beecham Corporation, and had come up through Biogen's operations organization, rising to Vice President, Operations in

---

[8] See Steven C. Wheelwright, "Biogen Inc.: rBeta Interferon Manufacturing Process Development," HBS case No. 696-083. This covers the early history of the company.

[9] Lawrence M. Fisher, "The Rocky Road From Startup to Big-time Player: Biogen's Triumphs Against the Odds," *Strategy & Business* (third quarter 1997), Booz Allen & Hamilton, http://www.strategy-business.com/casestudy/97305/page1.html.

[10] Ibid., page3.html.

[11] David Bovet and Joseph Martha, "Biogen Unchained," *Harvard Business Review* (May 2000), p. 28.

[12] Ronald Rosenberg, "Biogen Engineers Leadership Change of Company," *The Boston Globe*, February 15, 1997.

December 1991 until 1996 when he was appointed Vice President, International.[13] An engineer by training, Mullen had succeeded by focusing on operations and getting things done. In May 2000, Mullen became CEO, while Vincent remained Chairman of the Board of Directors.

A Continued Quest to Diversify

Despite its financial success, Biogen was in a delicate position. It continued to live primarily off Avonex. In FY 2000, the drug accounted for $761 million, or 82% of Biogen revenues. (Biogen's other income derived primarily from royalties on worldwide sales of products it had licensed out in its early days, including alpha interferon and hepatitis B vaccines as well as diagnostic products.)[14] In January 2001, CEO Jim Mullen told the financial community: "Biogen's next great challenge is to transition from being a one-product company into a multi-product company."

Nevertheless, the company believed its marketing relationships with physicians specializing in M.S. was one of its key competitive advantages, and it sought to extend its coverage of the M.S. market (**Exhibits 4** and **5**). As of early 2002, nine of the 14 products undergoing clinical trials at Biogen were Avonex related product extensions.[15]

For much of its history Biogen had prided itself on the quality of its biology and relied on the opportunistic research culture, as well as partnerships, to bring drug candidates into the pipeline. Lab biologists and the medical division were the key drivers in deciding where Biogen should put its research efforts. This culture centered on strong individual scientists with small, independent laboratories. Success was determined by finding good target genes using crack scientists, "wet" lab space, creativity, and a great deal of luck. By definition, biology remains a set of incredibly complex systems. Interactions between living creatures and drug molecules are often unpredictable and hard to understand fully. As Victor Koteliansky, Biogen's Director of Biological Research put it, "In biology it is a rule of the game that on Monday your experiment is working and the same experiment on Friday gives you partially a different thing! "

Identifying the genes responsible for human diseases remains a difficult, labor-intensive process. Finding a promising target that a small molecule or protein could interact with to block a disease without causing excessive side effects typically consumed large portions of a biologists' career. Scientists would rely on library research and clinical studies to develop a hypothesis regarding what genes and proteins were implicated in a disease. Then they would begin arduous lab work to test their hypothesis. Target validation was unpredictable and very time consuming. Even after identifying genes, compounds had to be isolated and introduced to living organisms." Test cycles on bacteria and mice would be run and meticulously recorded, then a further battery of extensive animal tests would eventually lead to a few human trials.

Faced with these challenges, Biogen's management had been quite conservative in funding projects throughout the nineties. Many scientists felt that a cumbersome decision making process had stifled projects. Although researchers could pursue a breadth of good ideas, getting a go ahead required multiple presentations, first to the research management and then to Joe Davie, head of R&D, as well as to the directors of the Medical and Marketing departments, and ultimately to Jim Vincent.

---

[13] Biogen, Inc., 10K, March 2001 (via Hoovers online).

[14] Biogen Annual Report, 2000.

[15] Biogen.com, Product Pipeline, January 24, 2002.

There were good reasons to be cautious. Developing therapeutic drugs remains a risky and expensive business (**Exhibit 6**). The Pharmaceutical Research and Manufacturers of America (PhRMA) trade group estimated that out of 5000 drugs that entered pre-clinical testing, only five would proceed to clinical trials.[16] Only one of these five clinical candidates would eventually reach the corner pharmacy.[17] Stringent tests and protocols, required by the Federal Drug Administration (FDA), meant it took twelve to fifteen years to bring a drug to market. Cost estimates for a new drug ranged from $500 million (PhRMA), to $802 million (Tufts Center for the Study of Drug Development), to even higher figures put forward by various consulting firms.[18]

Costs increase substantially throughout the clinical study phases, with the largest outlays occurring in Phase III, because a company must recruit thousands of people to test the drug. (See **Exhibit 7**.) Even after reaching Phase III, a significant number of candidates failed to demonstrate the efficacy and safety required to win FDA approval. This meant that companies had to invest in various failures and wait 12 to 15 years before having any hope of generating new revenues. Keeping a pipeline full of potential compounds at various development stages was critical to investor confidence. Despite these high hurdles, for large pharmaceutical companies, drug development remained a profitable investment.

Unfortunately, by the end of the nineties, much of Biogen's promising pipeline seemed to be drying up (**Exhibit 8**). Investor relations had a difficult time articulating Biogen's R&D strategy, and there was frequent confusion regarding how far along the development pipeline various projects were. In September 1999, a promising new drug, Antova, experienced safety issues during its Phase II evaluation. Trials were halted. Some on Wall Street came to believe that the pipeline was largely empty.

A New R&D Strategy

A strong focus on the bottom line and conservative attitude towards the technology had led senior management to turn down several opportunities to collaborate on leading-edge genomics projects. According to Rich Cate, Director, Gene Discovery, "We didn't make a big effort to get into the sequencing game. . . . We continued to make small decisions, to turn down opportunities as they arose, and eventually this came to be seen as a strategic decision."[19]

Meanwhile a revolution had occurred in the scale and detail of gene data available, and Biogen was not on board. Many of those involved in R&D, particularly those focusing on gene discovery, became increasingly frustrated with the Company's conservative approach. "There was a lot of data out there . . . we had to harness this information . . . and the use of it." Start-ups began trumping Biogen researchers; for instance, during 1997-1998 the company laboriously identified and cloned a receptor called RETL3. But while they were searching for its operating mechanism and binding site, they were beaten by Washington University and by a Copenhagen company, NsGene A/S. This

---

[16] PhRMA as cited by Biogen, www.Biogen.com, January 24, 2002.

[17] Tufts Center for the Study of Drug Development, "Backgrounder: How New Drugs Move through the Development and Approval Process," November 30, 2001.

[18] These various estimates include the costs of pursuing failed candidates abandoned at various stages of the development process.

[19] Biogen was not entirely out of genomics. It had partly funded an academic research institute in Switzerland; Jurg Tschopp was developing good bioinformatic tools, particularly algorithms used to for searching for proteins.

consortium won because it had used bioinformatics, *in silico* biology, to speed discovery.[20] Biogen eventually signed a research collaboration with the Copenhagen company to develop a drug candidate, Neublastin, but NsGene retained the rights to central nervous system indications.

Mullen realized that the company had fallen way behind in informatics and genomics. Biogen had outsourced most gene expression analysis to a biotech start-up, Curagen, in 1997. Aside from a couple of young scientists poached from genomics companies, Biogen had little internal ability to access, interpret, and apply geometrically expanding gene research databases. As the information available about diseases and genes became overwhelming, Biogen was forced to retool. According to Rich Cate, "Even if we weren't going to launch a genomics effort at that point, we had to know what genes were out there to compete on a level playing field."

By 1999, Mullen had started to make some fundamental changes within the company. He spent time working with the Boston Consulting Group to develop a proactive R&D strategy. The objectives were threefold: to define what Biogen wanted its scientists to work on, ensure senior management understood what and why research was focused on certain projects, and align R&D investment with the company's business development targets.

Eventually the R&D team identified four key areas: immunology, neurodegeneration, cancer and fibrosis. Having established a set of priorities, it became easier for senior management to delegate and streamline R&D project approval, provided projects fell into one of the agreed areas. Joe Davie, Vice President of Research, led the effort initially; a team of scientists began to filter a long list of potential targets down to 93 diseases of possible interest to the company. Their triaging criteria were based on market attractiveness and Biogen's perceived competitive strengths. Each disease was then re-examined in an effort to figure out which underlying pathologies might maximize "disease reach".

The Hard Road from *in vitro* to *in silico* Biology

But before launching these new R&D efforts, the company had to restructure its information systems department (I.S.) and, more important, the attitude of both management and bench scientists towards computer aided research.

Through the late 1990s, one rarely saw PCs on senior executive desks. As is typical in many companies, the I.S. department reported to the Chief Financial Officer and researchers were viewed as secondary customers. This is because the initial use for I.S. was automating the payroll and updating budgets. A desire to keep older finance applications running meant that the company used a legacy system based on a single VAX machine. Year 2000 (Y2K) demands forced Biogen to abandon the old VAX system in favor of two UNIX servers with 8 and 10 CPUs respectively. (But even this significant increase in capacity paled in contrast to Millenium, a Biogen competitor that was estimated to have 180 CPU capacity.)[21] Biogen's two computers ran everything—Oracle databases, data queries, gene research, but there were no shared services between the two machines.

The absence of adequate information technology became a source of constant friction between R&D and I.S., particularly as biologists began to try and assemble a map of all the genes in various organisms. Even downloading on-line data was too much for this system; Biogen had to ask the

---

[20] Bioinformatics takes masses of gene and protein sequence data gathered from various life forms and then compares known gene functions, potential treatments, and biologically active binding sites. This data is sometimes complemented by combinatorial chemistry, which designs and tries millions of slightly different molecules.

[21] Interview with Michael Rosenberg, November 28, 2001.

government for custom sets of a dozen compact disks to upload data to its increasingly obsolete system. As Rich Cate noted, "Through the summer and fall of 1998 the public database information was updated so infrequently on the VAX that we were out of date often."

Whenever researchers had requested system upgrades or internet access, the finance division had applied typical financial system criteria to those requests. This often ignored the burgeoning genomics field and its strategic importance for drug candidate discovery. Meanwhile many companies entered into a gene sequencing competition, investing heavily in expensive equipment and building a bioinformatics capability. They were comparing entire sets of genes, and using rapid computing power and sophisticated algorithms to detect gene-based differences between diseased tissue and "normal" tissue.

By 1999, the strategic planning process made it clear that Research I.T. had to be upgraded and responsibilities shared. The company hired a genomics expert, Michael Gilman, who left his position as Executive Vice President and Chief Scientific Officer for ARIAD Pharmaceuticals. (Gilman had previously trained and worked for eight years at James Watson's Cold Spring Harbor Laboratory.)[22] Soon after, Rainer Fuchs was recruited from Aventis to establish a bioinformatics group. Fuchs' new area inherited a number of I.S. people and the few employees who had some bioinformatics / combinatorial chemistry experience. Fuchs also brought in six new employees. Eventually the group achieved responsibility for all R& D informatics decisions.[23]

The Bioinformatics group spent its first year building up R&D information systems and infrastructure. They created a separate Oracle database server and introduced typical application / development protocols. A Linux farm, put together with relatively inexpensive servers, increased processing capacity approximately tenfold.[24] This investment was a fraction of what it would have cost to buy a similar sized mainframe a couple of years earlier. Links to the internet were vastly increased. Michael Rosenberg, Associate Director, Bioinformatics, estimated that with its increased data processing capacity and 4 Terabytes of dedicated storage, Biogen could theoretically download the entire NCBI (Genbank public database) in one hour.[25]

In parallel, Biogen was also building up its biology-genomics tool kit. John McCoy was hired in an effort to upgrade Biogen's discovery process; he took over as vice president of discovery biology and set to work expanding the company's database access capabilities. After buying a $2 million Sun Microsystems package, he subscribed to Incyte's proprietary database. Then, he substantially upgraded its access and use of public databases.

Hiring bioinformaticians is not easy or cheap. A recent graduate with a Masters degree would sometimes cost as much as a Biogen Senior Scientist and also required providing a package of stock options. Even then, some chose a different company at the last minute. Nevertheless, in two years Biogen went from having two specialists in research informatics to twenty-five and expected to

---

[22] Watson and Crick were the discoverers of DNA see: Watson, James and Francis Crick, "A Structure for Deoxyribose Nucleic Acid," *Nature,* April 2, 1953.

[23] This group retained a dotted line of responsibility to the new Chief Information Officer, a position established in 2000.

[24] Often biotech and genomics companies that require massive computing power chose to network powerful desk workstations rather than purchase a supercomputer.

[25] All publicly funded gene sequencing data is supposed to be deposited in a centralized public databases, within twenty-four hours. This data is supposed to be freely accessible. There are three key centers, one in Japan, on in the United States and one in Europe.

double this number over the next two years. The I.S. budget for R&D swelled to over $10 million per year.[26]

Tidal Waves of Data

As the company and its systems went both digital and genomic, researchers' access to and use of computers and data sets increased exponentially. Instead of controlling or Balkanizing computers, the Linux farm and databases were open to all researchers at Biogen. To put this change into context: after years of laborious genetic research, both in house and in partnership with Curagen, Biogen's internal database held about 200 proteins. Literally overnight, simply by subscribing to Incyte's database, Biogen's database expanded to 12,000 proteins.

Given a sixty-fold magnitude shift in data availability, a few scientists thrived, but many initially chose not to access the data. Even though all two hundred scientists reporting to the head of Research had access to the new information, only about 10% chose to get training in how to use it. Perhaps ten people became active users.[27] Those who were not on-line soon found themselves falling farther and farther behind. Those who did go on-line found that new information grew geometrically.

Research management tried to help researchers by developing a targeted "scouting" system. Rather than asking everyone to go fishing in overwhelming data pools, they asked everyone what genes, proteins, or targets they were looking for. Then, using a bioinformatics querying software package they built customized "sniffer" programs. These "bots" would search the enormous amount of data posted every night on public databases looking for matches with the company criteria. If they identified a "hit" this information was passed on to a scout, a scientist, who would follow up with additional preliminary analysis. If the information seemed valuable, then the particular researcher would get a summary, a site address, and relevant links. This approach, it was hoped, would ensure the advent of genomics at Biogen was firmly tied to market based objectives, and would not become an end to itself.

At first researchers were very happy. Searches that would have taken days or weeks were now automated and updated daily. For a few weeks everything seemed fine. Then a few e-mails began coming in from individual Biogen researchers. They all had a similar message; "we are overwhelmed, please shut the spigot." But this was not to be. It was merely the beginning of an onslaught of data, because Biogen was about to incorporate in-house gene profiling capability.

As Biogen's discovery-focused teams began developing experiments based on the reams of data flowing from public databases, they realized they had to expand their in-house expression-profiling lab massively. In late 2000, a former colleague of Fuchs' from Aventis, Steve Perrin, was hired to set up and run a gene expression-profiling lab within Biogen. This was a key tool in understanding why someone gets sick or stays healthy; various genes are expressed differentially, either up regulated or down regulated, in diseased tissue.

Obtaining and interpreting a gene expression profile used to be a complex and laborious process. Working with Curagen, Biogen scientists had been focusing on around one thousand genes that were differentially expressed in diseased tissues and healthy tissues. Biogen's scientists managed six to eight projects a year and had to wait three to six months for results.

---

[26] Fuchs interview.

[27] Carulli interview.

But as Biogen upgraded its labs and began using gene chips manufactured by Affymetrix, the amount of data exploded. Each Affymetrix chip created around forty megabytes of data and about one million new data points. Given that an average experiment employed about one hundred chips, and that Biogen's new lab could do six major experiments every three months, scientists were suddenly faced with analyzing and trying to make sense of billions of new data points showing how genes react when cells are attacked by a disease. (Single, complex experiments could consume five hundred gene chips and generate half a billion data points.)

As the data yielded up masses of new targets, workflow within the molecular biology labs changed dramatically. Previously, a scientist would walk a potential medicine through the entire development process, becoming an expert on a particular gene, protein, and condition. A star scientist and his assistants "owned" a disease and the molecules used to attack it. Now lab work became so vast and overwhelming that the process of validation had to be broken down into discrete steps. Not everyone was happy with this change. As John Carulli, Senior Scientist, Discovery Biology, explained, "fifteen minutes of fame turns into fifteen seconds of fame. This has a real effect on people's morale."

As various specialized groups began moving large numbers of targets through a particular step or cluster of steps, speed increased, but there was less opportunity for cross training, particularly for lab associates, and fewer opportunities to distinguish oneself through excellent and creative lab work.

But there were also significant advantages to having complete data sets. John McCoy, Vice President of Biogen's Discovery Biology Department thought: "Before you had these large scale profiling technologies, people would tend to think in a one-dimensional way, looking at one gene at a time, coming up with a hypothesis, testing the hypothesis and refining it before jumping on to the next gene. Now at the beginning of an experiment you have the opportunity to look at everything. That's a totally different way of approaching the problem. You start with all the information, and then devise filters to get to an answer. There's a lot of filtering going on."

ERPs, PEPs, and CIPs . . .

A flood of opportunities now superseded Biogen's and Wall Street's original concern, a lack of targets in the pipeline. This meant that instead of carefully watching and nursing a few compounds along, the company would have to quickly and efficiently develop a system to triage masses of targets and compounds. Biogen built a "value chain" stage gating process, which established explicit criteria for product development. Targets went through three stages: Exploratory Research Projects (ERPs), Prospect Evaluation Projects (PEPs), and Candidate Identification Projects (CIPs) before passing to the "Pre-clinical" stage.

Teams within the ERP stage focused on discovering prospective drug targets. Typically these teams had five to seven members, drawing heavily on the new technologies of data mining and expression profiling analysis. For many projects, about 80% of the work would be *in silico* and 20% *in vitro.*[28] Biogen might sponsor seven ERPs at a time, each costing around two million dollars per year.

After a formal review, prospective targets would achieve a PEP status where the team members would focus on validating; making sure the target was biologically active, through wet lab biology. These PEP teams generally were composed of about ten people with perhaps eight projects running

[28] About four-fifths of the research, molecule studies, manipulation, and comparison would take place within a computer and various databases. One-fifth would be validated by mixing compounds and trying out lab bench experiments.

at a given time. Each project cost around four million dollars per year. Currently about 10% of this work was *in silico* and 90% *in vitro*.

If a project was successfully validated, it was designated a CIP, or drug candidate, and the team members would again change to people focused on developing and, if the compound was a protein developed in a lab animal, "humanizing" the molecule. This stage took more time than the previous stages. Typically, a CIP team would have 12-15 people and about six projects would simultaneously.[29] In total, CIP teams could include up to seventy-five people and cost close to thirty million per year.[30]

R&D's overall evaluation was measured by the amount of pre-clinical targets that were validated and handed off to product development for clinical trials. This created some tension between the various teams. Differing computer skills and aptitude meant various ERP and PEP teams progressed at different rates as they attempted to make use of the onslaught of data. Steve Perrin, head of the expression-profiling lab, believed: "we flooded [the discovery] pipeline. Some groups managed the information effectively. They usually had one or two key scientists that sat within the disease group who knew what to do with the data once it arrived on their doorstep. Other disease projects were not even close to being prepared…and they've been very slow to push the project forward."

Perrin also felt: "Unfortunately, there is a big dichotomy between the qualifications of the scientists who are willing to embrace genomics and not. The ones who embrace it quickly are the ones who have a lot of interdisciplinary skills. They are good with computers, they are not afraid of technology, and they have a good biological background. Unfortunately some people who are outstanding biologists might not be very computer savvy so they'll want to come in and do a big profiling experiment and you give them the data and they freak out, they panic, they run, and they hide. Scientists that are very good with computers; they dive right in—they ask interesting questions—they know how to deal with the data. They are not afraid to learn things they don't quite understand and they are off and running. The biggest roadblock to my group's success is helping along the people behind the curve."

Biology Bites Back . . .

By its nature, biology, and the need to test drugs *in vivo*, remains a painstaking process.[31] Even after finding a promising target, determining it is biologically active and relatively non-toxic, it takes hundreds of thousands of hours of lab work with cells, animals, and, eventually, increasing number of human subjects to validate a drug and ask for permission to market it.[32] According to John McCoy, Biogen's Vice President Discovery Biology, the real change wrought by high throughput, data

---

[29] Time and people estimates come from John McCoy, Vice President Discovery Biology.

[30] Cost figures include 50% added for overhead.

[31] And even when you find a better procedure or molecule, a better method does not necessarily lead to rapid adoption. (Amy C. Edmondson; Richard Bohmer; Gary P. Pisano, "Speeding Up Team Learning," *Harvard Business Review* (October 2001). Furthermore, many scientists are very conservative and secretive because if there is anything a scientist hates it is to be shown to be wrong in public. The community ethos was built on credibility and accuracy; in the words of Victor Koteliansky, Director of Biological Research, "every day is championship day in science. You compete every day. You cannot put out bad data. This is not like politics or finance… two mistakes and no one will speak to you." Furthermore, within the competitive and secretive world of pre-publication science, it was sometimes hard to get researchers to openly share their findings and hypotheses, even within the same company.

[32] Robert Bazell provides a good description of this process in HER-2. Robert Bazell and Amy Bernstein, *HER-2: The Making of a Revolutionary Treatment for Breast Cancer* (Diane Publishing Co., 1998).

intensive processes was to alter the discovery phase. By introducing upfront profiling, scientists in weeks could identify desirable "up- or down-regulated" genes that previously could have taken a lifetime to find.[33]

While there was value to be gained from applying genomics across the value chain, McCoy noted that the research process downstream from initial discovery was not so amenable to high throughput approaches. "Perrin's group generates more data in one afternoon than all the data Biogen had previously generated... the question is how do we turn this into knowledge."

The overall plan was that by being more comprehensive in the initial target identification that there would be higher success rates downstream. But not everyone was happy with the rapid expansion of *in silico* biology. Rich Cate, Biogen's Director, Gene Discovery, felt that "sometimes there is too much hype associated with identifying targets using transcript profiling. They assume that somehow just knowing where a gene is expressed—the value of that information is distorted to the extent that people think you can somehow get function out of it. All you know is where the gene is expressed. To the extent it is a novel protein, you still don't know what it is doing."

