CHAPTER 1

THE OPPOSITE OF CERTAINTY

During the past century, research in the medical, social, and economic sciences has led to major improvements in longevity and living conditions. Statistical methods grounded in the mathematics of probability have played a major role in much of this progress. Our confidence in these quantitative tools has grown, along with our ability to wield them with great proficiency. We have an enormous investment of tangible and intellectual capital in scientific research that is predicated on this framework. We assume that the statistical methods as applied in the past so successfully will continue to be productive. Yet, something is amiss.

New findings often contradict previously accepted theories. Faith in the ability of science to provide reliable answers is being steadily eroded, as expert opinion on many critical issues flip-flops. Scientists in some fields seriously debate whether a majority of their published research findings are ultimately overturned[1]; the *decline effect* has been coined to describe how even strongly positive results often fade over time in the light of subsequent study[2]; revelations of errors in the findings published in prestigious scientific journals, and even fraud, are becoming more common.[3] Instead of achieving greater certainty, we seem to be moving backwards. What is going on?

Willful Ignorance: The Mismeasure of Uncertainty, First Edition. Herbert I. Weisberg.

Consider efforts to help disadvantaged children through early childhood educational intervention. Beginning around 1970, the U.S. government sponsored several major programs to help overcome social and economic disadvantage. The most famous of these, Project Head Start, aimed to close the perceived gap in cognitive development between richer and poorer children that was already evident in kindergarten. The aims of this program were admirable and the rationale compelling. However, policy debates about the efficacy and cost of this initiative have gone on for four decades, with no resolution in sight. Research on the impact of Head Start has been extensive and costly, but answers are few and equivocal.

Medical research is often held up as the paragon of statistical research methodology. Evidence-based medicine, based on randomized clinical trials, can provide proof of the effectiveness and safety of various drugs and other therapies. But cracks are appearing even in this apparently solid foundation. Low dose aspirin for prevention of heart attacks was gospel for years but is now being questioned. Perhaps the benefits are less and the risks, more than we previously believed. Hormone replacement therapy for postmenopausal women was considered almost miraculous until a decade ago when a landmark study overturned previous findings. Not a year goes by without some new recommendation regarding whether, how, and by whom, hormone replacement should be used.

These are not isolated instances. The ideal of science is an evolution of useful theory coupled with improved practice, as new research builds upon and refines previous findings. Each individual study should be a piece of a larger puzzle to which it contributes. Instead, research in the biomedical and social sciences is rarely cumulative, and each research paper tends to stand alone. We fill millions of pages in scientific journals with "statistically significant" results that add little to our store of practical knowledge and often cannot be replicated. Practitioners, whose clinical judgment should be informed by hard data, gain little that is truly useful to them.

TWO DEAD ENDS

If I am correct in observing that scientific research has contributed so little to our understanding of "what works" in areas like education, health care, and economic development, it is important to ask why this is the case. I believe that much of the problem lies with our research methodology. At one end of the spectrum, we have what can be called the quantitative approach, grounded in modern probability-based statistical methods. At the other extreme are researchers who support a radically different paradigm, one that is primarily qualitative and more subjective. This school of thought emphasizes the use of case studies and in-depth participatory observation to understand the dynamics of complex causal processes.

Both statistical and qualitative approaches have important contributions to make. However, researchers in either of these traditions tend to view those in the other with suspicion, like warriors in two opposing camps peering across a great divide. Nowadays, the statistical types dominate, because methods based on probability and statistics virtually define our standard of what is deemed "scientific." The perspective of qualitative researchers is much closer to that of clinicians but lacks the authority that the objectivity of statistics seems to provide.

Sadly, each side in this fruitless debate is stuck in a mindset that is too restricted to address the kinds of problems we face. Conventional statistical methods make it difficult to think seriously about causal processes underlying observable data. Qualitative researchers, on the other hand, tend to underestimate the value of statistical generalizations based on patterns of data. One approach willfully ignores all salient distinctions among individuals, while the other drowns in infinite complexity.

The resulting intellectual gridlock is especially unfortunate as we enter an era in which the potential to organize and analyze data is expanding exponentially. We already have the ability to assemble databases in ways that could not even be imagined when the modern statistical paradigm was formulated. Innovative statistical analyses that transcend twentieth century data limitations are possible if we can summon the will and imagination to fully embrace the opportunities presented by new technology.