Many core founders of Biogen's labs thought that new data sets might be useful in certain areas, but it would be very hard to speed up the actual biological components of drug research. John McCoy put it "Ultimately your research problem will resolve down to a handful of genes, your best candidates, and then it is back to good old-fashioned bench biology."

Having masses of data implied new risks. And not all biologists were pleased with the masses of new data available; to some it represented an unwelcome distraction and risk. Scientist Matvey Lukashev thought, "you cast a bigger net and therefore your chances of picking something up increase. It speeds up the process occasionally. But it sometimes creates a mess if the experiment isn't quite designed right. . . . Every single experiment generates . . . such an incredible amount of numbers and it is so easy to get false leads, and so many of them, that you can spend your life chasing them."

As biology labs faced an ever-larger number of targets to be validated, Biogen began to face a scarcity of researchers skilled in wet biology. Director of Biological Research, Victor Koteliansky argued, "unfortunately you need to go and make wet biology and you need to go on the bench. There is no virtual discovery. . . . Discovery is in wet science or functional stuff." This meant that new discoveries were forced through a big screen. As Steve Perrin, Senior Scientist, explained, "you start off with every gene in the genome, you ask a biological question and it gets honed down to a subset, you do a transcription profiling experiments and you've gone from 50,000 genes to 3,000 genes that might be biologically relevant based on the first experiment. You annotate those and now you are only interested in a subset of 300-400, the next step would probably be cloning those, but you can't clone them all so you prioritize 50, but the antibody guy says I can only make 5 antibodies—finally until you get to the product development guys who say then can only handle four projects per year."

And every time one of these projects began, it started to consume ever-larger amounts of cash (**Exhibits 9** and **10**).

---

[33] Genes turn on or off at various points within the cells of particular tissues. You can measure which genes are turned on (up-regulated) and causing a lot of biological activity in a particular tissue. Sometimes essential genes are turned off and fail to attack a disease (down-regulated). By looking at the pattern of gene expression, scientists can sometimes tell whether diseased tissues that exhibit the same symptoms and look the same under a microscope represent different diseases. Golub, et al., "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," *Science*, vol. 286, 15 October 1999.

The Eos Project

One way to look at the impact of genomics on Biogen in more detail is to look at its breast cancer research operation. During 2000, in an effort to develop drugs to fight breast cancer, Biogen began a partnership with Eos Biotechnology. To find appropriate antibody and protein therapeutic targets, the companies compared millions of data points from diseased and normal breast tissue. Potential targets were then culled using sophisticated statistical tools.

Steve Fawell, Biogen's Associate Director, Oncology, noted that traditionally target genes would crop up in ones or twos. Three years ago Biogen's oncology program had been working on two validated targets. By the end of 2001 it had 25 validated targets, had another 20 queued up for validation, and over 400 more that required initial evaluation. Hundreds more potential targets were piling up every month.

Fawell believed that he could accelerate Phase I slightly, but then drug development became business as usual; "understanding the pre-clinical models for cancer drug development proved to be not as predictive as one would like…if you apply our usual resourcing level, and timelines, and the degree of data we would normally collect prior to going to the clinic, you simply wouldn't be able to get through that list of targets with any speed unless you invested hugely more resources than we currently have."

Evaluating a much larger target list implied hard choices in balancing risk versus speed. In an effort to keep up with exploding opportunities, the oncology group began attempting to collapse steps in the Phase I and Phase II trials. The shift in scope of Stage I clinical trials had significant implications for product development downstream. Typically, process development started considering drug candidates before they completed Phase I so as to ensure quality material was available in a timely way for a Phase II trial. Setting up these manufacturing processes is complex and expensive. This used to be justified because drug candidates were exhaustively studied before entering clinical trials. Now, with many more, less well-understood candidates entering Phase 1, it was unclear when to begin manufacturing. There was also a question as to how to make small amounts of high quality materials for an ever-growing number of targets.

Part of Biogen began testing efficacy in Phase I rather than Phase II. Because Biogen began testing drugs before stringent data verification packets were in place, the company might someday face what could become a very expensive dilemma; the Food and Drug Administration (FDA), which approves all drug sales, makes it harder and harder for a company to change protocols the farther it progresses in clinical trials.

Nevertheless, many Biogen employees were very optimistic. Even though oncology was not within the company's original area of expertise, it was an attractive market because the company could test new drugs and get them to market reasonably fast. FDA approval times are different if the disease is deadly and there are few options.[34] Furthermore, oncology research was compatible with Biogen's new investments in *in silico* biology; the biological mechanisms underlying other specialties, such as neurology, were far less understood and would therefore be harder to industrialize.

[34] Various biotech and pharmaceutical companies are pursuing similar strategies focused on oncology. In 2000, there were 396 cancer drugs in Phase II and 144 cancer drugs in Phase III. Mark P. Mathieu (editor), *Parexel's Pharmaceutical R&D Statistical Sourcebook 2001*, p. 51.

Biogen's Future . . .

Jim Mullen knew his company was changing rapidly. Biogen had incorporated a lot of new tools in a short period of time. Five years ago, a large pharmaceutical company would generate around 100 Gigabytes of data per year. Biogen was now doing this every three months.[35] He now had access to enormous data sets on gene expression, toxicology, effectiveness and surrogate markers for clinical trials. New *in silico* tools, as well as novel technologies such as protein sequencing technologies (proteomics), promised to increase incoming data by orders of magnitude yet again.

But as more drug candidates entered the front end of the drug development pipeline, enthusiasm for genomics was tempered by the bottleneck created at slow biology stages. Those at the front end of Biogen's discovery chain felt that they were producing an enormous wealth of knowledge and were frustrated by the laborious process followed in clinical trials. Those at the back end of the discovery cycle felt that bioinformaticians did not understand what they did, nor did they understand the subtlety and complexity of biological organisms.

But if Biogen were to remain competitive, it would have to continue making large investments now in unproven targets. Already during 1999-2000, the company had spent $81 million expanding its Cambridge headquarters and laboratories. With the significant increase in targets to validate, work within molecular biology labs began to change quickly; rather than nurse one or two targets through all phases of the validation process, scientists and their technicians were specializing in certain "packets" or discreet steps. (See **Exhibit 11**.) Results were measured on how many targets they could process.

Through 2002 the company was planning to add 120,000 sq. ft. of new lab space. Plans for further expansions were moving ahead rapidly. Architects and facilities managers kept a frantic pace. Instead of a dry pipeline, Biogen now faced an "embarrassment of riches," uncovered during the last R&D product review. A much larger number of drug candidates might soon enter pre-clinical and clinical evaluation. Emphasis began to shift from identifying targets towards to how to utilize genomics tools to improve the success rate of drug candidates.

Mullen's publicly stated goal was to double revenues within five years. Yet as Biogen attempted to move from a few products to masses of targets, even this jump in revenues would be a drop in the bucket compared with the pharmaceutical behemoths. (See **Exhibit 12**.) As a tidal wave of data and of new opportunities washed over Biogen, how might the CEO have to restructure various areas including manufacturing, finance, marketing, and drug development? Would doubling revenues provide sufficient scale for Biogen to prosper in this new biotech environment? Would it make sense to adopt a niche strategy or grow into the various opportunities?[36]

---

[35] Estimate by Steve Perrin. A large pharma company would now generate at least five times more data than Biogen.

[36] On a larger scale Mullen also had to consider whether the changes occurring within Biogen were occurring elsewhere, how might this alter the overall drug market? Costs? Regulation? Relative competitiveness?

Exhibit 1a Biogen, Inc. and Subsidiaries Consolidated Financial Statements

Profit and Loss Statement
(in US $ thousands, except per share amounts)

	2001	2000	1999	1998	1997	1996
Revenues						
Product	$ 971,594	$ 761,079	$ 620,636	$ 394,863	$ 239,988	$ 78,202
Royalties	$ 71,766	$ 165,373	$ 173,799	$ 162,724	$ 171,921	$ 181,502
Total Revenues	$ 1,043,360	$ 926,452	$ 794,435	$ 557,587	$ 411,909	$ 259,704
Cost and Expenses						
Cost of Revenues	$ 136,510	$ 125,198	$ 111,005	$ 74,509	$ 50,188	$ 28,525
Research and development	$ 314,556	$ 302,840	$ 221,153	$ 177,228	$ 145,501	$ 132,384
Selling, general and administrative	$ 232,096	$ 170,058	$ 146,026	$ 115,211	$ 90,098	$ 73,632
Total Cost and Expenses	$ 683,162	$ 598,096	$ 478,184	$ 366,948	$ 285,787	$ 234,541
Income from Operations	$ 360,198	$ 328,356	$ 316,251	$ 190,639	$ (45,799)	$ (156,336)
Other income (expense) net	$ 29,299	$ 158,749	$ 12,765	$ 19,554	$ 194,767	$ 197,168
Income before income tax	$ 389,497	$ 487,105	$ 329,016	$ 210,193	$ 148,968	$ 40,829
Income Taxes	$ 116,814	$ 153,528	$ 108,566	$ 71,496	$ 59,801	$ 299
Net Income	$ 272,683	$ 333,577	$ 220,450	$ 138,697	$ 89,167	$ 40,530
Basic Earnings per share	$ 1.84	$ 2.24	$ 1.47	$ 0.94		
Diluted earnings per share	$ 1.78	$ 2.16	$ 1.40	$ 0.90	$ 0.58	$ 0.28
Shares used in calculating:						
Basic earnings per share	148,355	148,743	149,788	147,537		
Diluted earnings per share	152,916	154,602	157,788	154,270	152,999	146,442

Source: Biogen, Inc. 2000 Annual Report; SEC filings.

Exhibit 1b Biogen, Inc. and Subsidiaries Consolidated Financial Statements

Condensed Consolidated Balance Sheets
(in US $ thousands)

	2001	2000	1999
Assets			
Current Assets			
Cash and cash equivalents	)	$ 48,737	$ 56,920
Cash and marketable securities	$ 798,107	$ 633,675	$ 597,619
Accounts receivable, net	$ 177,582	$ 143,178	$ 137,363
Other current assets	$ 122,038	$ 102,681	$ 118,324
Total current assets	$ 1,097,727	$ 928,271	$ 910,226
Property and equipment, net	$ 555,998	$ 400,429	$ 239,777
Other assets	$ 67,321	$ 103,156	$ 127,970
Total Assets	$ 1,721,046	$ 1,431,856	$ 1,277,973
Liabilities and Shareholders' Equity			
Current Liabilities	$ 294,942	$ 221,021	$ 190,270
Long term debt & liabilities	$ 77,272	$ 104,433	$ 108,173
Shareholders' equity	$ 1,348,832	$ 1,106,402	$ 979,530
Total Liabilities and Shareholders' Equity	$ 1,721,046	$ 1,431,856	$ 1,277,973

Sources: Biogen, Inc. 2000 Annual Report
"Biogen Delivers on Operating Earnings Guidance of $1.90 per Share After One-Time Charges," *PR Newswire*, January 24, 2002.

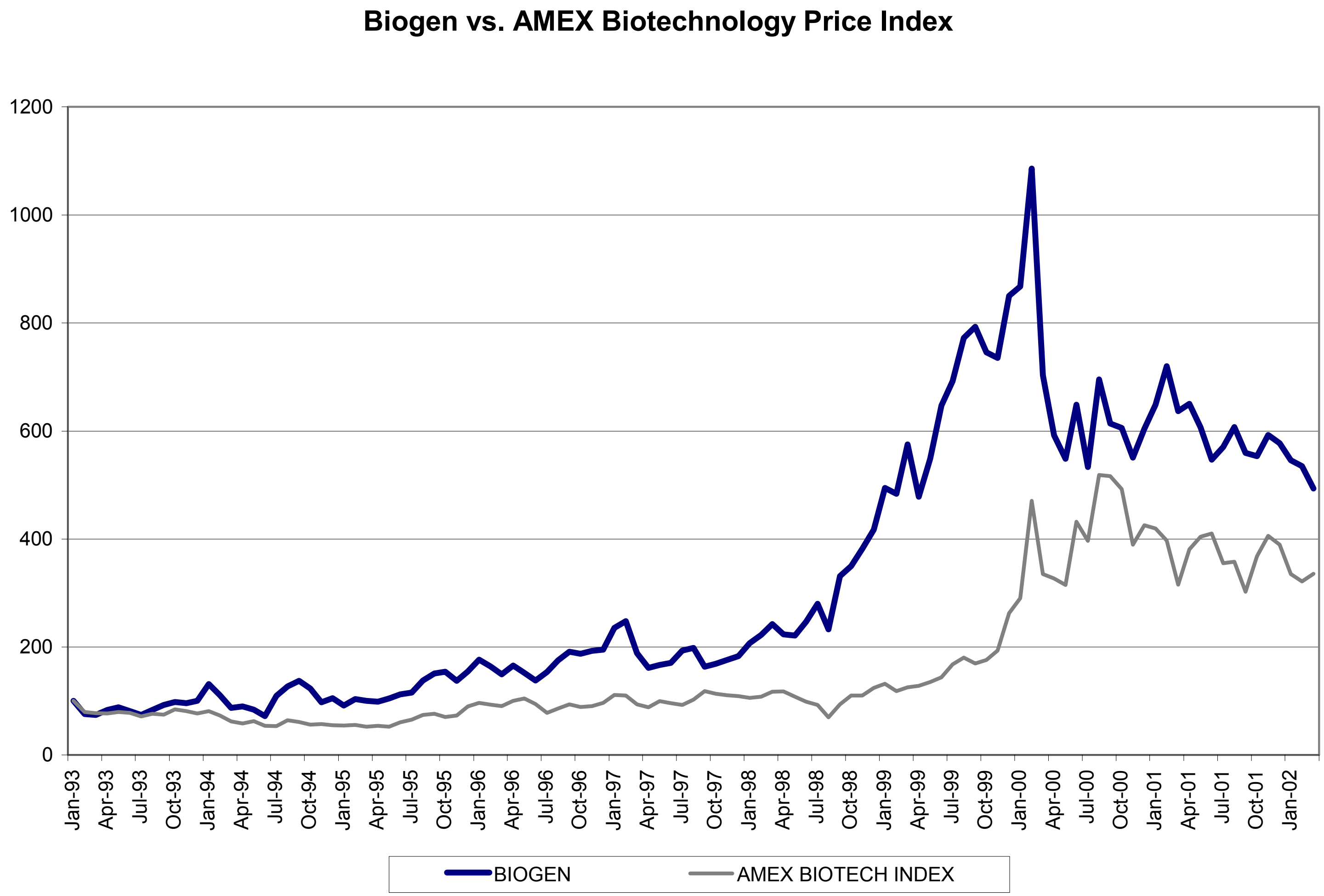
Exhibit 1c Biogen, Inc. and Subsidiaries Consolidated Financial Statements

Sources and Uses
(in US $ thousands)

For the years ended December 31,	**2000**	**1999**	**1998**
Cash Flow from Operations			
Net Income	$ 333,577	$ 220,450	$ 138,697
Adjustments to reconcile net income to net cash provided from operating activities			
Depreciation and amortization	$ 38,824	$ 31,099	$ 24,590
Other	$ (569)	$ 5,162	$ (888)
Deferred income taxes	$ 25,203	$ (23,981)	$ 7,486
Gain on sale of non-current marketable securities	$ (101,129)	-------	-------
Tax benefit of stock options	$ 81,023	$ 91,295	$ 19,595
Write-down of non-current marketable securities	-------	$ 15,287	-------
Changes in:			
Accounts receivable	$ (5,815)	$ (36,082)	$ (14,479)
Other current and other assets	$ (35,329)	$ (41,372)	$ (25,638)
Accounts payable, accrued expenses and other current and long-term liabilities	$ 30,154	$ 101,725	$ 38,077
Net cash flows from operating activities	$ 365,939	$ 363,583	$ 187,440
Cash Flow from Investing Activities			
Purchases of marketable securities	$ (627,168)	$ (1,120,218)	$ (574,021)
Proceeds from sales and maturities of marketable securities	$ 606,087	$ 1,006,465	$ 453,952
Proceeds from sales of non-current marketable securities	$ 120,199	-------	-------
Investment in collaborative partners	$ (5,000)	$ (10,000)	$ (5,000)
Acquisitions of property and equipment	$ (194,402)	$ (82,528)	$ (29,049)
Additions to patents			
Net cash flows from investing activities	$ (104,997)	$ (210,080)	$ (158,680)
Cash Flow from Financing Activities			
Repayments on note payable	-------	-------	$ (24,817)
Repayments on long-term debt	$ (4,888)	$ (4,887)	$ (4,886)
Purchases of treasury stock	$ (300,192)	$ (197,717)	$ (65,550)
Proceeds from put warrants	-------	$ 22,086	-------
Issuance of common stock and option exercises	$ 35,955	$ 58,490	$ 21,580
Net cash flows from financing activities	$ (269,125)	$ (122,028)	$ (73,673)
Net increase (decrease) in cash and cash equivalents	$ (8,183)	$ 31,475	$ (44,913)
Cash and cash equivalents, beginning of the year	$ 56,920	$ 25,445	$ 70,358
Cash and cash equivalents, end of the year	$ 48,737	$ 56,920	$ 25,445
Supplemental Cash Flow Data			
Cash paid during the year for:			
Interest	$ 4,314	$ 4,598	$ 5,909
Income taxes	$ 42,683	$ 4,787	$ 35,828

Source: Biogen, Inc. 2000 Annual Report.

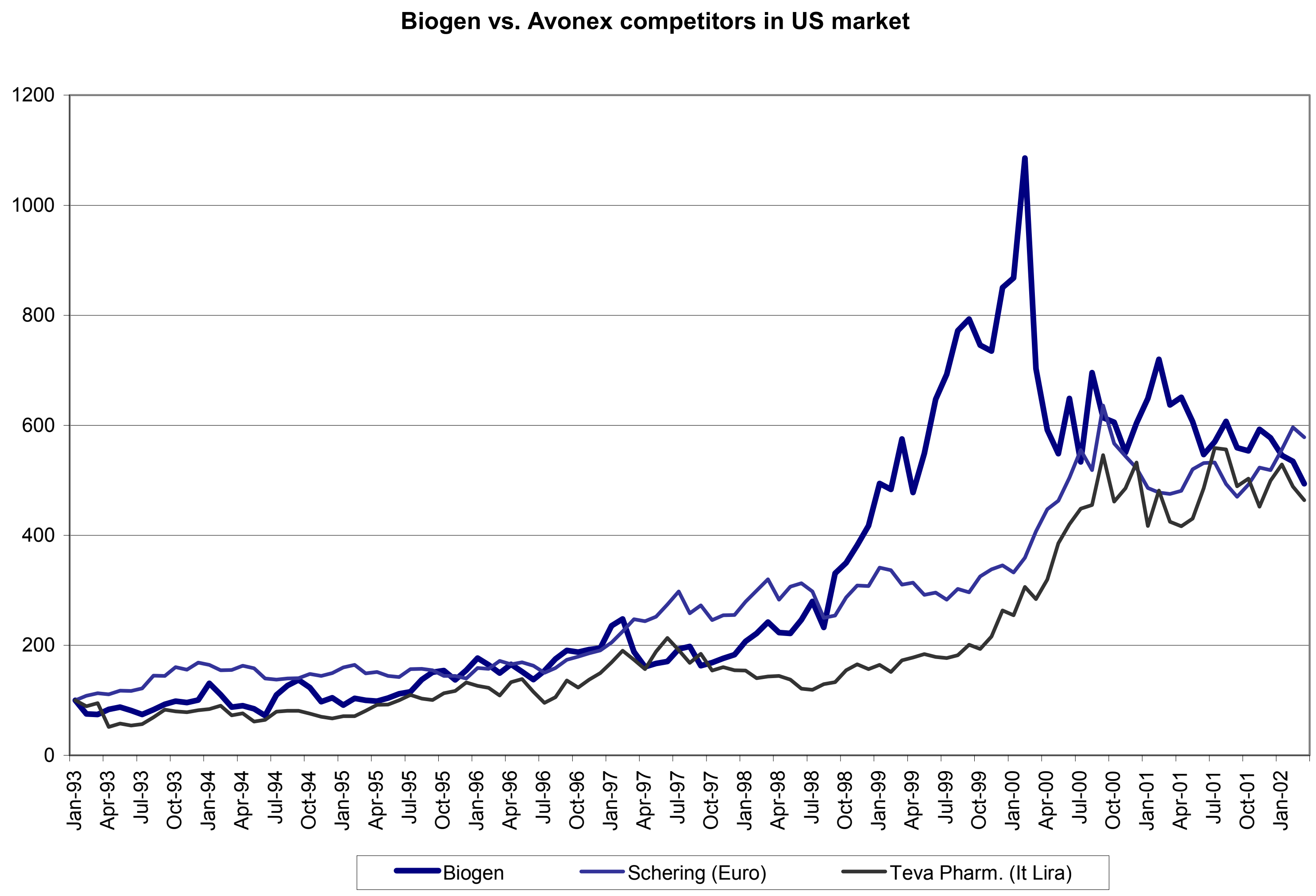
Exhibit 2a Biogen's Stock Performance Compared with Biotechnology Index and Direct Competitors



Source: Datastream.

Note: Prices have been indexed to 100.

Exhibit 2b Biogen's Stock Performance Compared with Biotechnology Index and Direct Competitors



Source: Datastream

Note: Prices have been indexed to 100.

Exhibit 3 Top Worldwide Biotech and Pharmaceutical Acquisitions 2000-to-Date

Date Effective	Acquiror Name	Target Name	Target Business Description		Value of Transaction ($mil)
06/19/00	Pfizer Inc	Warner-Lambert Co	Health care,consumer products	$	89,168
12/27/00	Glaxo Wellcome PLC	SmithKline Beecham PLC	Manufacture pharmaceuticals	$	75,961
03/31/00	Monsanto Co	Pharmacia & Upjohn Inc	Mnfr pharmaceutical prods	$	26,486
06/22/01	Johnson & Johnson	ALZA Corp	Manufacture pharmaceuticals	$	11,070
10/02/01	Bristol-Myers Squibb Co	DuPont Pharmaceuticals Co	Mnfr,whl pharmaceutical prods	$	7,800
03/02/01	Abbott Laboratories	Knoll AG(BASF AG)	Manufacture pharmaceuticals	$	6,900
08/06/01	Shareholders	Zimmer Holdings Inc	Mnfr,whl orthopedic implants	$	5,801
10/17/00	Tyco International Ltd	Mallinckrodt Inc	Mnfr diagnostic products	$	4,393
04/25/00	Linde AG	AGA AB	Mnfr industrial,medical gases	$	4,083
05/11/01	Shire Pharmaceuticals Group	BioChem Pharma Inc	Mnfr pharmaceuticals	$	3,748
08/31/00	King Pharmaceuticals Inc	Jones Pharmaceutical Inc	Mnfr drug products, vitamins	$	3,523
08/28/01	Medtronic Inc	MiniMed Inc	Mnfr,whl microinfusion systems	$	3,304
03/29/00	Cia di Partecipazioni Assicura	Montedison(Cie de Partecipazi)	Mnfr chemicals,pharmaceuticals	$	3,115
02/12/02	Millennium Pharmaceuticals Inc	COR Therapeutics Inc	Mnfr pharmaceuticals	$	2,417
07/10/00	Koninklijke Numico NV	Rexall Sundown Inc	Mnfr,whl vitamins	$	1,768
11/10/00	Elan Corp PLC	Dura Pharmaceuticals Inc	Mnfr pharmaceuticals	$	1,708
08/01/01	Koninklijke Philips Electronic	Agilent Tech-Healthcare Solut	Mnfr medical monitoring equip	$	1,700
01/15/02	MedImmune Inc	Aviron	Mnfr biological products	$	1,665

Exhibit 3 (continued)

Date Effective	Acquiror Name	Target Name	Target Business Description		Value of Transaction ($mil)
03/16/00	Adecco SA	Olsten Corp	Pvd temporary staffing svcs	$	1,475
09/14/00	Invitrogen Corp	Dexter Corp	Mnfr adhesives,coatings	$	1,468
04/27/01	Triad Hospitals Inc	Quorum Health Group Inc	Own,op acute care hospitals	$	1,400
01/31/02	WellPoint Health Networks Inc	RightCHOICE Managed Care Inc	Own,operate HMOs,PPOs	$	1,347
11/21/01	Johnson & Johnson	Inverness Medical-Diabetes	Mnfr diabetes-treatment prod	$	1,300
01/16/01	SmithKline Beecham PLC	Block Drug Co	Mnfr dental prod,pharmaceutica	$	1,235
09/28/01	Mayne Nickless Ltd	FH Faulding & Co Ltd	Mnfr,whl drugs,toiletries	$	1,218
11/10/00	GN Store Nord A/S	Photonetics SA	Mnfr photonic instruments	$	1,190
02/28/01	Sasol Ltd	Condea Chemie GmbH(RWE)	Mnfr chemicals and elastomers	$	1,142
03/11/02	L-3 Communications Holdings	Raytheon Co-Aircraft Intgrtion	Mnfr military aircraft	$	1,130
10/01/01	Welfide Corp	Mitsubishi-Tokyo Pharm	Mnfr synthetic organic fibers	$	1,129
10/19/01	Koninklijke Philips Electronic	Marconi plc-Medical Operations	Mnfr diagnostic imaging equip	$	1,100
12/14/00	Genzyme Corp	GelTex Pharmaceuticals Inc	Mnfr pharmaceuticals	$	1,052
10/03/00	Advance Paradigm Inc	PCS Health Systems(Rite Aid)	Pvd drug benefit mgmt svcs	$	1,022
09/01/00	Alcon Laboratories Inc(Nestle)	Summit Autonomous Inc	Mnfr,whl opthalmic laser sys	$	929
08/28/00	Watson Pharmaceuticals Inc	Schein Pharmaceutical Inc	Manfacture pharmaceuticals	$	916
01/24/00	Celltech Chiroscience PLC	Medeva PLC	Pvd medical research services	$	914
12/22/00	Corixa Corp	Coulter Pharmaceuticals Inc	Mnfr cancer pharmaceuticals	$	854
03/31/00	Shareholders	Baxter-Cardiovascular Bus	Mnfr cardiovascular prods	$	798
10/03/00	Ciba Vision Corp(Ciba-Geigy)	Wesley Jessen(Bain Capital)	Mnfr ophthalmic goods, equip	$	759
09/28/01	Investor Group	Carter-Wallace Inc-Consumer	Mnfr consumer products	$	739
09/22/00	Chiron Corp	PathoGenesis Corp	Mnfr pharmaceuticals	$	700

Source: Adapted from SDC Mergers and Acquisitions.

Exhibit 4 Multiple Sclerosis Population

Patient Population

Region	Estimated Population with MS
U.S.	350,000
Europe	350,000
Rest of World (ROW)	1,300,000

Source: Adapted from Morgan Stanley Research, Biogen Report, "MS Franchise Provides Near-Term Opportunity," September 18, 2001.

Biogen's AVONEX is approved for Relapsing Remitting Multiple Sclerosis. There are several other forms of the disease.

Categories of Multiple Sclerosis	% of US Disease Population (estimated 350K people)
Relapsing Remitting MS	45.0
Secondary Progressive MS	25.0
Primary Progressive MS	10.0
Relapsing Progressive MS	10.0
Monosymptomatic MS	10.0

Source: Adapted from Morgan Stanley Research, Biogen Report, "MS Franchise Provides Near-Term Opportunity," September 18, 2001.

Note: Drugs must be approved by the FDA for each category of multiple sclerosis.

Relapsing Remitting MS is characterized by "attacks " of the disease on nerve myelin coverings followed by periods of remission. *Secondary Progressive MS* is experienced similar to RRMS, although the nerves sustain more damage and this form of the disease is more difficult to treat than RRMS. *Primary Progressive MS* is characterized by a gradual decline in nerve function without acute attacks, and does not respond well to existing treatments. *Monosymptomatic MS* is characterized by a single attack by the disease. Biogen's CHAMPS study demonstrated efficacy in preventing recurrence of attacks in Monosymptomatic MS patients.

Exhibit 5 Multiple Sclerosis Market

Estimated Market Size: $2 billion worldwide in 2001, rising to $4 billion by 2005.

Source: Merrill Lynch Analyst Report.

Market Share

United States	
Biogen	53%
Teva	28%
Schering	19%
Total	100%

Europe	
Biogen	34%
Teva	5%
Schering	34%
Serono	27%
Total	100%

Source: Adapted from IMS Health via Morgan Stanley Research; Morgan Stanley Research Biogen Report, September 18, 2001.

"MS Franchise Provides Near-Term Opportunity," Caroline L. Copithorne, et al.

N.B. AVONEX was developed for the U.S. market under the Orphan Drug program. The underlying patent and the orphan drug status both expire in 2003, at which time Serono was to be allowed to enter the U.S. market with their version of a similar drug, Rebif. In early March 2002, the FDA determined that Serono would be permitted to sell Rebif in the U.S. Market affective immediately.

Company	Sales	R&D Exp.	Market Value
For Year Ending December 31, 2000			
(US $ Millions)			
Schering AB	$ 4,316	$ 1,267	$ 11,274
Serono S.A.	$ 1,147	$ 263	$ 21,700
Teva Pharmaceuticals	$ 1,750	$ 105	$ 9,370
Biogen	$ 927	$ 303	$ 8,878

Source: Hoover's Online; SEC filings.

Exhibit 6 Two Views on the Costs and Uncertainties of Drug Development

Stage	Avg. Duration (years)	% success	% cumulative success	Avg. $ per stage (US $ mil)	% R&D exp per stage
	Probabilities of Success by Phase				
Target Identification	1			165.0	19%
Target Validation	2	5	4.5	205.0	23%
Screening	0.4	13	5.8	40.0	5%
Optimization	2.7	53	3.0	120.0	14%
Pre Clinical	1.6	42	1.3	90.0	10%
Clinical	7	23	0.3	260.0	30%
Total	14.7			880.0	100%

Source: Adapted from "A Revolution in R&D: How Genomics and Genetics Are Transforming the Biopharmaceutical Industry," BCG Report November 2001. % success per stage derived from bar charts in Exhibit 5, p. 20. Duration and Cost per Stage from Exhibit 2, p. 12.

Costs include cost of failed projects.

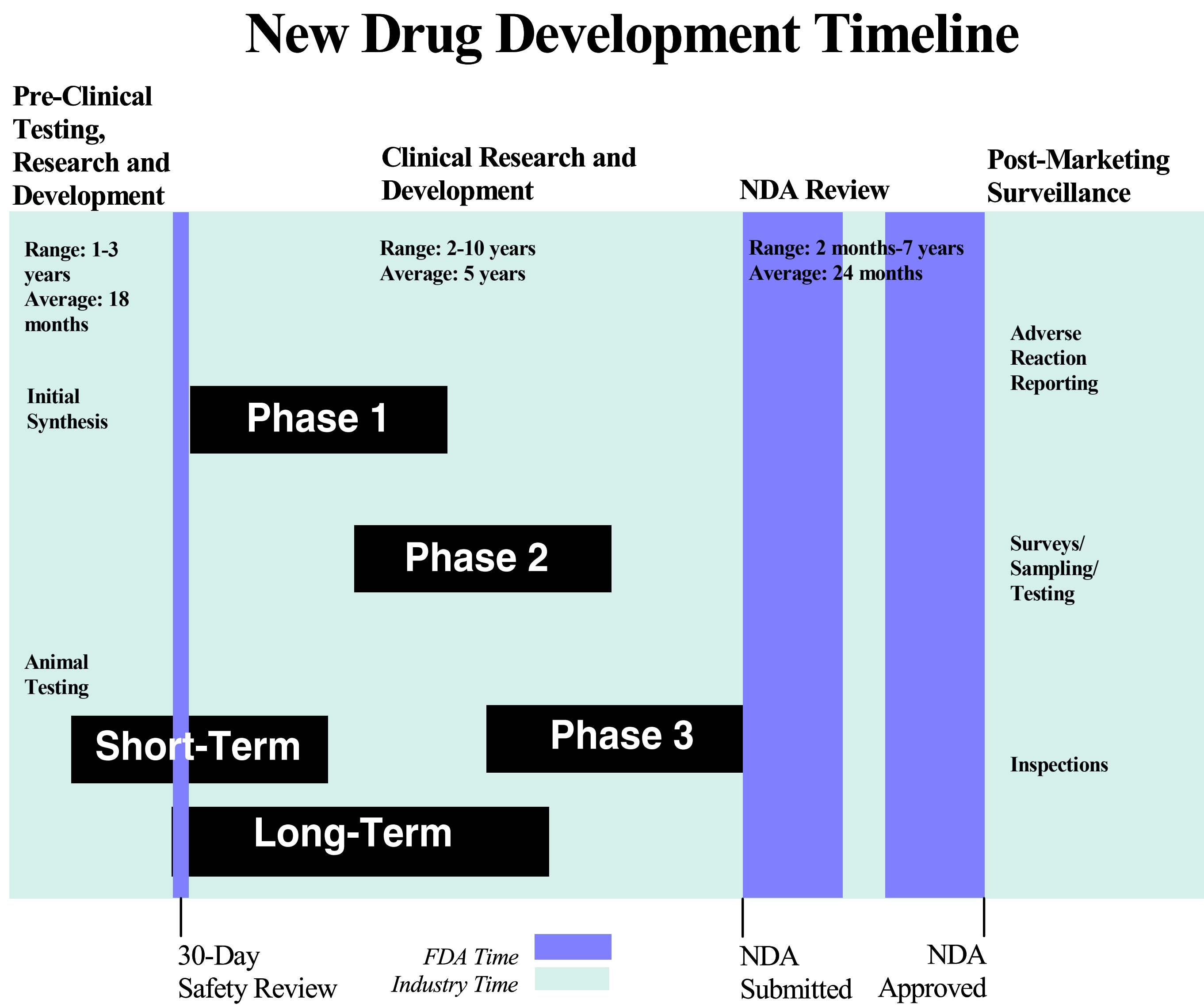
Stage		Avg. Duration (years)	% success	% cumulative success		Avg. $ per stage (US $ mil)	% R&D exp per stage
		Probabilities of Success by Phase					
Target Validation	)		30-35	30	)		
Hit Generation	)	2.5	90	27	)	60.4	4%
Lead optimization	))		90	24	))		
Biological validation	))	3	75	18	))	226.5	15%
Pre Clinical		1	50	9		151.0	10%
Phase 1		1.5	70	6		226.5	15%
Phase 2		2	50	3		332.2	22%
Phase 3		2.5	70	2		468.1	31%
FDA filing		1.5	90	2		45.3	3%
Total		14				1510.0	100%

Sources: Adapted from "The Fruits of Genomics," Lehman Brothers Study, Slide 3, p. 22. "The Fruits of Genomics II Analyst Forum," January 2001, Lehman Brothers Study. Note: $ per stage amounts derived using percents from Figure 22, p. 46, and total % per drug from Figure 24, p. 48.

N.B. Tufts Center for the Study of Drug Development announced that the average cost to develop a new drug is $802 million. This figure included the cost of project failures, the impact of long development times on investment costs, and was based on surveys of 10 drug companies.

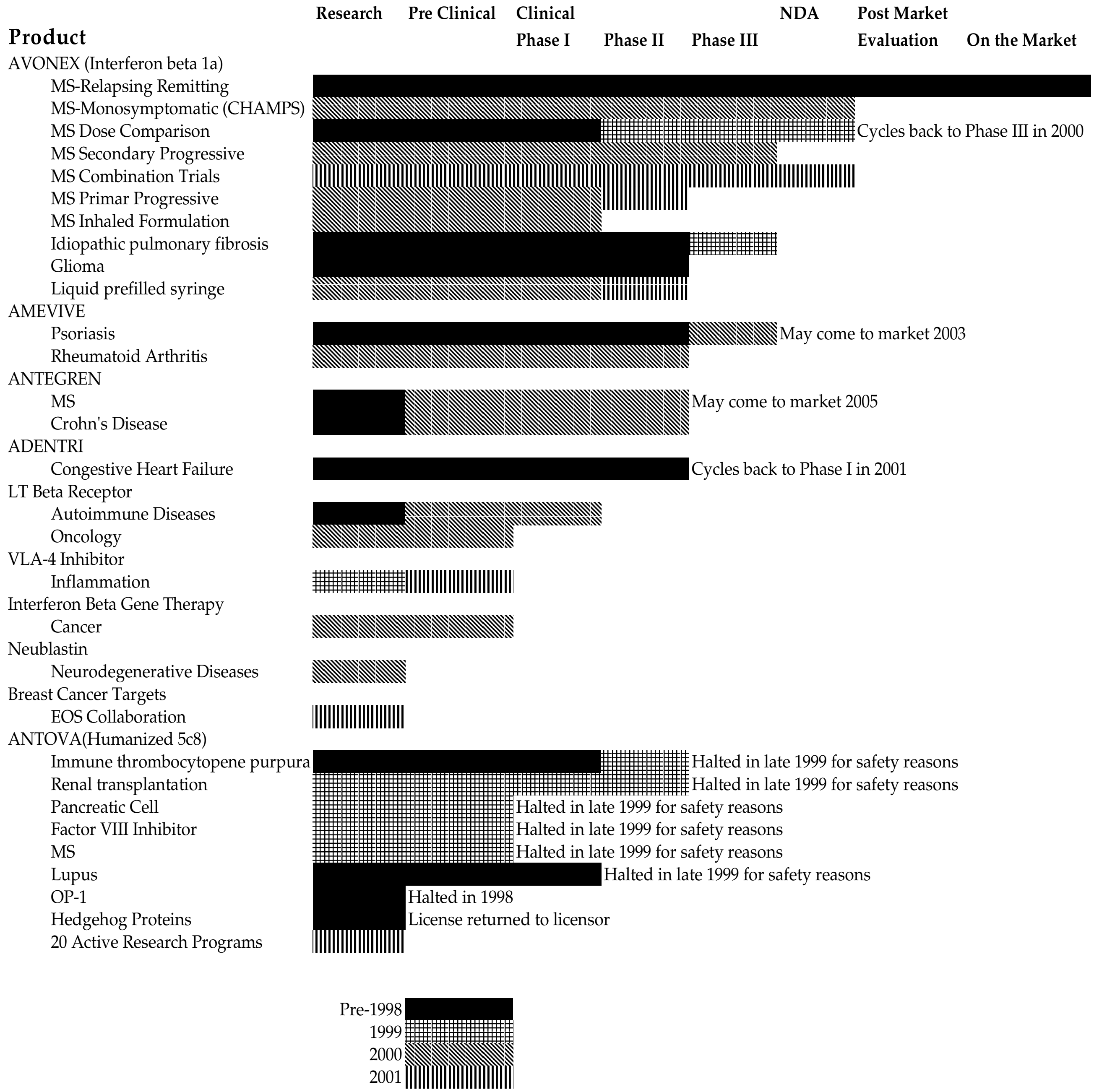
Source: Tufts Center for the Study of Drug Development press release, November 30, 2001.

Exhibit 7



Source: http://www.fda.gov/fdac/graphics/newdrugspecial/drugchart.pdf.

Exhibit 8 Biogen, Inc. Product Pipeline Evolution



Source: Compiled from Biogen, Inc. Annual Reports.

Exhibit 9 Breakdown of Biogen's R&D Expenditures

Estimated Distribution of R&D Spending by Development Phase

Development Stage	% R&D Spending*	Biogen 2001 (USD Mil)	# projects beg. 2001	$/project (USD Mil)
Basic Research	4%	$ 12.58	NA	
Discovery	15%	$ 47.18	NA	
Preclinical Development	10%	$ 31.46	2	$ 15.73
Clinical Development				
Phase 1	15%	$ 47.18	5	$ 9.44
Phase 2	22%	$ 69.20	3	$ 23.07
Phase 3	31%	$ 97.51	3	$ 32.50
FDA filing through approval	3%	$ 9.44	1	$ 9.44
Total	100%	$ 314.6		

Source: Created by casewriter based on data from "The Fruits of Genomics II," Lehman Brothers, January 2001. Biogen, Inc. Annual Report and company product pipeline data at www.Biogen.com, January 2001.

*Breakdown of R&D spending per "The Fruits of Genomics II," p. 46. Chart cites McKinsey & Co., Lehman Brothers, PhRMA and FDA as sources.

Biogen Expense by stage based on Lehman percentages. However, in company interviews, Biogen estimated it spends 30% of its R&D budget on research prior to the preclinical stage.

Biogen 2001 R&D expenditure preliminary figure released January 2002.

Exhibit 10 Biogen's Cost of Drug Development

Cost of R&D to bring 2000 pipeline to market

Projects per Phase per Year
(Full Year Equivalents)

Stage	Duration (years)	Prob of success	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Pre clinic	1.5	0.8	1.00	0.50								
Phase 1	1.5	0.7	4.00	2.40	0.80	0.00						
Phase 2	2	0.5	2.00	3.40	2.80	1.96	0.56	0.00				
Phase 3	2.5	0.7	8.00	8.00	5.00	1.70	1.90	1.68	0.28	0.14		
FDA filing	1.5	0.9	3.00	1.50	2.80	5.60	0.35		0.98	0.20		
Market				1.35			5.04	0.63		0.88	0.18	
Total	9											

Methodology: These figures were not provided by the company but reflect the casewriter's estimates based on industry information. Progress of drug candidates noted in Biogen 2000 Annual Report were tracked forward and updated to end of 2001. Projections utilizing industry probabilities and durations were then applied to this portfolio.

Duration: Durations for each phase from "The Fruits of Genomics II Analyst Forum", January 30, 2001, Fig. 22, p. 46. Duration does not include discovery stage activities.

Probability of success derived from the following sources:
Pre clinical: Company interviews
Clinical by Phase & FDA : "The Fruits of Genomics", Lehman Brothers, January 30, 2001, p. 22.

N.B. Five of the Phase 3 projects in 2002 are Avonex product extensions.

Cost per project per stage per year adjusted for inflation

Stage	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Pre clinic	$ 15,730	$ 16,517	$ 17,342	$ 18,209	$ 19,120	$ 20,076	$ 21,080	$ 22,134	$ 23,240	$ 24,402
Phase 1	$ 9,440	$ 9,912	$ 10,408	$ 10,928	$ 11,474	$ 12,048	$ 12,651	$ 13,283	$ 13,947	$ 14,645
Phase 2	$ 23,070	$ 24,224	$ 25,435	$ 26,706	$ 28,042	$ 29,444	$ 30,916	$ 32,462	$ 34,085	$ 35,789
Phase 3	$ 32,500	$ 34,125	$ 35,831	$ 37,623	$ 39,504	$ 41,479	$ 43,553	$ 45,731	$ 48,017	$ 50,418
NDA	$ 9,440	$ 9,912	$ 10,408	$ 10,928	$ 11,474	$ 12,048	$ 12,651	$ 13,283	$ 13,947	$ 14,645
Market										

Inflation factor 0.05
Per project per stage figures were derived by applying percentages from Exhibit 9 to each stage and dividing by the number of projects.