Unfortunately, as statistical methodology has matured, it has grown more timid. For many, the concept of scientific method has been restricted to a narrow range of approved techniques, often applied mechanically. The result is to limit the scope of individual creativity and inspiration in a futile attempt to attain virtual certainty. Already in 1962, the iconoclastic statistical genius John Tukey counseled that data analysts "must be willing to err moderately often in order that inadequate evidence shall more often *suggest* the right answer."[4]

Instead, to achieve an illusory pseudo-certainty, we dutifully perform the ritual of computing a significance level or confidence interval, having forgotten the original purposes and assumptions underlying such techniques. This "technology" for interpreting evidence and generating conclusions has come to replace expert judgment to a large extent. Scientists no longer trust their own intuition and judgment enough to risk modest failure in the quest for great success. As a result, we are raising a generation of young researchers who are highly adept technically but have, in many cases, forgotten how to think for themselves.

ANALYTICAL ENGINES

The dream of "automating" the human sciences by substituting calculation for intuition arose about two centuries ago. Adolphe Quetelet's famous treatise on his statistically based "social physics" was published in 1835, and Siméon Poisson's masterwork on probability theory and judgments in civil and criminal matters appeared in 1837.[5,6] It is perhaps not coincidental that in 1834 Charles Babbage first began to design a mechanical computer, which he called an *analytical engine*.[7] Optimism about the potential ability of mathematical analysis, and especially the theory of probability, to resolve various medical, social, and economic problems was at its zenith.

Shortly after this historical moment, the tide turned. The attempt to supplant human judgment by automated procedures was criticized as hopelessly naïve. Reliance on mathematical probability and statistical methods to deal with such subtle issues went out of favor. The philosopher John Stuart Mill termed such uses of mathematical probability "the real opprobrium of mathematics."[8] The famous physiologist Claude Bernard objected that "statistics teach absolutely nothing about the mode of action of medicine nor the mechanics of cure" in any particular patient.[9] Probability was again relegated to a modest supporting role, suitable for augmenting our reasoning. Acquiring and evaluating relevant information, and reaching final conclusions and decisions remained human prerogatives.

Early in the twentieth century, the balance between judgment and calculation began to shift once again. Gradually, mathematical probability and statistical methods based on it came to be regarded as more objective, reliable, and generally "scientific" than human theorizing and subjective weighing of evidence. Supported by rapidly developing computational capabilities, probability and statistics were increasingly viewed as methods to generate definitive solutions and decisions. Conversely, human intuition became seen as an outmoded and flawed aspect of scientific investigation.

Instead of serving as an adjunct to scientific reasoning, statistical methods today are widely perceived as a corrective to the many cognitive *biases* that often lead us astray. In particular, our naïve tendencies to misinterpret and overreact to limited data must be countered by a better understanding of probability and statistics. Thus, the genie that was put back in the bottle after 1837 has emerged in a new and more sophisticated guise. Poisson's ambition of rationalizing such activities as medical research and social policy development is alive and well. Mathematical probability, implemented by modern analytical engines, is widely perceived to be capable of providing scientific evidence-based answers to guide us in such matters.

Regrettably, modern science has bought into the misconception that probability and statistics can arbitrate truth. Evidence that is "tainted" by personal intuition and judgment is often denigrated as merely descriptive or "anecdotal." This radical change in perspective has come about because probability appears capable of objectively *quantifying* our uncertainty in the same unambiguous way as measurement techniques in the physical sciences. But this is illusory:

> Uncertain situations call for probability theory and statistics, the mathematics of uncertainty. Since it was precisely in those areas where uncertainty was greatest that the burden of judgment was heaviest, statistical tools seemed ideally suited to the task of ridding first the sciences and then daily life of personal discretion, with its pejorative associations of the arbitrary, the idiosyncratic, and the subjective. Our contemporary notion of objectivity, defined largely by the absence of these elements, owes a great deal to the dream of mechanized inference. It is therefore not surprising that the statistical techniques that aspire to mechanize inference should have taken on a normative character. Whereas probability theory once aimed to describe judgment, statistical inference now aims to replace it in the name of objectivity. … Of course, this escape from judgment is an illusion. … No amount of mathematical legerdemain can transform uncertainty into

> certainty, although much of the appeal of statistical inference techniques stems from just such great expectations. These expectations are fed … above all by the hope of avoiding the oppressive responsibilities that every exercise of personal judgment entails.[10]

Probability by its very nature entails ambiguity and subjectivity. Embedded within every probability statement are unexamined simplifications and assumptions. We can think of probability as a kind of devil's bargain. We gain practical advantages by accepting its terms but unwittingly cede control over something fundamental. What we obtain is a special kind of knowledge; what we give up is conceptual understanding. In short, by willingly remaining *ignorant*, in a particular sense, we may acquire a form of useful knowledge. This is the essential paradox of probability.

WHAT IS PROBABILITY?

Among practical scientists nowadays, the true *meaning* of probability is almost never discussed. This is really quite remarkable! The proper interpretation of mathematical probability within scientific discourse was a hotly debated topic for over two centuries. In particular, questions about the adequacy of mathematical probability to represent fully our uncertainty were deemed important. Recently, however, there has been virtually no serious consideration of this critical issue.

As late as the 1920s, a variety of philosophical ideas about probability and uncertainty were still in the air. The central importance of probability theory in a general sense was recognized by all. However, there was wide disagreement over how the basic concept of probability should be defined, interpreted, and applied. Most notably, in 1921 two famous economists independently published influential treatises that drew attention to an important theoretical distinction. Both suggested that the conventional concept of mathematical probability is incomplete.

In his classic, *Risk, Uncertainty and Profit*, economist Frank Knight described the kind of uncertainty associated with ordinary probability by the term *risk*.[11] The amount of risk can be deduced from mathematical theory (as in a game of chance) or calculated by observing many outcomes of similar events, as done, for example, by an insurance company. However, Knight was principally concerned with probabilities that pertain to another level of uncertainty. He had particularly in mind a typical business decision faced by an entrepreneur. The probability that a specified outcome will result from a certain action is ordinarily based on subjective judgment, taking into account all available evidence.

According to Knight, such a probability may be entirely intuitive. There may be no way, even in principle, to verify this probability by reference to a hypothetical reference class of similar situations. In this sense, the probability is completely subjective, an idea that was shared by some of his contemporaries. However, Knight went further by suggesting that this subjective probability also carries with it some sense of how much *confidence* in this estimate is actually entertained. So, in an imprecise but very important way, the numerical measure of probability is only a *part* of the full uncertainty assessment. "The action which follows upon an opinion depends as much upon the confidence in that opinion as upon the favorableness of the opinion itself." This broader but vaguer conception has come to be called Knightian uncertainty.

Knightian uncertainty was greeted by economists as a new and radical concept, but was in fact some very old wine being unwittingly rebottled. One of the few with even an inkling of probability's long and tortuous history was John Maynard Keynes. Long before he was a famous economist,[12] Keynes authored *A Treatise on Probability*, completed just before World War I, but not published until 1921. In this work, he probed the limits of ordinary probability theory as a vehicle for expressing our uncertainty. Like Knight, Keynes understood that some "probabilities" were of a different character from those assumed in the usual theory of probability. In fact, he conceived of probability quite generally as a measure of *rational belief* predicated on some particular *body of evidence*.

In this sense, there is no such thing as a unique probability, since the evidence available can vary over time or across individuals. Moreover, sometimes the evidence is too weak to support a firm numerical probability; our level of uncertainty may be better represented as entirely or partly *qualitative*. For example, my judgment about the outcome of the next U.S. presidential election might be that a Democrat is somewhat less likely than a Republican to win, but I cannot reduce this feeling to a single number between zero and one. Or, I may have no idea at all, so I may plead *complete ignorance*. Such notions of a non-numerical degree of belief, or even of complete ignorance (for lack of any relevant evidence), have no place in modern probability theory.

The mathematical probability of an event is often described in terms of the odds at which we should be willing to bet for or against its occurrence. For example, suppose my probability that the next president will be a Democrat is 40%, or 2/5. Then for me, the fair odds at which to bet on this outcome would be 3:2. So I will gain 3 dollars for every 2 dollars wagered if a Democrat actually wins, but lose my 2-dollar stake if a Republican wins. However, a full description of my uncertainty might also reflect how confident I would be about these odds. To force my expression of uncertainty into a precise specification of betting odds, as if I *must* lay a wager, may be artificially constraining.