Exhibit 10 (continued)

Estimated Cost of R&D to bring 2000 pipeline to market
(US$ '000)

Stage	Duration years	Prob of success	2002	2003	2004	2005	2006	2007	2008	2009	2010
Pre clinic	1.5	0.8	$ 15,730	$ 8,258	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Phase 1	1.5	0.7	$ 37,760	$ 23,789	$ 8,326	$ -	$ -	$ -	$ -	$ -	$ -
Phase 2	2	0.4	$ 46,140	$ 82,360	$ 71,217	$ 52,345	$ 15,703	$ -	$ -	$ -	$ -
Phase 3	2.5	0.7	$ 260,000	$ 273,000	$ 179,156	$ 63,959	$ 75,058	$ 69,685	$ 12,195	$ 6,402	$ -
NDA	1.5	0.12	$ 28,320	$ 14,868	$ 29,141	$ 61,197	$ 4,016	$ -	$ 12,397	$ 2,603	$ -
Market			$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total			$ 387,950	$ 402,275	$ 287,841	$ 177,500	$ 94,777	$ 69,685	$ 24,592	$ 9,006	$ -

Total required to pursue existing pipeline	$ 1,453,626	R&D investment required to support 2000 pipeline through launch
	$ 979,110	R&D investment from launch of Avonex up to 2000
	$ 2,432,736	Total R&D invested in new drugs since AVONEX launched

Source: Casewriter's estimates based on industry information.

Exhibit 11 Discovery and Exploration Process for the EOS Project

1)	Tissue Access	
2)	Running Affymetrix chips	Done by EOS
		Biogen developed access to tissue; some affymetrix chips run in-house
3)	Expression Mining	This step is a handoff from Affymetrix Lab / Bioinformatics specialists and wet lab scientists
4)	Sequence Mining	An overlapping team of “wet lab” scientists and “in silico” focused Scientists select best targets
5)	Cloning Group	Makes full length CDNA for every target selected
6)	Expression Group	Takes clones and puts into other vectors to support protein production
7)	Protein Purification Group	
8)	Monoclonal Antibody Facility	Inject protein into mice
9)	Pharmacology Group	Xenograft models (human tumors) test for efficacy
10)	Humanize antibody	
11)	Put Humanized Antibody into biology	Verify antibody still effective
12)	If Successful, Candidate Passes to Pre-Clinical Drug Development Stage	

Source: Company interview.

Exhibit 12 Top Pharmaceutical and Biotechnology Companies Ranked by Market Value End 2000 (in US$ millions)

	Sales/ Turnover (Net)	Market Value
PFIZER INC	$ 29,574	$ 290,444
MERCK & CO	$ 40,363	$ 216,049
GLAXOSMITHKLINE PLC	$ 27,413	$ 175,768
JOHNSON & JOHNSON	$ 29,139	$ 146,135
BRISTOL MYERS SQUIBB	$ 18,216	$ 144,440
NOVARTIS AG	$ 21,226	$ 115,214
LILLY (ELI) & CO	$ 10,862	$ 104,747
ASTRAZENECA PLC	$ 18,103	$ 89,034
AMERICAN HOME PRODUCTS CORP	$ 13,263	$ 83,363
SCHERING-PLOUGH	$ 9,815	$ 83,025
PHARMACIA CORP	$ 18,144	$ 79,074
ABBOTT LABORATORIES	$ 13,746	$ 74,881
AMGEN INC	$ 3,629	$ 66,329
SANOFI-SYNTHELABO	$ 5,510	$ 48,757
TAKEDA CHEMICAL INDUSTRIES	$ 8,729	$ 42,932
GENENTECH INC	$ 1,646	$ 42,826
IMMUNEX CORP	$ 862	$ 21,972
TEVA PHARMACEUTICALS	@NA	$ 17,686
ELAN CORP PLC	$ 1,302	$ 15,175
MILLENNIUM PHARMACTCLS INC	$ 196	$ 13,240
ALLERGAN INC	$ 1,626	$ 12,748
YAMANOUCHI PHARMACEUT CO LTD	$ 4,148	$ 12,450
NOVO NORDISK A/ S	$ 2,580	$ 11,597
SCHERING AG	$ 4,151	$ 11,246
SERONO SA	$ 1,147	$ 11,223
FOREST LABORATORIES	$ 1,175	$ 10,462
MEDIMMUNE INC	$ 540	$ 10,079
IDEC PHARMACEUTICALS CORP	$ 155	$ 9,280
SANKYO CO LTD (PHARMACEUTCL)	$ 4,938	$ 9,041
BIOGEN INC	$ 926	$ 8,879
KING PHARMACEUTICALS INC	$ 620	$ 8,830
HUMAN GENOME SCIENCES INC	$ 22	$ 8,677
GENZYME CORP	$ 752	$ 8,588

Source: Adapted by casewriter from Compustat Global Vantage.

Kuhn, T. 2012. The Structure of Scientific Revolutions. Chapter 12: The Resolution of Revolutions

This page intentionally left blank

XII. The Resolution of Revolutions

The textbooks we have just been discussing are produced only in the aftermath of a scientific revolution. They are the bases for a new tradition of normal science. In taking up the question of their structure we have clearly missed a step. What is the process by which a new candidate for paradigm replaces its predecessor? Any new interpretation of nature, whether a discovery or a theory, emerges first in the mind of one or a few individuals. It is they who first learn to see science and the world differently, and their ability to make the transition is facilitated by two circumstances that are not common to most other members of their profession. Invariably their attention has been intensely concentrated upon the crisis-provoking problems; usually, in addition, they are men so young or so new to the crisis-ridden field that practice has committed them less deeply than most of their contemporaries to the world view and rules determined by the old paradigm. How are they able, what must they do, to convert the entire profession or the relevant professional subgroup to their way of seeing science and the world? What causes the group to abandon one tradition of normal research in favor of another?

To see the urgency of those questions, remember that they are the only reconstructions the historian can supply for the philosopher's inquiry about the testing, verification, or falsification of established scientific theories. In so far as he is engaged in normal science, the research worker is a solver of puzzles, not a tester of paradigms. Though he may, during the search for a particular puzzle's solution, try out a number of alternative approaches, rejecting those that fail to yield the desired result, he is not testing the *paradigm* when he does so. Instead he is like the chess player who, with a problem stated and the board physically or mentally before him, tries out various alternative moves in the search for a solution. These trial attempts, whether by the chess player or by the scientist, are trials only of themselves, not of the rules of the game. They are possible only so long as the paradigm itself is taken for granted. Therefore, paradigm-testing occurs only after persistent failure to solve a noteworthy puzzle has given rise to crisis. And even then it occurs only after the sense of crisis has evoked an alternate candidate for paradigm. In the sciences the testing situation never consists, as puzzle-solving does, simply in the comparison of a single paradigm with nature. Instead, testing occurs as part of the competition between two rival paradigms for the allegiance of the scientific community.

Closely examined, this formulation displays unexpected and probably significant parallels to two of the most popular contemporary philosophical theories about verification. Few philosophers of science still seek absolute criteria for the verification of scientific theories. Noting that no theory can ever be exposed to all possible relevant tests, they ask not whether a theory has been verified but rather about its probability in the light of the evidence that actually exists. And to answer that question one important school is driven to compare the ability of different theories to explain the evidence at hand. That insistence on comparing theories also characterizes the historical situation in which a new theory is accepted. Very probably it points one of the directions in which future discussions of verification should go.

In their most usual forms, however, probabilistic verification theories all have recourse to one or another of the pure or neutral observation-languages discussed in Section X. One probabilistic theory asks that we compare the given scientific theory with all others that might be imagined to fit the same collection of observed data. Another demands the construction in imagination of all the tests that the given scientific theory might conceivably be asked to pass.[1] Apparently some such construction is necessary for the computation of specific probabilities, absolute or relative, and it is hard to see how such a construction can

[1] For a brief sketch of the main routes to probabilistic verification theories, see Ernest Nagel, *Principles of the Theory of Probability*, Vol. I, No. 6, of *International Encyclopedia of Unified Science*, pp. 60–75.

possibly be achieved. If, as I have already urged, there can be no scientifically or empirically neutral system of language or concepts, then the proposed construction of alternate tests and theories must proceed from within one or another paradigm-based tradition. Thus restricted it would have no access to all possible experiences or to all possible theories. As a result, probabilistic theories disguise the verification situation as much as they illuminate it. Though that situation does, as they insist, depend upon the comparison of theories and of much widespread evidence, the theories and observations at issue are always closely related to ones already in existence. Verification is like natural selection: it picks out the most viable among the actual alternatives in a particular historical situation. Whether that choice is the best that could have been made if still other alternatives had been available or if the data had been of another sort is not a question that can usefully be asked. There are no tools to employ in seeking answers to it.

A very different approach to this whole network of problems has been developed by Karl R. Popper who denies the existence of any verification procedures at all.[2] Instead, he emphasizes the importance of falsification, i.e., of the test that, because its outcome is negative, necessitates the rejection of an established theory. Clearly, the role thus attributed to falsification is much like the one this essay assigns to anomalous experiences, i.e., to experiences that, by evoking crisis, prepare the way for a new theory. Nevertheless, anomalous experiences may not be identified with falsifying ones. Indeed, I doubt that the latter exist. As has repeatedly been emphasized before, no theory ever solves all the puzzles with which it is confronted at a given time; nor are the solutions already achieved often perfect. On the contrary, it is just the incompleteness and imperfection of the existing data-theory fit that, at any time, define many of the puzzles that characterize normal science. If any and every failure to fit were ground for theory rejection, all theories ought to be rejected at all times. On the other hand, if only severe failure

[2] K. R. Popper, *The Logic of Scientific Discovery* (New York, 1959), esp. chaps. i–iv.

to fit justifies theory rejection, then the Popperians will require some criterion of "improbability" or of "degree of falsification." In developing one they will almost certainly encounter the same network of difficulties that has haunted the advocates of the various probabilistic verification theories.

Many of the preceding difficulties can be avoided by recognizing that both of these prevalent and opposed views about the underlying logic of scientific inquiry have tried to compress two largely separate processes into one. Popper's anomalous experience is important to science because it evokes competitors for an existing paradigm. But falsification, though it surely occurs, does not happen with, or simply because of, the emergence of an anomaly or falsifying instance. Instead, it is a subsequent and separate process that might equally well be called verification since it consists in the triumph of a new paradigm over the old one. Furthermore, it is in that joint verification-falsification process that the probabilist's comparison of theories plays a central role. Such a two-stage formulation has, I think, the virtue of great verisimilitude, and it may also enable us to begin explicating the role of agreement (or disagreement) between fact and theory in the verification process. To the historian, at least, it makes little sense to suggest that verification is establishing the agreement of fact with theory. All historically significant theories have agreed with the facts, but only more or less. There is no more precise answer to the question whether or how well an individual theory fits the facts. But questions much like that can be asked when theories are taken collectively or even in pairs. It makes a great deal of sense to ask which of two actual and competing theories fits the facts *better*. Though neither Priestley's nor Lavoisier's theory, for example, agreed precisely with existing observations, few contemporaries hesitated more than a decade in concluding that Lavoisier's theory provided the better fit of the two.

This formulation, however, makes the task of choosing between paradigms look both easier and more familiar than it is. If there were but one set of scientific problems, one world within which to work on them, and one set of standards for their solution, paradigm competition might be settled more or less routinely by some process like counting the number of problems solved by each. But, in fact, these conditions are never met completely. The proponents of competing paradigms are always at least slightly at cross-purposes. Neither side will grant all the non-empirical assumptions that the other needs in order to make its case. Like Proust and Berthollet arguing about the composition of chemical compounds, they are bound partly to talk through each other. Though each may hope to convert the other to his way of seeing his science and its problems, neither may hope to prove his case. The competition between paradigms is not the sort of battle that can be resolved by proofs.

We have already seen several reasons why the proponents of competing paradigms must fail to make complete contact with each other's viewpoints. Collectively these reasons have been described as the incommensurability of the pre- and postrevolutionary normal-scientific traditions, and we need only recapitulate them briefly here. In the first place, the proponents of competing paradigms will often disagree about the list of problems that any candidate for paradigm must resolve. Their standards or their definitions of science are not the same. Must a theory of motion explain the cause of the attractive forces between particles of matter or may it simply note the existence of such forces? Newton's dynamics was widely rejected because, unlike both Aristotle's and Descartes's theories, it implied the latter answer to the question. When Newton's theory had been accepted, a question was therefore banished from science. That question, however, was one that general relativity may proudly claim to have solved. Or again, as disseminated in the nineteenth century, Lavoisier's chemical theory inhibited chemists from asking why the metals were so much alike, a question that phlogistic chemistry had both asked and answered. The transition to Lavoisier's paradigm had, like the transition to Newton's, meant a loss not only of a permissible question but of an achieved solution. That loss was not, however, permanent either. In the twentieth century questions about the qualities of chemical substances have entered science again, together with some answers to them.

More is involved, however, than the incommensurability of standards. Since new paradigms are born from old ones, they ordinarily incorporate much of the vocabulary and apparatus, both conceptual and manipulative, that the traditional paradigm had previously employed. But they seldom employ these borrowed elements in quite the traditional way. Within the new paradigm, old terms, concepts, and experiments fall into new relationships one with the other. The inevitable result is what we must call, though the term is not quite right, a misunderstanding between the two competing schools. The laymen who scoffed at Einstein's general theory of relativity because space could not be "curved"—it was not that sort of thing—were not simply wrong or mistaken. Nor were the mathematicians, physicists, and philosophers who tried to develop a Euclidean version of Einstein's theory.[3] What had previously been meant by space was necessarily flat, homogeneous, isotropic, and unaffected by the presence of matter. If it had not been, Newtonian physics would not have worked. To make the transition to Einstein's universe, the whole conceptual web whose strands are space, time, matter, force, and so on, had to be shifted and laid down again on nature whole. Only men who had together undergone or failed to undergo that transformation would be able to discover precisely what they agreed or disagreed about. Communication across the revolutionary divide is inevitably partial. Consider, for another example, the men who called Copernicus mad because he proclaimed that the earth moved. They were not either just wrong or quite wrong. Part of what they meant by 'earth' was fixed position. Their earth, at least, could not be moved. Correspondingly, Copernicus' innovation was not simply to move the earth. Rather, it was a whole new way of regarding the problems of physics and astronomy,

[3] For lay reactions to the concept of curved space, see Philipp Frank, *Einstein, His Life and Times*, trans. and ed. G. Rosen and S. Kusaka (New York, 1947), pp. 142–46. For a few of the attempts to preserve the gains of general relativity within a Euclidean space, see C. Nordmann, *Einstein and the Universe*, trans. J. McCabe (New York, 1922), chap. ix.

one that necessarily changed the meaning of both 'earth' and 'motion.'[4] Without those changes the concept of a moving earth was mad. On the other hand, once they had been made and understood, both Descartes and Huyghens could realize that the earth's motion was a question with no content for science.[5]

These examples point to the third and most fundamental aspect of the incommensurability of competing paradigms. In a sense that I am unable to explicate further, the proponents of competing paradigms practice their trades in different worlds. One contains constrained bodies that fall slowly, the other pendulums that repeat their motions again and again. In one, solutions are compounds, in the other mixtures. One is embedded in a flat, the other in a curved, matrix of space. Practicing in different worlds, the two groups of scientists see different things when they look from the same point in the same direction. Again, that is not to say that they can see anything they please. Both are looking at the world, and what they look at has not changed. But in some areas they see different things, and they see them in different relations one to the other. That is why a law that cannot even be demonstrated to one group of scientists may occasionally seem intuitively obvious to another. Equally, it is why, before they can hope to communicate fully, one group or the other must experience the conversion that we have been calling a paradigm shift. Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time, forced by logic and neutral experience. Like the gestalt switch, it must occur all at once (though not necessarily in an instant) or not at all.

How, then, are scientists brought to make this transposition? Part of the answer is that they are very often not. Copernicanism made few converts for almost a century after Copernicus' death. Newton's work was not generally accepted, particularly on the Continent, for more than half a century after the *Prin-*

[4] T. S. Kuhn, *The Copernican Revolution* (Cambridge, Mass., 1957), chaps. iii, iv, and vii. The extent to which heliocentrism was more than a strictly astronomical issue is a major theme of the entire book.

[5] Max Jammer, *Concepts of Space* (Cambridge, Mass., 1954), pp. 118–24.

cipia appeared.[6] Priestley never accepted the oxygen theory, nor Lord Kelvin the electromagnetic theory, and so on. The difficulties of conversion have often been noted by scientists themselves. Darwin, in a particularly perceptive passage at the end of his *Origin of Species*, wrote: "Although I am fully convinced of the truth of the views given in this volume . . . , I by no means expect to convince experienced naturalists whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. . . . [B]ut I look with confidence to the future,—to young and rising naturalists, who will be able to view both sides of the question with impartiality."[7] And Max Planck, surveying his own career in his *Scientific Autobiography*, sadly remarked that "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."[8]

These facts and others like them are too commonly known to need further emphasis. But they do need re-evaluation. In the past they have most often been taken to indicate that scientists, being only human, cannot always admit their errors, even when confronted with strict proof. I would argue, rather, that in these matters neither proof nor error is at issue. The transfer of allegiance fom paradigm to paradigm is a conversion experience that cannot be forced. Lifelong resistance, particularly from those whose productive careers have committed them to an older tradition of normal science, is not a violation of scientific standards but an index to the nature of scientific research itself. The source of resistance is the assurance that the older paradigm will ultimately solve all its problems, that nature can be shoved

[6] I. B. Cohen, *Franklin and Newton: An Inquiry into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof* (Philadelphia, 1956), pp. 93–94.

[7] Charles Darwin, *On the Origin of Species* . . . (authorized edition from 6th English ed.; New York, 1889), II, 295–96.

[8] Max Planck, *Scientific Autobiography and Other Papers*, trans. F. Gaynor (New York, 1949), pp. 33–34.

into the box the paradigm provides. Inevitably, at times of revolution, that assurance seems stubborn and pigheaded as indeed it sometimes becomes. But it is also something more. That same assurance is what makes normal or puzzle-solving science possible. And it is only through normal science that the professional community of scientists succeeds, first, in exploiting the potential scope and precision of the older paradigm and, then, in isolating the difficulty through the study of which a new paradigm may emerge.

Still, to say that resistance is inevitable and legitimate, that paradigm change cannot be justified by proof, is not to say that no arguments are relevant or that scientists cannot be persuaded to change their minds. Though a generation is sometimes required to effect the change, scientific communities have again and again been converted to new paradigms. Furthermore, these conversions occur not despite the fact that scientists are human but because they are. Though some scientists, particularly the older and more experienced ones, may resist indefinitely, most of them can be reached in one way or another. Conversions will occur a few at a time until, after the last holdouts have died, the whole profession will again be practicing under a single, but now a different, paradigm. We must therefore ask how conversion is induced and how resisted.

What sort of answer to that question may we expect? Just because it is asked about techniques of persuasion, or about argument and counterargument in a situation in which there can be no proof, our question is a new one, demanding a sort of study that has not previously been undertaken. We shall have to settle for a very partial and impressionistic survey. In addition, what has already been said combines with the result of that survey to suggest that, when asked about persuasion rather than proof, the question of the nature of scientific argument has no single or uniform answer. Individual scientists embrace a new paradigm for all sorts of reasons and usually for several at once. Some of these reasons—for example, the sun worship that helped make Kepler a Copernican—lie outside the apparent sphere of science entirely.[9] Others must depend upon idiosyncrasies of autobiography and personality. Even the nationality or the prior reputation of the innovator and his teachers can sometimes play a significant role.[10] Ultimately, therefore, we must learn to ask this question differently. Our concern will not then be with the arguments that in fact convert one or another individual, but rather with the sort of community that always sooner or later re-forms as a single group. That problem, however, I postpone to the final section, examining meanwhile some of the sorts of argument that prove particularly effective in the battles over paradigm change.

Probably the single most prevalent claim advanced by the proponents of a new paradigm is that they can solve the problems that have led the old one to a crisis. When it can legitimately be made, this claim is often the most effective one possible. In the area for which it is advanced the paradigm is known to be in trouble. That trouble has repeatedly been explored, and attempts to remove it have again and again proved vain. "Crucial experiments"—those able to discriminate particularly sharply between the two paradigms—have been recognized and attested before the new paradigm was even invented. Copernicus thus claimed that he had solved the long-vexing problem of the length of the calendar year, Newton that he had reconciled terrestrial and celestial mechanics, Lavoisier that he had solved the problems of gas-identity and of weight relations, and Einstein that he had made electrodynamics compatible with a revised science of motion.

Claims of this sort are particularly likely to succeed if the new paradigm displays a quantitative precision strikingly better than

[9] For the role of sun worship in Kepler's thought, see E. A. Burtt, *The Metaphysical Foundations of Modern Physical Science* (rev. ed.; New York, 1932), pp. 44–49.

[10] For the role of reputation, consider the following: Lord Rayleigh, at a time when his reputation was established, submitted to the British Association a paper on some paradoxes of electrodynamics. His name was inadvertently omitted when the paper was first sent, and the paper itself was at first rejected as the work of some "paradoxer." Shortly afterwards, with the author's name in place, the paper was accepted with profuse apologies (R. J. Strutt, 4th Baron Rayleigh, *John William Strutt, Third Baron Rayleigh* [New York, 1924], p. 228).

its older competitor. The quantitative superiority of Kepler's Rudolphine tables to all those computed from the Ptolemaic theory was a major factor in the conversion of astronomers to Copernicanism. Newton's success in predicting quantitative astronomical observations was probably the single most important reason for his theory's triumph over its more reasonable but uniformly qualitative competitors. And in this century the striking quantitative success of both Planck's radiation law and the Bohr atom quickly persuaded many physicists to adopt them even though, viewing physical science as a whole, both these contributions created many more problems than they solved.[11]

The claim to have solved the crisis-provoking problems is, however, rarely sufficient by itself. Nor can it always legitimately be made. In fact, Copernicus' theory was not more accurate than Ptolemy's and did not lead directly to any improvement in the calendar. Or again, the wave theory of light was not, for some years after it was first announced, even as successful as its corpuscular rival in resolving the polarization effects that were a principal cause of the optical crisis. Sometimes the looser practice that characterizes extraordinary research will produce a candidate for paradigm that initially helps not at all with the problems that have evoked crisis. When that occurs, evidence must be drawn from other parts of the field as it often is anyway. In those other areas particularly persuasive arguments can be developed if the new paradigm permits the prediction of phenomena that had been entirely unsuspected while the old one prevailed.