Knight and Keynes were among a minority who perceived that uncertainty embodies something more than mere "risk." They understood that uncertainty is inherently *ambiguous* in ways that often preclude complete representation as a simple number between zero and one. William Byers eloquently articulates in *The Blind Spot* how such ambiguity can often prove highly generative and how attempts to resolve it completely or prematurely have costs.[13]

As a prime example, Byers discusses how the ancient proto-concept of "quantity" evolved over time into our current conception of numbers:

> A unidirectional flow of ideas is at best a reconstruction. It is useful and interesting but it misses something. It inevitably takes the present situation to be definitive. It tends to show how our present knowledge is superior in every way to the knowledge of the Greeks, for example. In so doing, it ignores the possibility that the Greeks knew things we do not know, that we have forgotten or suppressed. It seems heretical to suggest, but is nevertheless conceivable, that the Greek conception of quantity was in a certain way richer than our own, that their conception of number was deeper than ours. It was richer in the sense that a metaphor can be rich—because it comes with a large set of connoted meanings. It may well be that historical progress in mathematics is in part due to the process of abstraction, which inevitably involves narrowing the focus of attention to precisely those properties of the situation that one finds most immediately relevant. This is the way I shall view the history of mathematics and science—as a process of continual development that involves gain and loss, not as the triumphant march toward some final and ultimate theory.

In very much the same way, our modern idea of probability emerged from earlier concepts that were in some respects richer, and perhaps deeper.[14]

When Knight and Keynes wrote, the modern interpretation of probability had already almost completely crystallized. Shortly afterwards, further "progress" took the form of "narrowing the focus" even more. Although the broader issues addressed by Knight and Keynes were ignored, there remained (and still remains) one significant philosophical issue. Should probability be construed as essentially *subjective* or *objective* in nature? Is probability purely an aspect of personal thought and belief *or* an aspect of the external world? I will suggest that this is a false dichotomy that must be transcended.

As statistical methods became more prominent in scientific investigations, objectivity became paramount. It became widely accepted that science must not reflect any subjective considerations. Rather, it must deal with things that we can measure and count objectively. Thus, mathematical probability, interpreted as the frequency with which observable events occur, became the yardstick for measurement in the context of scientific research. This link to empirical reality created a false sense of objectivity that continues to pervade our research methodology today, although a more subjective interpretation has recently made some limited inroads.

From our modern viewpoint, there appears to be a sharp distinction between the subjective and objective interpretations of probability. However, to the originators of mathematical probability, these two connotations were merged in a way that can seem rather muddled to us. Were they confused, or do we fail to grasp something meaningful for them now lost to us? Is there, as Byers intimates, a "transcendental" perspective from which this distinction would no longer seem meaningful? If so, it might point the way toward a resolution of the conflict between apparently opposing ways of thinking about science. That, in turn, could help bridge the gap between scientific research and clinical practice.

UNCERTAINTY

At the core of science is the desire for greater certainty in a highly unpredictable world. Probability is often defined as a measure of our degree of certainty, but what is certainty? If I am certain that a particular event will occur, what does that mean? For concreteness, suppose I have just enrolled in a course on a subject with which I am not very familiar. For the moment, let us assume I am absolutely *certain* of being able to pass this course.

Dictionary definitions of the word "certain" contain phrases like "completely confident" and "without any doubt." But what conditions would allow us to be in such a state of supreme confidence? Obviously, our knowledge of the situation or circumstances must be adequate for us to believe that the event *must* occur. My certainty about passing the course would rest on a matrix of information and beliefs that justify (for me) the necessary confidence.

Now, suppose that, on the contrary, I am *not certain* that I can pass this course. Clearly, this implies that I am lacking in certainty, but what does this mean? I would submit that uncertainty has two quite different connotations, or aspects. On one hand, my uncertainty can arise from *doubt*. So, the opposite of being sure of passing the course is being extremely doubtful. On the other hand, being uncertain could also mean that I just *do not know* whether I will be able to pass the course. I may suffer from confusion, because the situation facing me seems *ambiguous*. So the opposite of being highly confident would be something like having no idea, being literally "clueless."

The situation can be described graphically as in Figure 1.1. We can conceptualize our degree of uncertainty as the resultant of two psychological "forces." The horizontal axis represents our degree of *doubt* and the vertical axis, the degree of *ambiguity* we perceive. In general, certainty corresponds to the absence of *both* doubt and ambiguity. Our degree of uncertainty increases according to the "amounts" of doubt and ambiguity.

To be more specific, ambiguity pertains generally to the clarity with which the situation of interest is being conceptualized. How sure am I about the mental category, or classification, in which to place what I perceive to be happening? In order to exercise my judgment about what is likely to occur, I must have a sense of the relevant features of the situation. So, reducing uncertainty by resolving ambiguity, at least to some extent, seems to be a necessary prerequisite for assessing doubt. However, the relationship between ambiguity and doubt can be complex and dynamic. We certainly cannot expect to eliminate ambiguity completely before framing a probability.