Copernicus' theory, for example, suggested that planets should be like the earth, that Venus should show phases, and that the universe must be vastly larger than had previously been supposed. As a result, when sixty years after his death the telescope suddenly displayed mountains on the moon, the phases of Venus, and an immense number of previously unsuspected stars,

[11] For the problems created by the quantum theory, see F. Reiche, *The Quantum Theory* (London, 1922), chaps. ii, vi–ix. For the other examples in this paragraph, see the earlier references in this section.

those observations brought the new theory a great many converts, particularly among non-astronomers.[12] In the case of the wave theory, one main source of professional conversions was even more dramatic. French resistance collapsed suddenly and relatively completely when Fresnel was able to demonstrate the existence of a white spot at the center of the shadow of a circular disk. That was an effect that not even he had anticipated but that Poisson, initially one of his opponents, had shown to be a necessary if absurd consequence of Fresnel's theory.[13] Because of their shock value and because they have so obviously not been "built into" the new theory from the start, arguments like these prove especially persuasive. And sometimes that extra strength can be exploited even though the phenomenon in question had been observed long before the theory that accounts for it was first introduced. Einstein, for example, seems not to have anticipated that general relativity would account with precision for the well-known anomaly in the motion of Mercury's perihelion, and he experienced a corresponding triumph when it did so.[14]

All the arguments for a new paradigm discussed so far have been based upon the competitors' comparative ability to solve problems. To scientists those arguments are ordinarily the most significant and persuasive. The preceding examples should leave no doubt about the source of their immense appeal. But, for reasons to which we shall shortly revert, they are neither individually nor collectively compelling. Fortunately, there is also another sort of consideration that can lead scientists to reject an old paradigm in favor of a new. These are the arguments, rarely made entirely explicit, that appeal to the individual's sense of the appropriate or the aesthetic—the new theory is said to be "neater," "more suitable," or "simpler" than the old. Probably

[12] Kuhn, *op. cit.*, pp. 219–25.

[13] E. T. Whittaker, *A History of the Theories of Aether and Electricity*, I (2d ed.; London, 1951), 108.

[14] See *ibid.*, II (1953), 151–80, for the development of general relativity. For Einstein's reaction to the precise agreement of the theory with the observed motion of Mercury's perihelion, see the letter quoted in P. A. Schilpp (ed.), *Albert Einstein, Philosopher-Scientist* (Evanston, Ill., 1949), p. 101.

such arguments are less effective in the sciences than in mathematics. The early versions of most new paradigms are crude. By the time their full aesthetic appeal can be developed, most of the community has been persuaded by other means. Nevertheless, the importance of aesthetic considerations can sometimes be decisive. Though they often attract only a few scientists to a new theory, it is upon those few that its ultimate triumph may depend. If they had not quickly taken it up for highly individual reasons, the new candidate for paradigm might never have been sufficiently developed to attract the allegiance of the scientific community as a whole.

To see the reason for the importance of these more subjective and aesthetic considerations, remember what a paradigm debate is about. When a new candidate for paradigm is first proposed, it has seldom solved more than a few of the problems that confront it, and most of those solutions are still far from perfect. Until Kepler, the Copernican theory scarcely improved upon the predictions of planetary position made by Ptolemy. When Lavoisier saw oxygen as "the air itself entire," his new theory could cope not at all with the problems presented by the proliferation of new gases, a point that Priestley made with great success in his counterattack. Cases like Fresnel's white spot are extremely rare. Ordinarily, it is only much later, after the new paradigm has been developed, accepted, and exploited that apparently decisive arguments—the Foucault pendulum to demonstrate the rotation of the earth or the Fizeau experiment to show that light moves faster in air than in water—are developed. Producing them is part of normal science, and their role is not in paradigm debate but in postrevolutionary texts.

Before those texts are written, while the debate goes on, the situation is very different. Usually the opponents of a new paradigm can legitimately claim that even in the area of crisis it is little superior to its traditional rival. Of course, it handles some problems better, has disclosed some new regularities. But the older paradigm can presumably be articulated to meet these challenges as it has met others before. Both Tycho Brahe's earth-centered astronomical system and the later versions of the phlogiston theory were responses to challenges posed by a new candidate for paradigm, and both were quite successful.[15] In addition, the defenders of traditional theory and procedure can almost always point to problems that its new rival has not solved but that for their view are no problems at all. Until the discovery of the composition of water, the combustion of hydrogen was a strong argument for the phlogiston theory and against Lavoisier's. And after the oxygen theory had triumphed, it could still not explain the preparation of a combustible gas from carbon, a phenomenon to which the phlogistonists had pointed as strong support for their view.[16] Even in the area of crisis, the balance of argument and counterargument can sometimes be very close indeed. And outside that area the balance will often decisively favor the tradition. Copernicus destroyed a time-honored explanation of terrestrial motion without replacing it; Newton did the same for an older explanation of gravity, Lavoisier for the common properties of metals, and so on. In short, if a new candidate for paradigm had to be judged from the start by hard-headed people who examined only relative problem-solving ability, the sciences would experience very few major revolutions. Add the counterarguments generated by what we previously called the incommensurability of paradigms, and the sciences might experience no revolutions at all.

But paradigm debates are not really about relative problem-solving ability, though for good reasons they are usually couched in those terms. Instead, the issue is which paradigm should in the future guide research on problems many of which neither competitor can yet claim to resolve completely. A decision between alternate ways of practicing science is called for, and in the circumstances that decision must be based less on

[15] For Brahe's system, which was geometrically entirely equivalent to Copernicus', see J. L. E. Dreyer, *A History of Astronomy from Thales to Kepler* (2d ed.; New York, 1953), pp. 359–71. For the last versions of the phlogiston theory and their success, see J. R. Partington and D. McKie, "Historical Studies of the Phlogiston Theory," *Annals of Science*, IV (1939), 113–49.

[16] For the problem presented by hydrogen, see J. R. Partington, *A Short History of Chemistry* (2d ed.; London, 1951), p. 134. For carbon monoxide, see H. Kopp, *Geschichte der Chemie*, III (Braunschweig, 1845), 294–96.

past achievement than on future promise. The man who embraces a new paradigm at an early stage must often do so in defiance of the evidence provided by problem-solving. He must, that is, have faith that the new paradigm will succeed with the many large problems that confront it, knowing only that the older paradigm has failed with a few. A decision of that kind can only be made on faith.

That is one of the reasons why prior crisis proves so important. Scientists who have not experienced it will seldom renounce the hard evidence of problem-solving to follow what may easily prove and will be widely regarded as a will-o'-the-wisp. But crisis alone is not enough. There must also be a basis, though it need be neither rational nor ultimately correct, for faith in the particular candidate chosen. Something must make at least a few scientists feel that the new proposal is on the right track, and sometimes it is only personal and inarticulate aesthetic considerations that can do that. Men have been converted by them at times when most of the articulable technical arguments pointed the other way. When first introduced, neither Copernicus' astronomical theory nor De Broglie's theory of matter had many other significant grounds of appeal. Even today Einstein's general theory attracts men principally on aesthetic grounds, an appeal that few people outside of mathematics have been able to feel.

This is not to suggest that new paradigms triumph ultimately through some mystical aesthetic. On the contrary, very few men desert a tradition for these reasons alone. Often those who do turn out to have been misled. But if a paradigm is ever to triumph it must gain some first supporters, men who will develop it to the point where hardheaded arguments can be produced and multiplied. And even those arguments, when they come, are not individually decisive. Because scientists are reasonable men, one or another argument will ultimately persuade many of them. But there is no single argument that can or should persuade them all. Rather than a single group conversion, what occurs is an increasing shift in the distribution of professional allegiances.

At the start a new candidate for paradigm may have few supporters, and on occasions the supporters' motives may be suspect. Nevertheless, if they are competent, they will improve it, explore its possibilities, and show what it would be like to belong to the community guided by it. And as that goes on, if the paradigm is one destined to win its fight, the number and strength of the persuasive arguments in its favor will increase. More scientists will then be converted, and the exploration of the new paradigm will go on. Gradually the number of experiments, instruments, articles, and books based upon the paradigm will multiply. Still more men, convinced of the new view's fruitfulness, will adopt the new mode of practicing normal science, until at last only a few elderly hold-outs remain. And even they, we cannot say, are wrong. Though the historian can always find men—Priestley, for instance—who were unreasonable to resist for as long as they did, he will not find a point at which resistance becomes illogical or unscientific. At most he may wish to say that the man who continues to resist after his whole profession has been converted has *ipso facto* ceased to be a scientist.

Beyond the Numbers: Building Your Qualitative Intelligence

This page intentionally left blank

The world is not responding to our attempts to control it with quantitative models. Our chaotic environment demands a new approach that pays attention to *qualities* in addition to *quantities.*

Beyond the Numbers: Building Your Qualitative Intelligence

by **Roger Martin**

WE LIVE IN A WORLD THAT IS PREOCCUPIED with predictability and enraptured with quantitative analysis. In the realm of business, the evidence is clear: forecasters crank out precise predictions of economic growth using massive econometric models, and CEOs give to-the-penny guidance to capital markets on the next quarter's predicted earnings. In the world of health care, geneticists sequence the human genome and predict the elimination of numerous diseases. In our day-to-day lives, we are governed by adages like "Show me the numbers" and "If you can't measure it, it doesn't count."

Where has this obsession gotten us? Not far, I would argue; the economists, for one, have gotten it consistently wrong. As late as the middle of 2008, none of the world's leading macroeconomists or forecasting organizations were predicting that the economy would shrink that year, let alone that it would crater as disastrously as it did. But undaunted, the same economists who totally missed the recession turned back to the same quantitative, scientific models to predict how the economy would recover, only to be mainly wrong again.

Likewise, CEOs continue to give quarterly guidance based on their sophisticated financial planning systems and they, too, continue to be wrong – getting slammed not for bad performance but for their failure to predict performance exactly months ahead of time. And in the health care realm, once the human genome was sequenced, the scientists who predicted the solving of the world's medical mysteries had to admit that the entire project raised more questions about the complicated interaction among genes than it answered.

Our deep-seated desire to quantify the world is not surprising: given the complexity we face on a regular basis, we naturally seek ways to understand and control whatever we can. But the world is not responding well to our attempts, making it clear that it refuses to be organized, understood and controlled in a purely quantitative fashion. Not unlike a spirited race horse, the harder we tug on the reins to control it, the more it resists the bit to prove a point: it is not a machine that responds the way a user manual suggests it will; rather, it is an intensely complicated, ambiguous system of systems that defies comprehensive quantification. Such an environment demands a lesser-known form of analysis; one that pays attention to *qualities* rather than just *quantities*.

The Quantitative vs. the Qualitative Paradigm

There is no denying the strengths of the quantitative paradigm: when used in a suitable context, its methods produce reliable data. However, the paradigm breaks down when the phenomenon under study is either complex or ambiguous to any significant degree. The greatest weakness of the quantitative approach is that it decontextualizes human behaviour, removing an event from its real-world setting and ignoring the effects of variables not included in the model.

It would be lovely if, when setting out to predict Gross Domestic Product, macroeconomists could simply add up all existing loans to determine 'credit outstanding' and then plug those quantities into an economic model. But it isn't that easy. Unfortunately, as we found out in 2008, not all loans are the same, and some – especially subprime mortgages – aren't worth the proverbial paper they're written on. Life would be so much easier if, when trying to predict next quarter's sales, CEOs could simply extrapolate from last month's sales. But it may well turn out that those sales aren't as solid a base for growth as we'd hope – especially if the customer relationships underpinning them are weak. Likewise, it would have been fantastic if, after the genome was sequenced, scientists could have said 'This particular gene causes Lymphoma and this one causes Alzheimer's.' But it didn't turn out that way.

The fundamental shortcoming in each of these examples is that they depend entirely on quantities to produce answers, and in so doing, they are blissfully ignorant of qualities. As education scholar **Elliot Eisner** has pointed out, "Not everything that matters can be measured, and not everything that can be measured matters." Unlike quantities, qualities cannot be objectively measured. We can count the number of people in a room, but that tells us nothing about the qualities – upbeat, flat, intense, contentious – of the group's interaction. We need to better understand the role of qualities to help us navigate the ambiguous and uncertain world in which we live.

Unfortunately, garnering an understanding of and a facility with qualities is a real challenge, given the sharp swing that our formal education system has taken toward quantitative thinking. As Rotman Adjunct Professor **Hilary Austen** has pointed out: "Results on tests with right and wrong answers clearly rank us against other members of our class. Quantitative thinking allows us to be precise and to share understanding; we use it to define fairness and rationality and effectiveness, and it is this utility that has led so many people to equate it with intelligence."

The allure of defining intelligence as equivalent to quantitative thinking has led higher education to force millions of students to take the SAT, MCAT, LSAT or GMAT to earn their spots in a university or graduate school. Whether purportedly testing Mathematics or English, these tests are dominantly assessments of single-answer problem solving. And despite their rather famous inability to predict anything about the likely life-performance of the test taker, these tests are sacrosanct. Prospective students tend to become a single-point, uni-dimensional creature in the eyes of admissions departments – as in "She *is* a 750 GMAT", not "She scored 750 on her GMAT."

Unfortunately, that is just the beginning of this obsession for students. Typically, once the walls of higher education are breached, yet more quantitative tools, models and methodologies are poured into the minds of our students, and more single-point-answer testing ensues to ensure that the quantitative toolbox provided is both usable and well-used. As a result, the majority of students graduate capable of using their newly-gained conceptual knowledge as a recipe or paint-by-numbers kit to produce analyses that tell them what is right and wrong, true or false. And if they happen to graduate in one of the so-called STEM fields (Science, Technology, Engineering or Mathematics) the **National Research Council** (NRC) and the **National Science Foundation** (NSF) will jump for joy because these are the fields that they have declared to be of most import to a country's competitiveness in the world.

The NRC and NSF may indeed be right that we are producing too few STEM graduates – though there is (ironically) no particular quantitative evidence given to buttress the claim. Not surprisingly, whenever comparisons are made, the relative quantities of STEM graduates are presented with no regard to their qualities. But if indeed the challenges of the world are increasingly going to feature ambiguity, complexity, uniqueness and indeterminacy – which all signs point to – STEM training *per se* will not be our salvation. A different form of intelligence needs to be built up to handle the unique problems we face, and that is qualitative intelligence.

As Hilary Austen describes it: "Whenever you allow action to generate outcomes rather than use action to pursue pre-established goals; whenever you reason with sensory experience rather than with abstract symbols; whenever you act without hesitation with

I have become utterly convinced that we need to train and develop 'business artists' more than we need to develop business analysts.

what you know, while courting the possibilities of surprise; and whenever you use a combination of immediate and remembered experience to predict and then revise immediate action – you are exercising qualitative intelligence."

For Austen and others, including the late management scholar **Donald Schön,** qualitative intelligence is synonymous with artistry and is at the very heart of excellence in any profession. As Schön once said, "The artistry of painters, sculptors, musicians, dancers and designers bears a strong family resemblance to the artistry of extraordinary lawyers, physicians, managers and teachers. It is no accident that professionals often refer to an 'art' of teaching or management."

A Growing Demand for Artistic Capacity

Effectively dealing with the challenges of the modern world – rather than with narrow sub-segments of them – demands artistic capacity. Without the explicit development of qualitative thought, sophisticated mental operations like judgment in the face of uncertainty, coping with ambiguity, balancing consequences, and responding effectively to surprise will remain elusive. No matter what we do for living, we need to be able to go beyond using our knowledge as a recipe and aim higher than crunching quantitative data to produce single-point answers.

To some, this may seem an extreme claim, and of course, there are many legitimate and conflicting views on the subject. Implicitly, institutions such as the NRC and NSF see the situation differently, and countless popular management books have been written exhorting business people to become even more analytical and quantitative in their approaches. But from my vantage point of business strategy and management education, I have become utterly convinced that we need to train and develop 'business artists' more than we need to develop business analysts.

While it has not always been a popular view, more and more business executives are seeing the shortcomings of a quantitatively-obsessed world. In June of 2008, I interviewed **Scott Cook,** the cerebral founder of **Intuit,** on stage at a design conference in San Francisco. In a reflective mood, he freely admitted: "What I learned about the old style of business – about analytic and deductive models – [is that] it's no longer up to the task. I find it destroying value, instead of creating value." Cook was not disparaging his company's ability to crunch the numbers, but indicating that it just wasn't enough anymore: his people needed to observe more, and to let themselves be surprised.

Cook shared the story of the origins of QuickBooks, the product that transformed his company and drove its tremendous growth. Prior to QuickBooks, Intuit's leading product was Quicken, a personal financial management program. Early consumer research showed that half of Quicken users were using it not at home, but at the office. It was a curious finding, one the company at first chose to explain away as meaningless or just plain wrong. "We ignored it," Cook says "and continued on our merry way."

But the same strange result kept showing up and after a few years, Cook decided to stop ignoring it. He set out to investigate, visiting and talking to users to figure out what was going on. His team discovered that there was indeed a large segment of customers using the product in ways Intuit had not intended. These were small businesses that chose to use the simple, friendly Quicken personal software instead of complex professional accounting software. Conventional wisdom was that these businesses wanted and needed traditional debit-and-credit accounting software – and so nothing else was available to them – until Quicken.

Faced with this new understanding, Intuit could have continued ahead as it had before, focusing on the home user and enjoying the additional boon of these newly-discovered office users. Instead, Cook and his team chose to embrace the surprise. "I'm a big believer in savoring surprises," he says. "If something emerges that's really a surprise – upside or downside – generally, this is the real world speaking to you, saying there's something that you don't yet understand. And in that thing that you don't yet understand could come a major mindset change, a paradigm shift that could rock your world."

So it was for Intuit. It created a Quicken-like, easy-to-use financial management software specifically targeted to the needs of small businesses, and within a month, it was the market leader. Only by

In Design, there is never one right answer. It's a very different view from Science.

cultivating surprise and being willing to imagine new and different conceptions of its customer was Intuit able to build on and maintain its success. Cook summed up the transformation as follows:

> "We've had to swing the pendulum drastically from a deliberative, top-down model with lots of debates and PowerPoint presentations to an emergent model – where solutions and decisions emerge from individual action based on observation and experimentation, not based on what the boss says. And the more I see patterns in successful firms, the more I see that this is actually underneath it all."

At the same conference, I held an on-stage discussion with **Claudia Kotchka**, who masterminded a similar pendulum swing at **Procter & Gamble.** In explaining how the most qualitatively sensitive designers differed from her more quantitatively-driven colleagues at P&G, she commented: "(Designers) don't look for 'the' answer, because in Design, there is no one right answer. It's a very different view from Science."

What are the implications of the insights of Austen, Eisner, Schön, Cook and Kotchka for business and for management education? The key message for business is to learn to appreciate qualities. That means not obsessing about measurement so much that you exclude essential but un-measurable qualities from your understanding of a given situation. Consider the possibility that if you can't measure something, it might very well be the most important aspect of the problem on which you are working. Across industries, business people must strive to understand the qualities as much as the quantities of the environment in which they work.

In closing

I will leave you with three pieces of advice regarding the qualitative paradigm. First, I hope it is now clear that decision making is not only about equations and symbols. We must use all of our senses as we form opinions and make decisions. Numbers can help to describe sensory experience, but they cannot serve as a substitute for it. Former P&G Chairman and CEO **A.G. Lafley** – one of the best CEOs in the world during his tenure – once told me: "The analysis never tells you the answer. The best it can do is inform your judgment in a helpful way. If you expect the analysis to give you the answer, you will be disappointed with the answer you get."

Second, never dismiss strong feelings that you have but cannot explain. In areas in which your qualitative capacity is nascent, your feelings will run ahead of your ability to explain them to another person – largely because you cannot yet explain them to yourself. But that doesn't mean that they are wrong. The task, rather than dismissing these feelings, is to integrate them into your quantitative analysis.

Third, cultivate surprise and learn to embrace it. If the course of action you have chosen – i.e. your model – produces an outcome that you didn't expect, don't get upset and throw out the experiment. Instead, learn from it and adjust your model.

All of this poses a stiff challenge for both organizations and management education, which have been set up to develop and worship quantitative models and findings. At the moment, professors, students and pedagogy are not oriented toward the development of qualitative capacity. But educating only half of a student's brain is a crying shame. We must embrace the education of the full mind – the analytical and the artistic, the quantitative and the qualitative. Students need to learn how to think critically and creatively every bit as much as they need to learn how to crunch numbers. In the end, that is the only way we will create leaders that are capable of facing our toughest problems and answering our most difficult questions. **R**

Roger Martin is Dean, Premier's Research Chair and professor of Strategic Management at the Rotman School of Management. His latest book is *The Design of Business: Why Design Thinking is the Next Competitive Advantage* (Harvard Business Publishing, 2009).

Salsburg, D. 2002. The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. Chapter 2: The Skew Distributions

This page intentionally left blank

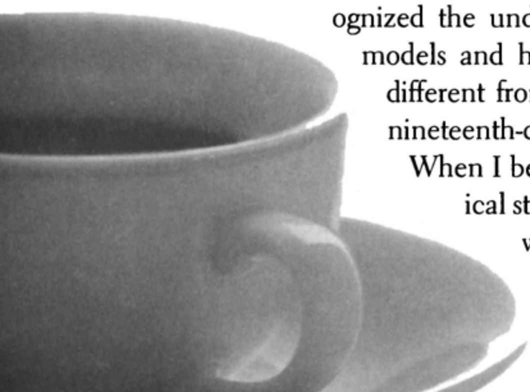
CHAPTER

2

THE SKEW DISTRIBUTIONS

As with many revolutions in human thought, it is difficult to find the exact moment when the idea of a statistical model became part of science. One can find possible specific examples of it in the work of the German and French mathematicians of the early nineteenth century, and there is even a hint of it in the papers of Johannes Kepler, the great seventeenth-century astronomer. As indicated in the preface to this book, Laplace invented what he called the error function to account for statistical problems in astronomy. I would prefer to date the statistical revolution to the work of Karl Pearson in the 1890s. Charles Darwin recognized biological variation as a fundamental aspect of life and made it the basis of his theory of the survival of the fittest. But it was his fellow Englishman Karl Pearson who first recognized the underlying nature of statistical models and how they offered something different from the deterministic view of nineteenth-century science.

When I began the study of mathematical statistics in the 1960s, Pearson was seldom mentioned in my



classes. As I met and talked with the major figures in the field, I heard no references to Pearson or his work. He was either ignored or treated as a minor figure whose activities had long since been outmoded. Churchill Eisenhart, from the U.S. National Bureau of Standards, for instance, was studying at University College, London, during the final years of Karl Pearson's life. He remembered Pearson as a dispirited old man. The pace of statistical research had swept him by, dashing him and most of his work into the dustbin of the past. The bright young students at University College were flocking to study at the feet of the newer great men, one of them Karl Pearson's own son, but no one was coming to see old Karl in his lonely office far from the bustle of new, exciting research.

It wasn't always like this. In the 1870s, young Carl [sic] Pearson had left England to pursue his graduate studies in political science in Germany. There he became enamored of the work of Karl Marx. In tribute to Marx, he changed the spelling of his own first name to Karl. He returned to London with a doctorate in political science, having written two respectable books in the field. In the very heart of stuffy Victorian England, he had the audacity to organize a Young Men's and Women's Discussion Club. At the club, young men and women gathered together (unchaperoned), in an equality of the sexes modeled after the salons of upper-class German and French society. There they discussed the great political and philosophical problems of the world. The fact that Pearson met his wife in this environment suggests that there may have been more than one motive for founding the club. This little social venture provides some insight into Karl Pearson's original mind and his utter disregard for established tradition.

Although his doctorate was in political science, Pearson's main interests were in the philosophy of science and the nature of mathematical modeling. In the 1880s, he published *The Grammar of Science*, which went through a number of editions. For much of the period prior to World War I, this was considered one of the great books on the nature of science and mathematics. It is filled

Karl Pearson, 1857–1936, with a bust of Raphael Weldon in the background

with brilliant, original insights, which make it an important work in the philosophy of science. It was also written in a smooth, simple style that makes it accessible to anyone. You don't have to know mathematics to read and understand the *Grammar of Science*. Although, at this writing, the book is over a hundred years old, the insights and the ideas found in it are pertinent to much mathematical research of the twenty-first century and provide an understanding of the nature of science that holds true even today.

The Galton Biometrical Laboratory

At this point in his life, Pearson fell under the influence of the English scientist Sir Francis Galton. Most people who have heard of Galton know him as the "discoverer" of fingerprints. The realization that fingerprints are unique to each individual and the methods usually used to classify and identify them are Galton's. The uniqueness of fingerprints lies in the occurrence of irregular marks and cuts in the finger patterns, which are called "Galton Marks." Galton did far more. Independently wealthy, he was a dilettante scientist who sought to bring mathematical rigor into the science of biology through the study of patterns of numbers. One of his first investigations involved the inheritance of genius. He collected information about pairs of fathers and sons who were reputed to be highly intelligent. He found the problem very difficult, however, because there was no good measure of intelligence at the time. He decided to look at the inheritance of traits that were more easily measured, like height.

Galton set up a biometrical laboratory (*bio* for biology, *metric* for measurement) in London and advertised for families to come and be measured. At the biometrical laboratory, he collected heights, weights, measurements of specific bones, and other characteristics of family members. He and his assistants tabulated these data and examined and reexamined them. He was looking for some way to predict measures from parents to children. It was obvious, for instance, that tall parents tended to have tall children, but was there some mathematical formula that would predict how tall the children would be, using only the heights of the parents?

CORRELATION AND REGRESSION

In this way, Galton discovered a phenomenon he called "regression to the mean." It turned out that sons of very tall fathers tended to be shorter than their fathers and sons of very short fathers tended to be taller than their fathers. It was as if some mysterious force were causing human heights to move away from the extremes and toward the mean or average of all humans. The phenomenon of regression to the mean holds for more than human heights. Almost all scientific observations are bedeviled by regression to the mean. We shall see in chapters 5 and 7 how R. A. Fisher was able to turn Galton's regression to the mean into statistical models that now dominate economics, medical research, and much of engineering.

Galton thought about his remarkable finding and then realized that it had to be true, that it could have been predicted before making all his observations. Suppose, he said, that regression to the mean did not occur. Then, on the average, the sons of tall fathers would be as tall as their fathers. In this case, some of the sons would have to be taller than their fathers (in order to average out the ones who are shorter). The sons of this generation of taller men would then average their heights, so some sons would be even taller. It would go on, generation after generation. Similarly, there would be some sons shorter than their fathers, and some grandsons even shorter, and so on. After not too many generations, the human race would consist of ever taller people at one end and ever shorter ones at the other.

This does not happen. The heights of humans tend to remain stable, on the average. This can only happen if the sons of very tall fathers average shorter heights and the sons of very short fathers average greater heights. Regression to the mean is a phenomenon that maintains stability and keeps a given species pretty much the "same" from generation to generation.

Galton discovered a mathematical measure of this relationship. He called it the "coefficient of correlation." Galton gave a specific formula for computing this number from the type of data he collected at the biometrical laboratory. It is a highly specific formula for measuring one aspect of regression to the mean, but tells us nothing whatsoever about the cause of that phenomenon. Galton first used the word *correlation* in this sense. It has since moved into popular language. Correlation is often used to mean something much more vague than Galton's specific coefficient of correlation. It has a scientific ring to its sound, and nonscientists often bandy the word around as if it described how two things are related. But unless you are referring to Galton's mathematical measure, you are not being very precise or scientific if you use the word correlation, which Galton used for this specific purpose.

DISTRIBUTIONS AND PARAMETERS

With the formula for correlation, Galton was getting very close to this new revolutionary idea that was to modify almost all science in the twentieth century. But it was his disciple, Karl Pearson, who first formulated the idea in its most complete form.

To understand this revolutionary idea, you have to cast aside all preconceived notions about science. Science, we are often taught, is measurement. We make careful measurements and use them to find mathematical formulas that describe nature. In high school physics, we are taught that the distance a falling body will travel versus time is given by a formula involving a symbol g, where g is the constant of acceleration. We are taught that experiments can be used to determine the value of g. Yet, when the high school student runs a sequence of experiments to determine the value of g, rolling small weights along an inclined plane and measuring how long it takes them to get to different places on the ramp, what happens? It seldom comes out right. The more times the student runs the experiment, the more confusion occurs, as different values of g emerge from different experiments. The teacher looks down from his superior knowledge and assures the students that they are not getting the right answer because they are sloppy or not being careful or have copied incorrect numbers.

What he does not tell them is that all experiments are sloppy and that very seldom does even the most careful scientist get the number right. Little unforeseen and unobservable glitches occur in every experiment. The air in the room might be too warm and the sliding weight might stick for a microsecond before it begins to slide. A slight breeze from a passing butterfly might have an effect. What one really gets out of an experiment is a scatter of numbers, not one of which is right but all of which can be used to get a close estimate of the correct value.

Armed with Pearson's revolutionary idea, we do not look upon experimental results as carefully measured numbers in their own right. Instead, they are examples of a scatter of numbers, a *distribution* of numbers, to use the more accepted term. This distribution of numbers can be written as a mathematical formula that tells us the probability that an observed number will be a given value. What value that number actually takes in a specific experiment is unpredictable. We can only talk about probabilities of values and not about certainties of values. The results of individual experiments are random, in the sense that they are unpredictable. The statistical models of distributions, however, enable us to describe the mathematical nature of that randomness.

It took some time for science to realize the inherent randomness of observations. In the eighteenth and nineteenth centuries, astronomers and physicists created mathematical formulas that described their observations to a degree of accuracy that was acceptable. Deviations between observed and predicted values were expected because of the basic imprecision of the measuring instruments, and were ignored. Planets and other astronomical bodies were assumed to follow precise paths determined by the fundamental equations of motion. Uncertainty was due to poor instrumentation. It was not inherent in nature.

With the development of ever more precise measuring instruments in physics, and with attempts to extend this science of measurement to biology and sociology, the inherent randomness of nature became more and more clear. How could this be handled? One way was to keep the precise mathematical formulas and treat the deviations between the observed values and the predicted values as a small, unimportant error. In fact, as early as 1820, mathematical papers by Laplace describe the first probability distribution, the error distribution, that is a mathematical formulation of the probabilities associated with these small, unimportant errors.

This error distribution has entered popular parlance as the "bell-shaped curve," or the normal distribution.[1]

It took Pearson to go one step beyond the normal, or error, distribution. Looking at the data accumulated in biology, Pearson conceived of the measurements themselves, rather than errors in the measurement, as having a probability distribution. Whatever we measure is really part of a random scatter, whose probabilities are described by a mathematical function, the distribution function. Pearson discovered a family of distribution functions that he called the "skew distributions" and that, he claimed, would describe any type of scatter a scientist might see in data. Each of the distributions in this family is identified by four numbers.

The numbers that identify the distribution function are not the same type of "number" as the measurements. These numbers can never be observed but can be inferred from the way in which the measurements scatter. These numbers were later to be called parameters—from the Greek for "almost measurements." The four parameters that completely describe a member of the Pearson System are called

1. the mean—the central value about which the measurements scatter,
2. the standard deviation—how far most of the measurements scatter about the mean,
3. symmetry—the degree to which the measurements pile up on only one side of the mean,
4. kurtosis—how far rare measurements scatter from the mean.

[1]It is sometimes called the Gaussian distribution, in honor of the man once believed to have first formulated it, except that it was not Carl Friedrich Gauss but an earlier mathematician named Abraham de Moivre who first wrote down the formula for the distribution. There is good reason to believe Daniel Bernoulli came across the formula before this. All of this is an example of what Stephen Stigler, a contemporary historian of science, calls the law of misonomy, that nothing in mathematics is ever named after the person who discovered it.

There is a subtle shift in thinking with Pearson's system of skew distributions. Before Pearson, the "things" that science dealt with were real and palpable. Kepler attempted to discover the mathematical laws that described how the planets moved in space. William Harvey's experiments tried to determine how blood moved through the veins and arteries of a specific animal. Chemistry dealt with elements and compounds made up of elements. However, the "planets" that Kepler tried to tame were really a set of numbers identifying the positions in the sky where shimmering lights were seen by observers on earth. The exact course of blood through the veins of a single horse was different from what might have been seen with a different horse, or with a specific human being. No one was able to produce a pure sample of iron, although it was known to be an element.

Pearson proposed that these observable phenomena were only random reflections. What was real was the probability distribution. The real " things" of science were not things that we could observe and hold but mathematical functions that described the randomness of what we could observe. The four parameters of a distribution are what we really want to determine in a scientific investigation. In some sense, we can never really determine those four parameters. We can only estimate them from the data.

Pearson failed to recognize this last distinction. He believed that if we collected enough data the estimates of the parameters would provide us with true values of the parameters. It took his younger rival, Ronald Fisher, to show that many of Pearson's methods of estimation were less than optimal. In the late 1930s, as Karl Pearson was approaching the end of his long life, a brilliant young Polish mathematician, Jerzy Neyman, showed that Pearson's system of skew distributions did not cover the universe of possible distributions and that many important problems could not be solved using the Pearson system.

But let us leave the old, abandoned Karl Pearson of 1934 and return to the vigorous man in his late thirties, who was filled with enthusiasm over his discovery of skew distributions. In 1897, he took over Galton's biometrical laboratory in London and marshaled legions of young women (called "calculators") to compute the parameters of distributions associated with the data Galton had been accumulating on human measurements. At the turn of the new century, Galton, Pearson, and Raphael Weldon combined their efforts to found a new scientific journal that would apply Pearson's ideas to biological data. Galton used his wealth to create a trust fund that would support this new journal. In the first issue, the editors set forth an ambitious plan.

THE PLAN OF *BIOMETRIKA*

Galton, Pearson, and Weldon were part of an exciting cadre of British scientists who were exploiting the insights of one of their most prominent members, Charles Darwin. Darwin's theories of evolution postulated that life forms change in response to environmental stress. He proposed that changing environments gave a slight advantage to those random changes that fit better into the new environment. Gradually, as the environment changed and life forms continued to have random mutations, a new species would emerge that was better fit to live and procreate in the new environment. This idea was given the shorthand designation "survival of the fittest." It had an unfortunate effect on society when arrogant political scientists adapted it to social life, declaring that those who emerged triumphant from the economic battle over riches were more fit than those who plunged into poverty. Survival of the fittest became a justification for rampant capitalism in which the rich were given the moral authority to ignore the poor.

In the biological sciences, Darwin's ideas seemed to have great validity. Darwin could point to the resemblances among related species as suggesting a previous species out of which these modern ones had emerged. Darwin showed how small birds of slightly different species and living on isolated islands had many anatomical commonalities. He pointed to the similarities among embryos of different species, including the human embryo, which starts with a tail.

The one thing Darwin was unable to show was an example of a new species actually emerging within the time frame of human history. Darwin postulated that new species emerge because of the survival of the fittest, but there was no proof of this. All he had to display were modern species that appeared to "fit" well within their environment. Darwin's proposals seemed to account for what was known, and they had an attractive logical structure to them. But, to translate an old Yiddish expression, "For instance is no proof."

Pearson, Galton, and Weldon set out in their new journal to rectify this. In Pearson's view of reality as probability distributions, Darwin's finches (an important example he used in his book) were not the objects of scientific investigation. The random distribution of all finches of a species was the object. If one could measure the beak lengths of all the finches in a given species, the distribution function of those beak lengths would have its own four parameters, and those four parameters would *be* the beak length of the species.

Suppose, Pearson said, that there was an environmental force changing a given species by providing superior survivorship to certain specific random mutations. We might not be able to live long enough to see a new species emerge, but we might be able to see a change in the four parameters of the distribution. In their first issue, the three editors declared that their new journal would collect data from all over the world and determine the parameters of their distributions, with the eventual hope of showing examples of shifts in parameters associated with environmental change.

They named their new journal *Biometrika*. It was funded lavishly by the Biometrika Trust that Galton set up, and was so well funded that it was the first journal to publish full color photographs and foldout glassine sheets with intricate drawings. It was printed on high-quality rag paper, and the most complicated mathematical formulas were displayed, even if they meant extremely complicated and expensive typesetting.

For the next twenty-five years, *Biometrika* printed data from correspondents who plunged into the jungles of Africa to measure tibia and fibula of the natives; sent in beak lengths of exotic tropical birds caught in the rain forests of Central America; or raided ancient cemeteries to uncover human skulls, into which they poured buckshot to measure cranial capacity. In 1910, the journal published several sheets of full color photographs of flaccid penises of pygmy men, laid on a flat surface against measuring sticks.

In 1921, a young female correspondent, Julia Bell, described the troubles she underwent when she tried to get anthropomorphic measurements of recruits for the Albanian army. She left Vienna for a remote outpost in Albania, assured that she would find German-speaking officers to help her. When she arrived, there was only a sergeant, who spoke three words of German. Undaunted, she took out her bronze measuring rods and got the young men to understand what she wanted by tickling them until they lifted their arms or legs as she desired.

For each of these data sets, Pearson and his calculators computed the four parameters of the distributions. The articles would display a graphical version of the best-fitting distribution and some comments about how this distribution differed from the distribution of other related data. In retrospect, it is difficult to see how all this activity helped prove Darwin's theories. Reading through these issues of *Biometrika*, I get the impression that it soon became an effort that was done for its own sake and had no real purpose other than estimating parameters for a given set of data.

Scattered throughout the journal are other articles. Some of them involve theoretical mathematics dealing with problems that arise from the development of probability distributions. In 1908, for instance, an unknown author, publishing under the pseudonym of "Student," produced a result that plays a role in almost all modern scientific work, "Student"'s t-test. We will meet this anonymous author in later chapters and discuss his unfortunate role in mediating between Karl Pearson and Ronald Fisher.

Galton died in 1911, and Weldon died in a skiing accident in the Alps earlier. This left Pearson as the sole editor of *Biometrika* and the sole dispenser of the trust's money. In the next twenty years, it was Pearson's personal journal, which published what Pearson thought was important and did not publish what Pearson thought was unimportant. It was filled with editorials written by Pearson, in which he let his fertile imagination range over all sorts of issues. Renovation of an ancient Irish church uncovered bones in the walls, and Pearson used involved mathematical reasoning and measurements made on those bones to determine whether they were, in fact, the bones of a particular medieval saint. A skull was found that was purported to be the skull of Oliver Cromwell. Pearson investigated this in a fascinating article that described the known fate of Cromwell's body, and then compared measurements made on pictures painted of Cromwell to measurements made on the skull.[2] In other articles, Pearson examined the lengths of reigns of kings and the decline of the patrician class in ancient Rome, and made other forays into sociology, political science, and botany, all of them with a complicated mathematical gloss.

Just before his death, Karl Pearson published a short article entitled "On Jewish–Gentile Relationships," in which he analyzed anthropomorphic data on Jews and Gentiles from various parts of the world. He concluded that the racial theories of the National Socialists, the official name of the Nazis, were sheer nonsense, that there was no such thing as a Jewish race or, for that matter, an Aryan race. This final paper was well within the clear, logical, carefully reasoned tradition of his previous work.

[2]After the restoration of the monarchy, following Cromwell's dictatorship, a truce between the two factions in the civil war in England meant that the new rulers could not prosecute any of the living followers of Cromwell. However, there was nothing in the truce about the dead. So the bodies of Cromwell and two of the judges who had ordered the execution of Charles I were dug up and tried for the crime of regicide. They were convicted, and their heads were chopped off and placed on pikes above Westminster Abbey. The three heads were left there for years and eventually disappeared. A head, supposedly that of Cromwell, showed up in a "museum" in London. It was that head which Pearson examined. He concluded that it was, indeed, the head of Oliver Cromwell.

Pearson used mathematics to investigate many areas of human thought that few would consider the normal business of science. To read through his editorials in *Biometrika* is to meet a man with a universal range of interests and a fascinating capacity to cut to the heart of any problem and find a mathematical model with which to attack it. To read through his editorials is also to meet a strong-willed, highly opinionated man, who viewed subordinates and students as extensions of his own will. I think I would have enjoyed spending a day with Karl Pearson—provided I did not have to disagree with him.

Did they prove Darwin's theory of evolution through survival of the fittest? Perhaps they did. By comparing the distributions of cranial capacity from skulls in ancient cemeteries to those of modern men and women, they managed to show that the human species has been remarkably stable across many thousands of years. By showing that anthropomorphic measurements on aborigines had the same distribution as measurements taken on Europeans, they disproved the claims of some Australians that the aborigines were not human. Out of this work, Pearson developed a basic statistical tool known as the "goodness of fit test," which is an indispensable tool for modern science. It enables the scientist to determine whether a given set of observations is appropriate to a particular mathematical distribution function. In chapter 10, we shall see how Pearson's own son used this goodness of fit test to undermine much of what his father had accomplished.

As the twentieth century advanced, more and more of the articles in *Biometrika* dealt with theoretical problems in mathematical statistics and fewer dealt with distributions of specific data. When Karl Pearson's son, Egon Pearson, took over as editor, the shift to theoretical mathematics was complete, and today *Biometrika* is a preeminent journal in that field.

But did they prove survival of the fittest? The closest they came to it occurred early in the twentieth century. Raphael Weldon conceived of a grand experiment. The development of china factories in southern England in the eighteenth century had caused some of the rivers to become silted with clay, so the harbors of Plymouth and Dartmouth had changed, with the interior regions more silted than those closer to the sea. Weldon took several hundred crabs from these harbors and put them into individual glass jars. In half the jars he used the silted water from the inner harbors. In the other half of the jars he used clearer water from the outer harbors. He then measured the carapaces of the crabs that survived after a period of time and determined the parameters of the two distributions of crabs: those that survived in clear water and those that survived in silted water.

Just as Darwin had predicted, the crabs that survived in the silted jars showed a change in distribution parameters! Did this prove the theories of evolution? Unfortunately, Weldon died before he could write up the results of his experiment. Pearson described the experiment and its results in a preliminary analysis of the data. A final analysis was never run. The British government, which had supplied the funds for the experiment, demanded a final report. The final report never came. Weldon was dead, and the experiment was ended.

Eventually, Darwin's theories were shown to be true for short-lived species like bacteria and fruit flies. Using these species, the scientist could experiment with thousands of generations in a short interval of time. Modern investigations of DNA, the building blocks of heredity, have provided even stronger evidence of the relationships among species. If we assume that the rate of mutation has been constant over the past ten million or more years, studies of DNA can be used to estimate the time frame of species emergence for primates and other mammals. At a minimum, it runs into the hundreds of thousands of years. Most scientists now accept Darwin's mechanism of evolution as correct. No other theoretical mechanism has been proposed that matches all known data so well. Science is satisfied, and the idea that one needs to determine the shift in distribution parameters to show evolution on a short time scale has been dropped.

What remains of the Pearsonian revolution is the idea that the "things" of science are not the observables but the mathematical distribution functions that describe the probabilities associated with observations. Today, medical investigations use subtle mathematical models of distributions to determine the possible effects of treatments on long-term survival. Sociologists and economists use mathematical distributions to describe the behavior of human society. In the form of quantum mechanics, physicists use mathematical distributions to describe subatomic particles. No aspect of science has escaped the revolution. Some scientists claim that the use of probability distributions is a temporary stopgap and that, eventually, we will be able to find a way to return to the determinism of nineteenth-century science. Einstein's famous dictum that he did not believe that the Almighty plays dice with the universe is an example of that view. Others believe that nature is fundamentally random and that the only reality lies in distribution functions. Regardless of one's underlying philosophy, the fact remains that Pearson's ideas about distribution functions and parameters came to dominate twentieth-century science and stand triumphant on the threshold of the twenty-first century.

Williams, H. S. 1991. Innovating Reprint. Informing vs. Persuading

This page intentionally left blank

INNOVATING Reprint

Volume 1, Number 2 Winter 1991

Informing vs Persuading

by Harold S. Williams

Programs in government and non-profit groups often begin with the premise of education. We speak of the need to "educate" people to many things—including the dangers of smoking and teen pregnancy, the problems of solid waste and the potentials of a drug-free life. At first blush, who could disagree with such a noble intent? But we need to look deeper to see how the high road in concept becomes the low road in result.

When people speak of educating others, the probability is very high that they do not have information gain as the objective. Nor is the goal knowledge or even wisdom. Their intent is not simply to inform, but to persuade. They want to get people to do something that they are not doing now, or to prevent them from doing something they might otherwise do.

The point of this paradigm is to show just how far apart are the purposes of educating to inform and educating to persuade. When the distinction is appreciated, a number of strategic implications for program planners and implementers will follow.

The paradigm shapes up as follows:

Informing	**Persuading**
Focus on data	Focus on data use
Ends with knowledge	Ends with behavior
Objective (facts)	Subjective (meanings)
What people are presumed to need	What people want and will use
Begins with answers	Begins with questions
Extols virtues	Sells benefits

The differences are key. To begin, the informing mind-set sees information as its commodity. Public-sector and non-profit prevention programs churn out facts and figures on every imaginable issue. Many millions of dollars are spent annually on videos, pamphlets, speeches, workshops, and other means repeatedly used to disseminate data.

The assumption is that if people become knowledgeable on a given issue, they will do the right thing. This core tenet is clear but misguided. From diverse social science research studies comes a key finding: information is rarely sufficient to change behavior. Evaluations of prevention programs in many critical social problem areas conclude that unless other steps are taken, the likelihood is low that a

INNOVATING, the quarterly journal, is for people in organizations and communitites who lead change by example. We are published by The Rensselaerville Institute (TRI), an educational center defined at www.tricampus.org. INNOVATING is our registered trademark. Annual subscriptions are $30. Send requests for subscription or author guidelines to Kathy Reinhart at The Rensselaerville Institute, Rensselaerville, NY 12147. Telephone: (518) 797-3783. Fax: (518) 797-5270. Email: kreinhart@tricampus.org

person hearing the message—whether on drugs, illiteracy, pollution or another topic—will respond with a new behavior. Further, this remains true even if the person gets and retains the information!

The persuading mind-set starts and ends not with data but with the person who must use it if "education" is to succeed. It recognizes that the role of information is no more critical than many other factors—such as recipient values, sense of options, peer pressure, and wants and desires—in determining responsive behavior.

Informing focuses on the objective world of facts and figures. The problem is that people act not on data but on what it means to them. Management consultant Zig Ziglar makes this point by telling a story from European history. Aristotle believed that if two different weights of the same material were dropped from a uniform height, they would fall at different speeds. This belief was taught at the University of Pisa. Years later, Galileo arrived not only to challenge this theory but prove equal drop rates by a demonstration from the Leaning Tower of Pisa. He convinced the students and teachers beyond any doubt. But Aristotle's theory continued to be taught at that university. Galileo had informed them but had not persuaded them.

Informing, like much of current education, is based on a premise that is tidy but self-limiting: the educator knows what the student needs. It is self-evident, according to this view, that people need jobs, self-esteem, literacy, connection to parks and cultural assets, or whatever else we have to offer. In reality, what people want and what they are prepared to use is often very different from what others think they should have. The most fundamental premise in marketing is that what the customer wants to buy is more important than what the vendor wants to sell. This is as true for social and human services as for computer software and cars.

It is not necessary to end up with what a person first desires, but it is essential to start there. The entry point is questions, not answers. Until we learn about the concerns and issues as seen by potential recipients of our "educating," much of what we tell them will prove irrelevant. When the informer focuses on value and virtues of a given course of action, his or her speech can be endless. When he or she focuses on benefits meaningful to the respondent, the words are fewer but far more effective.

While this essay speaks to "educating" beyond schools, a number of people in formal education are coming to this understanding as well. Classrooms filled with lectures and rote memorization are gradually giving way to sessions of active inquiry and implementation by students. For example, while schools believe that many young people need physics, those same young people are much more likely to say that what they want to know about is cars. A few bright physics teachers solve this by recognizing that the study of a car and its parts can include virtually every key principle of physics.

One teacher discovered recipient meaning when faced with the task of getting his students to memorize a passage of the U.S. Constitution. His efforts at both mechanical and humanistic approaches failed. Then, by looking closely at the world of the students, he came to the realization that most of them were perfectly capable of memorizing—if only the context were theirs and not his. He asked the kids to form into groups and to do the passages as a rap song. Almost everyone passed!

Some readers may resist the informing to persuading shift on ethical grounds. Persuading is manipulative compared to the elevated ground of education. My response is that the intent of persuading cloaked as "educating" is the greater worry.

The New Framework for Persuading

The paradigm of persuading needs a framework that may at first glance seem an uneasy fit in the public sector. The framework is called marketing and it is useful for three reasons. The first is its insistence that the clients, patients, students, offenders, and at-risk persons whom public and non-profit groups served be considered as customers. "Customer" is a useful term in connoting the choices that recipients of social and other public programs actually do have. They may choose or not choose to get the information. They may choose or not choose to retain it. They may choose or not choose to use it. Indeed, program offerers are actually dependent on voluntary acts of potential customers to be achieve their results.

The second advantage of marketing comes from the inclusiveness of definition. Marketing means the design, pricing, promotion, distribution, and sale of ideas, goods, and services to create exchanges that satisfy both offerer and recipient. Marketing starts with the problem of shaping products that will meet customer needs and wants, not as the problem of selling someone something entirely predetermined. Many forms of public-sector "educating" presume that the product is perfect and, indeed, if there are defects, they must lie with the user!

As a third factor, marketing connects product with customer in an exchange process that is rarely successful when viewed as exploitation. Unless the customer also wins, the exchange generally won't happen. And if it does happen, it won't last. Marketing is at heart <u>cooperative</u>, not <u>competitive</u>.

Marketing as the framework for persuading has three successive elements.

1. <u>Informing</u>. People are given sufficient information that they are able to draw a conclusion and become interested in doing so. In the marketing framework, information is more through compelling examples and illustrations than in statistical profiles.

2. <u>Convincing</u>. In this step, customers put together information to form conclusions concerning not only belief but disposition to act on belief. The recipient now has connected information bits into a pattern—a conviction and a disposition.

3. <u>Action Prompting</u>. The disposition to do something is now translated into behavior. Persons convinced they should get a job, fight drugs, lose weight, stop abusing, etc., now act on their conviction. They do something.

An example from my world of fund raising at The Rensselaerville Institute may help put these steps in perspective. To raise over $500,000 annually from a diverse set of donors, we start by <u>informing</u> them. Through newsletters, annual reports and the like, we tell people what The Institute does and stress the results of our work. While information is not enough, it is an essential starting point. We must ensure that a lively and compelling tone will lead people to read it. If they don't, the next steps are probably irrelevant.

Our next step is <u>convincing</u> those people who get informed to support us. For contributions to come, those whom we inform must decide that The Institute's programs: a) deserve their support, and b) are more worthy of support than others to which they might give instead. When this happens, people are disposed to contribute. This step takes follow-up by us. It will not happen solely from our mailing.

But if I relax at my desk waiting for big bucks to come in, I'm sure to be disappointed. The gap between belief and action can be great. Most paths are paved with good intentions. I must now convert conviction to behavior by <u>prompting action</u>. In marketing terms, this step is called "the sale." At least for potential donors of larger amounts, I need to find a way to say to them: "I'm delighted you find us worthy of your support." This last step makes the difference between a gift to us or a gift to someone else...or to no one at all.

A number of strategies are defined by this progression. Here are five which are broadly relevant to public service programs.

1. <u>Knowing when to end a step is as important as knowing when to start it</u>. The timing issue in classroom education is knowing when to begin. Thresholds are deemed important in suggesting "reading readiness" and the like. In persuading, the key is less to know when to conclude.

There is almost no end to the facts and figures we might communicate and help others to retain in the step of informing. From a persuading perspective, however, there is a level of information that is sufficient to enable persons to become convinced. Continuing to inform people once that point is reached is actually counterproductive. Clarity becomes confusion; data becomes overload.

If informing has its useful limits, convincing has its end point—when a conviction is reached. Once a person has reasoned, analyzed, felt or in other ways come to the desired conclusion, he or she is disposed to act. Additional time spent on developing the conviction may well lead to new questions or a reconsideration of decision. Advocates are among those who can linger too long. They are more preoccupied with the principle of understanding rights than the behavior of meeting needs. Others fall in this category because they believe that conviction will invariably lead to behavior if only it is sufficiently strengthened and enforced.

The final step, prompting action, recognizes that there is a <u>peak point</u> at which a person is most likely to act on his or her conviction. Once past, readiness actually declines! Salespersons (including evangelists and fund raisers just as readily as those who sell cars or computers) have made a science of knowing just when the customer is most likely to buy the product. They carefully wait for that moment to close the sale.

Persons who would "educate" people to do something tend to be anchored in the timeless and durable nature of the problems which they, as helpers, face. They fail to understand that the world of those who have (and occasionally cause) problems is not so constant. Interveners frequently fail to understand that, if the endpoint is action rather than accrual, less can certainly be more.

2. <u>Persuading builds on choice</u>. When advertising compares the good or service promoted with other brands, it recognizes that people do have choices and that it is useful to acknowledge and speak to them. We make a mistake when we think of those served by government and non-profit groups as being a captive audience forced to listen and respond, whether by necessity or mandate. While money may not be needed, people have a choice concerning time. They may attend your program, watch TV, play poker, enjoy their kids or anything else.

   Even if a program could have required attendance, this would not be helpful. Indeed, there is little basis for forming a conclusion of any kind if no choice is at hand. One simply follows the one available path. Even in formal education, lessons that assume only one right answer are generally less effective than those that enable learners to actively engage content to analyze and conclude in more than one way.

   In convincing, the structure moves from options to choice. Recipients apply reasoning and judgment to draw a conclusion concerning the content and their disposition to it. Choice-making is decision-making. In this context, the persuader must often find a way to call the question. Whether it's choosing one car over another or yes or no to drug use, people can continually seek more information and defer deciding unless prompted to do so.

   Since choice is to lead to action, it is important to show the connection between decision and the behaviors needed to implement or express it. If conviction precedes this understanding, it is at best an attitude in search of a behavior. People can hold many of those. Expression of choice is especially critical in that in most instances the choice made is not simply a behavior but a <u>change</u> in behavior. People are not only asked to do something new, but to stop doing something old.

3. <u>The recipient's problem must be solved</u>. Persuading is usefully defined as problem-solving. What problem does my customer have for which my program can become a solution? Once

the problem is defined, some of the benefits of the program or message become critical; others become irrelevant. In weight reduction, for example, the problem individuals is rarely the pounds themselves. Rather, it is the perception of being unattractive to others, being uncomfortable in groups, or something else. A clinical discussion of the facts and figures of obesity in such terms as health risk and mortality are not likely to be relevant to the wants of many overweight people.

4. Disposition is the linchpin connecting interest and action. A person is disposed to do something when he or she is clear not only on what the information means but on how it helps to make a choice and implement it. In general, conviction involves an emotional as well as cognitive commitment. It ties to feelings and values as well as to facts and proof. For this reason, effective marketers use illustrations that will personally touch the lives of those whom they would persuade.

   A related distinction is the continuum of objective informing to normative convincing to subjective action. Information gains credibility from its objectivity, often defined as constancy over time and place. At the convincing step, data becomes meaningful as it is anchored to the beliefs and values the customer holds to be important. A choice that must fly in the face of previously held convictions creates cognitive dissonance, and will not likely be made or implemented. Thus, a legislator who believes that most people on public assistance are lazy will not readily form the conviction that an expanded public welfare program should be funded. At the same time, a belief that ties to an existing disposition will predictably be embraced. That same legislator, for example, may believe that government is too large and intrusive, and support a self-help program that enables people on public assistance to solve their own problems.

   Action—the exercise of choice once made— becomes personal. While it is guided by objective data and shaped by normative meanings, it happens subjectively. Indeed, impulses for action often best explain purchasing behavior. At some point of evident readiness, the right car, spouse, program or other choice comes along and is promptly embraced. If explanations are needed, most of us are highly capable of constructing them after the fact. The line between rationale and rationalization is a thin one.

5. Inform a population, convince people, prompt action from individuals. Informing can and is done on a mass-media basis. Once designed, information can be widely distributed. You need not see, hear, or touch the persons with whom you communicate. For convincing, some type of two-way communication is often needed...a telephone call, a visit, a highly tailored letter. Convincing often takes meetings or other activities which bring together those selected (and, generally, self-selected) for interest. At this point, names become faces and faces become differences. At the level of prompting action—the sale, interpersonal moves to personal. In general, this is a one-on-one relationship.

   The increased cost per person of convincing and prompting action suggests careful resource allocation. An "educational" campaign that spends all its money informing 20,000 persons may end up with fewer results than one that starts by informing only 5,000 persons, but with enough time and money left to follow-up with those interested to get at least 500 persons to do something differently as a result.

These and other distinctions among the three stages can be summarized as follows:

The Persuading Progression

Inform———>	Convince————>	Prompt action
options	choice	use of choice
what people need	what persons want	what individual uses
interest	disposition	behavior
objective	normative	subjective
retain data	integrate data	apply convictions
mass media	interpersonal media	personal media

As with all paradigm shifts, the change from informing to persuading is not easy. We must begin by questioning some cherished beliefs—including the reality that information is not enough and that providers always know best.

If this paradigm interests you, learn it by application. For yourself and for others, apply the paradigm to this shift itself. How would you move from informing others about the persuading approach to convincing them to try it and then prompting that action? When to begin? The next time you hear someone say, "We need to educate people to....."

The author is President of The Rensselaerville Institute and teaches workshops on marketing government programs. More detailed publications on marketing in public and non-profit sectors are available from
The Institute

Neustadt, R. E., & May, E. R. 1988. Thinking in Time: The Uses of History for Decision-Makers. Chapter 1: Success Story

This page intentionally left blank

THINKING IN TIME

The Uses of History for Decision-Makers

Richard E. Neustadt
Ernest R. May

THE FREE PRESS

1
Success Story

"They're too busy. Can't read what they get now. They'll glance at papers in the limousine, thumb them while someone is talking, or just wing it. If you do get their attention, you can't keep it. They will have to catch a plane or go to a press conference."

So an experienced diplomat responded to our argument that government officials should make better use of history. He was right. It may be easier to get a million dollars of public money than a minute from a president or a cabinet officer. The Bay of Pigs debacle (discussed later) occurred in part because President John F. Kennedy and his key advisers could never give it sustained attention for more than forty-five minutes at a time.[1] The same strains work on governors, mayors, and many officials much less exalted.

But we are not asking a lot. In government and outside, decision-makers use history now. They draw every day on the past experience of other people. They assign aides bits and pieces of historical research: going to the files or checking memories and comparing recollections. They look at a great many words on paper. A former high official wrote us, "Although the public impression is that Presidents and Secretaries of State have no time to read or think, the truth is that most of them spend an enormous amount of time reading material generated both in the government and outside."[2] We argue chiefly that uses now made of history can be more reflective and systematic, hence more helpful.

This book is about how to do it. With stories of success and failure we suggest practices which, if made routine, could at least protect against common mistakes. We have tried to make the stories entertaining. We think them also instructive, even for readers too young to vote. Our particular target audience, however, consists of decision-makers and the women and men who work for them (or hope to do so) as direct or personal staff. Almost every executive has a split personality. He or she wants to act and feels impatient with those who block action. Presidents feel so about Congress, the bureaucracy, foreign allies, and the press; cabinet officers feel so about Presidents; assistant secretaries feel so about cabinet officers; and so on down to the bottom rungs of management. At the same time, every executive fears being hustled into action by those impatient people down below. The same holds true for legislators; they also make decisions and have decisions thrust upon them. Good staff work consists of helping a boss with both sets of concerns—clearing obstacles on one side while setting them up on the other. This book is intended to be a manual for such staff work. We hope the bosses will read it and tell their aides to put its recommendations into practice. We hope the aides will use it and use it and use it.

We start with a pair of stories about successes: the Cuban missile crisis of 1962 and the social security reform of 1983. Not everyone will agree that those were successes, but the participants so regarded them, and the majority of journalists so label them. Knowing mostly tales with less happy endings, we are not inclined to apply more rigid criteria. In the one case the missiles were withdrawn and nuclear war didn't happen. In the other, the system didn't run out of cash and wage-earners weren't penalized. Both met immediate issues without ending longer-run concerns. Castro remains an unrepentant Communist; cost-of-living adjustments still give Budget Directors fits. The priorities, however, appeared sound to most contemporaries and appear so still, in retrospect. The results are enough for us. Besides, one case is foreign, one domestic, one occurred under Democrats, the other under Republicans. They thus argue that effective use of history is independent of policy area or party.

We turn now to the first of the two stories; the other follows in our second chapter. Then we sum up what both show about using history better.

For President Kennedy, the acute phase of the missile crisis started about 8:45 A.M. on Tuesday, October 16, 1962, when his National Security Assistant, McGeorge Bundy, came to his bedroom to report that a U-2, a high-flying reconnaissance airplane, had brought back photographs showing Russians at work in Cuba on launch sites for medium-range nuclear missiles.[3]

Kennedy reacted with a mixture of alarm and anger. Five years earlier the Russians had startled the world by sending "Sputnik" rockets into space. The Soviet leader, Nikita Khrushchev, said this showed Russia could destroy the United States with intercontinental nuclear missiles. Americans feared a "missile gap" opening in the Russians' favor. Elected in 1960 partly on a promise to close that gap, Kennedy as President gave high priority to a big defense buildup. By the time new intelligence had proved Khrushchev to be bluffing, the United States was on the way to creating a missile gap about two hundred to one in its own favor. Relations had been tense, especially when the Russians suddenly put up a wall between East and West Berlin. More recently tension had eased. Kennedy reached a few agreements with Khrushchev. He hoped for more. Now this!

And in *Cuba!* The revolution of 1959 putting in power Fidel Casto and a Communist regime had shocked Americans at least as much as Sputnik had. Kennedy in 1960 held out hope that he would also get Cuba back into Washington's orbit. His failure to do so gave Republicans an issue for 1962. Castro helped them by asking for—and getting—Soviet military aid. Republican Senator Kenneth Keating of New York charged that the Russians were going to base nuclear missiles in Cuba. Other Republicans echoed him. Senator Homer Capehart of Indiana, running for reelection, called for an invasion of the island. No one, however, produced solid evidence of anything but defensive, nonnuclear air defense missiles. CIA analysts pointed out to Kennedy that the Russians had never placed nuclear missiles even in Eastern Europe: Why would they put them in Cuba? (The answer probably was that a medium-range missile in Eastern Europe could reach the Soviet Union; one in Cuba could not. But that point was easier to see after the fact.)

By late August Kennedy felt worried. He began daily reviews of relevant intelligence. On September 4 he assured the public that the government had no evidence that any Soviet offensive weapons were going into Cuba. For Khrushchev's ears he added, "Were it to be otherwise, the gravest issues would arise." Afterward he received reassurances not only from his intelligence services but directly from Soviet Ambassador Anatoly Dobrynin. To Theodore Sorensen, the President's chief domestic policy adviser and speechwriter, Dobrynin said that everything the Russians were doing in Cuba was "defensive in nature." Dobrynin said the same to Robert F. Kennedy, the Attorney General (and the President's brother).

The news Bundy brought to the President's bedroom that Tuesday morning was thus not utterly unforeseen. It was none the less shocking. Kennedy's immediate response was to name a handful of men with whom he wanted to take counsel. The group would come to be called the Executive Committee of the National

Security Council—ExComm for short. It included Bundy, Sorensen, Secretary of State Dean Rusk, Secretary of Defense Robert McNamara, and Secretary of the Treasury Douglas Dillon. Robert Kennedy took part continuously. Others were eventually asked in.

For a week, the President and ExComm managed to keep the matter secret. Kennedy preserved a noncommittal smile when Soviet Foreign Minister Andrei Gromyko repeated to his face assurances that Russia would do nothing provocative before November's congressional elections. By various ruses, Kennedy and the others kept the White House press corps ignorant of their day-and-night debates.

When Kennedy and his ExComm first went to work, they used history—and did not use it—in very standard ways. In cases of which we know, debate in serious decision situations starts at least nine times out of ten with the question: What do we *do?* Background and context get skipped. The past comes in, if at all, in the form of analogy, with someone speaking of the current situation as like some other. That may be to put a familiar face on something strange. It may be for advocacy—because the analogue's supposed lesson supports the speaker's preference as to what to do. Otherwise, all concern is for the present, with seldom a glance backward or, in any focused way, toward the future. Of such usual practice we shall offer many examples. Here, even in the missile crisis, one sees it at the outset.

Recordings of ExComm's first meetings are now publicly available. Anyone visiting the John F. Kennedy Library in Boston can hear excerpts. The group assembles in the Cabinet Room of the White House a little before noon. Experts from the Central Intelligence Agency explain the U-2 photographs. A few questions are asked about details. Then, in a measured Southern accent, part rural Georgian and part Rhodes Scholar, Secretary Rusk starts substantive discussion by setting forth two choices: give an ultimatum for withdrawal of the missiles or stage a quick surprise strike to destroy them. The crackling, confident voice of McNamara asserts that "any air strike must be directed not solely against the missile sites, but against the missile sites plus the airfields plus the aircraft . . . plus all potential nuclear storage sites." Joint Chiefs Chairman General Maxwell Taylor says clearly, "What we'd like to do is . . . take 'em out without any warning whatever," but he tallies other military options, including a naval blockade. After some back and forth, Kennedy himself, his famous Boston cadence soft-voiced and hesitating, sets the terms for the rest of the day's debate. He specifies three choices: "One would be just taking out these missiles. Number two would be to take out all the airplanes. Number three is to invade." His conclusion as the group recesses is, "We're certainly going to do Number One. We're going to take out these missiles."

During the initial meeting analogies make an appearance. Saying that the Russians may be trying to draw attention to Cuba because they plan a move elsewhere, perhaps against Berlin, Rusk speaks of the "Suez-Hungary combination," alluding to 1956, when Western preoccupation with Suez had made it easier for the Soviets to use tanks to crush a revolution in Hungary. Subsequently, "Suez" becomes shorthand for such a diagnosis.

For subsequent days' debates, we do not yet have verbatim transcripts. We have to reconstruct from contemporary memoranda and later reminiscence. Wednesday saw members of ExComm hold various meetings with Kennedy not present. He had concluded that second-level people such as Rusk's deputy, George Ball, or McNamara's, Roswell Gilpatric, were more likely to speak up with the President not in the room. The scene shifted too. An antiseptic conference room on the seventh floor of the new State Department building became from then on the principal meeting place.

From some early point Robert Kennedy had begun to feel queasy about an air strike. On Tuesday he spoke against going for both missiles and bombers. "I would say that, uh, you're dropping bombs all over Cuba if you do the second. . . . You're going to kill an awful lot of people, and, uh, you're going to take an awful lot of heat on it." Expressing similar doubts, George Ball invoked an analogy. "This, uh, come in there on Pearl Harbor just frightens the hell out of me." Robert Kennedy later recalled passing his brother a note which said, "I know now how Tojo felt when he was planning Pearl Harbor." On Wednesday Robert Kennedy emphasized this analogy. Arguing now against any surprise air strike at all he said that it would be "a Pearl Harbor in reverse, and it would blacken the name of the United States in the pages of history." Notes on the Wednesday meetings prepared for the President by Sorensen referred several times to "Pearl Harbor."[4]

All this parallels what we see as standard practice, far from any ideal. The records of the ExComm suggest myopic concentration on what to do tomorrow. Reference is made now and then, partly for word-saving, partly for advocacy, to analogies from recent history. Looking back now, one can see signs of practice contrary to the usual. If action had been taken either the first day or the second, however, those signs would be scarcely noticeable. The decision would almost surely have been for an air strike. Whether the President would have chosen to hit not only the missile sites but also bombers and air defenses, we cannot guess. Whatever his choice, and whatever happened in the longer term, historians looking back (assuming there were any) would see Kennedy's decision as a product of usual practice.

In fact, Kennedy was not to announce a decision until Monday, October 22—after more than six days of nearly continuous debate. Then, telling the world what the Russians were doing, he was to proclaim a naval "quarantine." That course of action, initially mentioned in passing by General Taylor, had found its first champion in Vice President Lyndon Johnson. By the evening of the first day it had also become McNamara's favorite option—"this alternative doesn't seem to be a very acceptable one," he said, "but wait until you work on the others."[5] At some point—probably early on—the President came to the same opinion. By the weekend there was near-consensus. The U.S. Navy would stop any new missiles from going to Cuba. Kennedy would thus buy time for trying to talk the Russians into removing the missiles already there. By the following weekend, however, having used the time to no avail, it seemed, Kennedy was back at his starting point. The question again was whether to bomb only the missile sites or to go also for airfields. But on the second Sunday Khrushchev announced that he would withdraw the missiles. The story thus became one of success.

It may be that the only decision-making that mattered was Moscow's. The main American contribution may have been delay that allowed the Soviets to collect themselves. We suspect that American decisions and nondecisions had some more independent influence on the outcome. Whatever the case, as we look back, it seems clear to us that deliberate prolonging of the crisis, together with various moves aimed at producing a peaceful settlement, originated in or were at least much influenced by resort to history in ways not ordinary for American government officials. If the happy outcome was due even in part to those choices by Kennedy and his ExComm, then *un* usual uses of history perhaps deserve part of the credit.

Kennedy and his ExComm departed from standard practice first of all in subjecting analogies to serious analysis. The President early invited into ExComm former Secretary of State Dean Acheson, at that time a lawyer in private practice. Acheson favored a quick air strike. Hearing the Pearl Harbor analogy, he judged it, as he was to write later, "silly" and "thoroughly false and pejorative." He told the President that there were no points of similarity and many points of difference, to wit:

> [A]t Pearl Harbor the Japanese without provocation or warning attacked our fleet thousands of miles from their shores. In the present situation the Soviet Union had installed missiles ninety miles from our coast—while denying they were doing so—offensive weapons that were capable of lethal injury to the United States. This they were doing a hundred and forty years after the warning given in [the Monroe Doctrine]. How much warning was necessary to avoid the stigma of "Pearl Harbor in reverse."[6]

For ExComm and perhaps for the President, the effect of Acheson's analysis was the reverse of that intended. By stripping away all the dissimilarities, Acheson exposed the analogy's relevant point. Robert Kennedy responded to Acheson by saying, "For 175 years we had not been that kind of country. A sneak attack was not in our traditions." Then—not earlier—Secretary of the Treasury Dillon was won over. "I felt that I was at a real turning point in history," he recalled later, "I knew then that we should not undertake a strike without warning."[7]

All in all, the proceedings of ExComm are distinguished by the extent—unusual—to which analogies were invoked sparingly and, when invoked, were subjected to scrutiny. "Suez" did not last. A State Department lawyer referred to FDR's "Quarantine Address" of 1937 when suggesting that "quarantine" be substituted for "blockade," but no one represented the situations as analogous. Though Sorensen recalls talk of the Berlin blockade of 1948–49 and of the Bay of Pigs affair of 1961, possible points of comparison do not seemed to have gripped anyone's imagination.[8] When Kennedy went on television he referred to the "clear lesson" of the 1930s as one reason for demanding that the Russians back off. But that was rhetoric. The available records of ExComm debates are innocent of any allusion to "lessons" of the 1930s.

ExComm's second noteworthy departure from usual practice took the form of attention to the issue's history—to its sources and its context.

Kennedy himself had much to do with this, in part just by the choices he made in forming ExComm. He put a high premium on secrecy. "Maybe a lot of people know about what's there," he said at the initial meeting, "but what we're going to do about it ought to be, you know, the tightest of all, because otherwise we bitch it up." Nevertheless, he included in ExComm men who did not have to be there. Dillon is one example. The Treasury Department had no title to representation. Of course, Kennedy could see a partisan storm coming. “We’ve just elected Capehart . . . and Ken Keating will probably be the next President,” he said to one aide soon after discovery of the missile sites. Since Dillon had been Under Secretary of State for Eisenhower and was the most conspicious Republican in the subsequent Administration, Kennedy may have wanted him for the sake of seeming bipartisan. The same could hold true for his inviting former Defense Secretary Robert A. Lovett to join ExComm, for Lovett was a leader of New York’s Republican establishment. Or Kennedy may have turned to Dillon and Lovett just because he valued their judgment. Whatever the case, he got as a bonus the benefit of long and wide-ranging experience. He had around him men whose memories of dealing with the Soviet Union reached all the way back to World War II. He also called in Charles Bohlen and Llewellyn Thompson, two of the most senior serving members of the State Department’s Russian service, and Edwin Martin from the State Department’s Latin American bureau. Those three had memories, the first two of Russia and the third of Cuba, which also went far back.[9]

We suspect that this result was not accidental. Looking at the whole record of ExComm, one sees Kennedy himself repeatedly raising questions about the actual history of the issue. “I don’t know enough about the Soviet Union,” he said on the very first day, “but if anybody can tell me any other time since the Berlin blockade where the Russians have given us so clear a provocation, I don’t know when it’s been because they’ve been awfully cautious really.” He went on to wonder aloud whether the crisis might have been averted if he had said something more clearly, earlier (in retrospect, a telling criticism). He kept trying to find out when the Russians had decided to install the missiles, seeking in the timing some clue to their possible motives. On his orders the CIA produced a detailed review of the history of Soviet military aid to Cuba. During the terrifying six days between his public speech and Khrushchev’s backdown, Kennedy also asked an ExComm planning subcommittee to give high priority to a paper on “the Cuban base problem in perspective.”[10]

Third, in unusual degree Kennedy and his ExComm looked hard at key presumptions. During the initial meetings, the President said “it doesn’t make any difference if you get blown up by an ICBM flying from the Soviet Union or one that was ninety miles away.”[11] When Sorensen summarized for Kennedy the first two days' ExComm deliberations, he wrote: "It is generally agreed that these missiles, even when fully operational, do not significantly alter the balance of power. . . . Nevertheless it is generally agreed that the United States cannot tolerate the known presence of offensive nuclear weapons in a country ninety miles from our shore, if our courage and commitments are ever to be believed by either allies or adversaries." Though no one paused over it at the time, an early exchange between Kennedy and Bundy exposed a weakness in that particular pair of presumptions:

> KENNEDY: It's just as if we suddenly began to put a major number of MRBMs [medium range missiles] in Turkey. Now that'd be goddamn dangerous, I would think.
>
> BUNDY: Well we did, Mr. President.

In fact, since 1957 the United States had had in Turkey Jupiter missiles of greater range than most of the Soviet missiles going into Cuba. Fifteen were still there.[12] The Soviets had "tolerated" them throughout.

In later ExComm sessions the President included one of McNamara's subordinates, Assistant Secretary of Defense Paul Nitze. In doing so, Kennedy dropped another tap into the past, for Nitze had been Acheson's chief planner back in Truman's time. Nitze challenged the proposition that the Cuban missiles did not affect the balance of power. The existing missile gap, he said, gave the United States an unquestionable "second-strike" capacity. The Russians knew the United States could devastate their country even if they successfully staged an all-out surprise attack. That knowledge presumably made them cautious about running any risk of war. With missiles in place in Cuba, Nitze argued, the Russians might reason differently. They might suppose that an all-out surprise attack could destroy enough American missiles and bombers so that Soviet territory would not suffer terrible damage. In any case, they might suppose that their home-based and Cuban-based missiles together posed such danger to the United States that the American government would not risk war over, for example, Berlin. Nitze argued that the missiles in Cuba thus made a real difference.

Given the President's puzzlement as to why the Soviet Union had suddenly ceased to behave conservatively, Nitze's argument had some force with Kennedy. He at least altered his previous presumption, taking thereafter the position that the missiles were more than symbolic. As one result he became clear in his own mind that the missiles mattered much more than did the Soviet bombers also going into Cuba. He pressed for removal of all "offensive" weapons, including the bombers, and the Russians in fact withdrew both; but Kennedy told ExComm that "we should not get 'hung up' on the . . . bombers."[13]

Another presumption tested and changed concerned U.S. capacity for a "surgical" air strike, one that would take out only the missile sites. Since military planners wanted to protect U.S. bombers by suppressing Cuban and Soviet air defenses, they exaggerated somewhat the difficulty of effectively bombing only the missile sites. Because McNamara had misgivings akin to Robert Kennedy's about any air strikes, he may have encouraged that exaggeration. Nevertheless, at least through the first few days, several members of ExComm believed that Kennedy should order a "surgical" strike and would end up doing so.

Like the presumption that missiles were missiles, wherever placed, the presumption that a "surgical" strike could be effected gave way less because scrutinized or explicitly tested against the historical record than because it was questioned by men who had lived relevant history. No one tallied up the precision of past air operations, but some of those present had seen a lot of them. Lovett, himself a one-time Navy flier, had been in World War II the civilian in charge of U.S. land-based air forces. That fact counted when he spoke for a naval blockade in preference to an air strike. Robert Kennedy was ever afterward to treasure Lovett's use of the quotation, "Good judgment is usually the result of experience. And experience is frequently the result of bad judgment."[14]

The thirteen days of the missile crisis saw many other presumptions challenged. McNamara and the Chief of Naval Operations exchanged furious words because McNamara questioned the presumption that the Navy knew how to put in effect the quarantine the President had ordered. Secretary Rusk provoked a lot of paper-writing in his own department and in the CIA by voicing doubt as to whether Castro was truly in the Russians' pocket. In fact, the Navy knew exactly what to do, and, though Castro was sometimes angry with the Russians, he never showed for a moment an inclination to strike a deal at their expense.[15] Nevertheless, Kennedy and his ExComm seem to us exemplary for the extent to which they asked: How well-founded are the presumptions on which we plan to act?

Fourth, Kennedy and ExComm showed uncommon interest in the history in the heads of their adversaries. Kennedy's questions at the first ExComm meeting were about the Soviet Union, conceived as a single rational actor. He asked, in effect, Why is *he* doing this to *me?* Most high-ranking officials involved in international disputes ask that type of question. Early speculation is anthropomorphic. “This is a left hook designed to make him tougher when he comes at us in November, presumably on Berlin,” hazarded one ExComm participant endorsing the “Suez” thesis.[16] All that distinguished Kennedy’s initial formulation was his retrospection—his interest in when the rational actor had decided to depart from a previous line of conduct. As the crisis continued, however, Kennedy and others began to conceive of the Soviet government more as a collection of individuals.

Coached chiefly by Thompson, members of ExComm began to consider the possibility that certain U.S. actions might provoke Khrushchev to act impulsively rather than out of cool reasoning. ExComm members also took into some account the possibility that pride might affect the Soviet military in case of an attack on *their* missile sites. So far as we can tell, neither Kennedy nor any member of ExComm wondered aloud about the Russian history that Khrushchev and other Soviet leaders had experienced—the Revolution, the civil war, the Great Purge, World War II, de-Stalinization, the split with China, and other great events. On the other hand, Thompson surely had some of that history in his own mind. Probably remembering the Soviet struggle for full diplomatic recognition and the Soviet role in designing the UN, Thompson talked to ExComm of how the Russians might be influenced by a vote of the Organization of American States. They set high store on legal formalities, he said. Also, Thompson predicted that they would press for removal of the U.S. missiles in Turkey: “they like parallels.”[17]

According to Robert Kennedy, the President tried constantly to put himself in Khrushchev’s position. Once during the crisis he even described to Ben Bradlee of the *Washington Post* how he thought he would feel if in the Kremlin, but he cautioned Bradlee that his words were off the record. “It isn’t wise politically to understand Khrushchev’s problem in quite this way.”[18]

In the climactic hours of the crisis Kennedy received two messages from Khrushchev, the first a rambling four-part cable seeming to offer withdrawal of the missiles in return for a U.S. pledge not to invade Cuba, the second, more curt and formal, seeming to retract that offer. Instead of returning to “left hook” imagery, the President and members of ExComm speculated about factionalism in the Kremlin. They visualized Khrushchev, stamping around his giant office in the Kremlin, possibly not altogether sober, dictating to a secretary, and sending off the text without showing it to anyone. They imagined other members of the Politburo bending over the second cable and tightening its wording. All that made easier their decision to ignore the second cable and simply say yes to the first. Some of them thought later that this tactic was the source of their success, the means to bring the crisis to a close, yet they probably would not have settled on it had they not by then begun to think of Khrushchev as a person, with a history of his own.

Fifth, Kennedy and his ExComm paid attention to organizational histories. They did not do so in quite the way we shall advocate later. They thought of how organizations behaved without asking explicitly how they had behaved over time, and why. But the fact that they took organizational behavior into account at all distinguishes them from ninety-odd percent of the decision-making groups of which we have personal knowledge.

Again, Kennedy himself gave ExComm its cue. He seemed to understand in his bones the tendency of large organizations to act today as they acted yesterday. He pursued his own hunches about American performance. Among other things, he sent the CIA to photograph Air Force planes at Florida bases. The pictures showed that, contrary to his orders, the planes were lined up in the highly vulnerable standard position—wing tip to wing tip—just as at Manila twenty-one years before. Schooled in the inertia of military procedures as a junior officer in World War II, Kennedy was annoyed but not surprised.

Kennedy and ExComm were encouraged toward the quarantine option by Thompson's reminder that Russian military organizations practiced extreme secretiveness. Built into organizational routines, that secretiveness would make the Russians hesitant, they hoped, to risk having their ships boarded and searched.[19] In fact the Russians halted all missile-carrying ships well outside the quarantine line.

Thompson and other Sovietologists also helped Kennedy and the members of ExComm appreciate the possibility that events on the Soviet side could be products of organizational routine or momentum rather than deliberate purpose. Just when Kennedy and his advisers were trying to puzzle out the differences between the two Khrushchev cables, a U.S. U-2 plane was shot down over Cuba. It would have been easy, even natural, to see that as a signal confirming a hardened Soviet line. Kennedy, however, accepted Thompson's counsel not to read political significance into what could well have been just a Soviet air defense unit acting according to the book. Others urged at least tit-for-tat retaliation, but Kennedy chose to wait. As a result, no U.S. strike on a Soviet air defense site complicated Khrushchev's decision to accept Kennedy's terms. (He meanwhile must have had to show equal good sense, for another American U-2 blundered coincidentally into Soviet air space, moving Kennedy to explain, "There's always some son-of-a-bitch who doesn't get the word.")[20]

After the crisis ended, Kennedy said he thought the odds on war had been "between one out of three and even." At the same time, according to Robert Kennedy, the President believed Khrushchev to be "a rational, intelligent man who, if given sufficient time and shown our determination, would alter his position." The historian Arthur M. Schlesinger, who knew the Kennedy brothers well and has written movingly of both, offers an explanation for the seeming contradiction in terms that seem plausible to us, namely: "Kennedy's grim odds were based on fear, not of Khrushchev's intentions, but of human error, of something going terribly wrong down the line."[21] If that is accurate, then the taking into account of historical patterns in organizational conduct may have been exceptionally important among the unusual practices exemplified by Kennedy and ExComm.

But a final peculiarity in their practice strikes us as perhaps most important of all. In unusual degree, Kennedy and his ExComm saw the issues before them as part of a time sequence beginning long before the onset of crisis and continuing into an increasingly indistinct future. The more Kennedy and ExComm deliberated, the more they weighed consequences and the more they shifted from the simple question of what to do *now* to the harder question: How will today's choices appear when they are history—when people look back a decade or a century hence?

The initial debate in ExComm involved no evident thought beyond the next week or so. As early as the evening of the first day, however, a few participants had lifted their sights. "I don't know what kind of a world we live in after we've struck Cuba," McNamara said. And Bundy: "Our principal problem is to try and imaginatively to think what the world would be like if we do this and what it will be like if we don't."[22]

The President's own way of looking ahead appeared most clearly in his eventual handling of the parallel problem—those U.S. missiles in Turkey. In early sessions the notion of a swap had been dismissed as unthinkable. By the tenth and eleventh days of the crisis, on the other hand, Kennedy and his advisers talked about the possibility in terms of how it might fit a long sequence of events. Most of Kennedy's advisers still argued against removal, predicting that the Turks would protest and that other NATO governments would then make endless trouble. While those advisers were looking back and looking ahead, they did so with the eyes of men whose worlds were made of foreign offices and defense ministries. Kennedy saw the question more broadly. As the minutes of one ExComm meeting record:

> The President recalled that over a year ago we wanted to get the Jupiter missiles out of Turkey because they had become obsolete and of little military value. If the missiles in Cuba added 50 percent to the Soviet nuclear capability, then to trade these missiles for others in Turkey would be of great military value. But we are now in the position of risking war in Cuba and in Berlin over missiles in Turkey which are of little military value. From the political point of view, it would be hard to get support on an airstrike against Cuba because many would think we could make a good trade if we offered to take the missiles out of Turkey in the event the Russians would agree to remove the missiles from Cuba. We are in a bad position if we appear to be attacking Cuba for the purpose of keeping useless missiles in Turkey. We . . . have to face up to the possibility of some kind of a trade over missiles.

Robert Kennedy found a way around the dilemma. In very private conversations with Dobrynin, he promised that the U.S. missiles would be out of Turkey in four to five months. He also said not only that he would deny ever making such a promise but that, if any Russian revealed it, all deals would be off. The bargain was struck. No word was said of any trade other than Soviet withdrawal of missiles from Cuba in return for assurances that the United States would not invade Cuba. Five months later the U.S. missiles came out of Turkey.[23]

That the President came to see such issues in a stream of time is still more sharply illustrated by remarks he made to his brother about World War I. He had recently read a book on the outbreak of that war. It had reminded him of having heard in college of a former German Chancellor who, asked about the reasons for World War I, had replied, "Ah, if we only knew." Kennedy was not invoking an analogy, not even in the vein of his brother's reference to Pearl Harbor. Instead, we think, 1914 came to his mind because he saw himself as part of a long procession of political leaders on whose decisions many lives might depend. The book had been Barbara Tuchman's *Guns of August,* and Kennedy said to his brother, "I am not going to follow a course which will allow anyone to write a comparable book about this time, *The Missiles of October*. If anybody is around to write after this, they are going to understand that we made every effort to find peace and every effort to give our adversary room to move."[24]

The missile crisis may have been only accidentally a success story. We do not know—may never know—why the Russians decided as they did, and different decisions by them could have led toward a horrible ending. To the extent that American decisions shaped the outcome, uncharacteristic ways of using—and avoiding—history do not suffice as explanations of their clarity and cogency. Those choices were products of extraordinary conditions: intense concentration; effective secrecy sustained by media cooperation (after Watergate that would be thought treasonous to the First Amendment); a high average of mind—these people were not tagged "the best and the brightest" for nothing—along with breadth of experience. While some staff work could have been better, the run of the mill seldom is as good as the poorest was then. Similar conditions are not often likely to obtain at any level. Few issues can carry on their faces the blazing show of novelty and gravity combined—arresting the attention needed for frontal exploration of concerns and options—as did the first directly military confrontation between Washington and Moscow in the missile age.

Even so, the uses made of history appear to have contributed, demonstrably, to the high quality of analysis and management apparent during the missile crisis. Right or wrong, Kennedy had the wherewithal for reasoned and prudent choice, and resort to history helped produce it. One cannot expect that lesser choices on more mundane matters, either at the top level or down below, will often, if ever, benefit from the special factors present in 1962. One cannot even count on those factors in the next crisis. But why not hope that in choice-making, low-level or high, the preparatory work takes heed of history in ways to emulate—or, better still, improve upon—this Cuban instance?