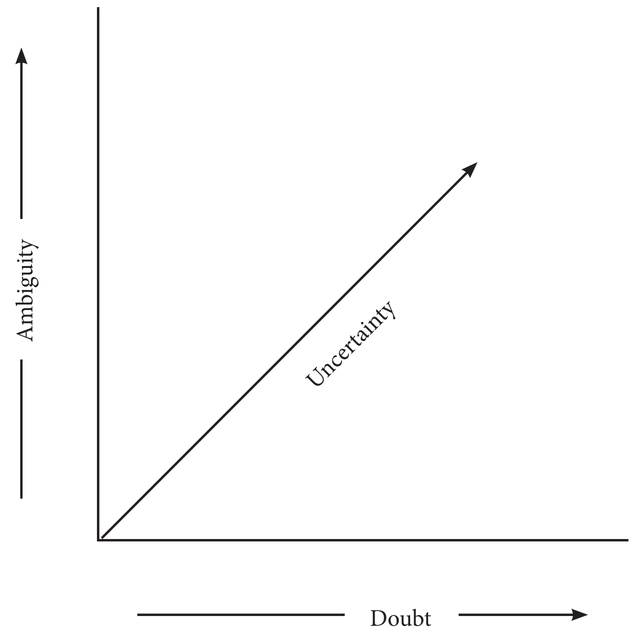

FIGURE 1.1 The two dimensions of uncertainty: ambiguity and doubt.

In relation to probability, there is an important difference between these two dimensions of uncertainty. It seems natural to think of a degree of doubt as a *quantity*. We can, for example, say that our doubt that the Chicago Cubs will win the World Series this year is greater than our doubt that the New York Yankees will be the champions. We may even be able to assign a numerical value to our degree of doubtfulness. Ambiguity, on the other hand, seems to be essentially *qualitative*. It is hard to articulate what might be meant by a "degree of ambiguity."

Probability in our modern mathematical sense is concerned exclusively with the doubt component of uncertainty. For us, the probability of a certain event is assigned a value of 1.0, or 100%. At the opposite end of the spectrum, an event that is deemed impossible, or virtually impossible, has a probability value of 0.0, or 0%. When we say that something, such as passing a course, has a probability of 95%, we mean that there exists a small degree of doubt that it will actually occur. Conversely, a probability of 5% implies a very strong doubt. In order to achieve such mathematical precision, we must suppress some subtlety or complexity that creates ambiguity. That way, our uncertainty can be ranged along a single dimension; our degree of confidence that the event *will* happen becomes identical to our lack of confidence that it *will not* happen.

As I will be explaining, this modern *mathematical* conception of probability emerged quite recently, about three centuries ago. Prior to its invention, there had existed for thousands of years earlier concepts of probability that were, indeed, "in a certain way richer than our own." Especially important, these archaic ideas about uncertainty encompassed *both* dimensions of uncertainty, often without clearly distinguishing between them. Dealing with uncertainty implicitly entailed two challenges: attempting to *resolve the ambiguity* and to *evaluate the doubtfulness* of what we know. By essentially ignoring ambiguity in order to quantify doubt, we have obtained the substantial benefits of mathematical probability.

Since the 1920s, the equally important issue of ambiguity has been left outside the pale of scientific (and most philosophical) thinking about probability. The concerns raised by Keynes, Knight, and others back then were never addressed. In essence, they perceived the dangers in reducing probability to a technology for measuring doubt that ignores ambiguity. Failing to address this issue has led to the moribund state in which many areas of science now find themselves.

WILLFUL IGNORANCE

Suppose you are an emergency-room physician confronted by a new patient who displays an unusual constellation of symptoms. Rapid action is required, as the patient's condition is life-threatening. You are uncertain about the appropriate course of treatment. Your task is twofold: resolve your confusion about what type of illness you are observing and decide on the optimal therapy to adopt.

The diagnosis aims to eliminate, or at least minimize, any ambiguity pertaining to the patient's condition and circumstances. The physician's methodology may include a patient history, a physical examination, and a variety of clinical testing procedures. All of the resulting information is evaluated and integrated subjectively by the physician and possibly other specialist colleagues. The usual outcome is a classification of the patient into a specific disease category, along with any qualifying details (e.g., disease duration and severity, concomitant medications, allergies) that may be relevant to various potential treatment options. The process of attempting to resolve ambiguity in this situation, or in general, draws mainly on the clinician's expertise and knowledge. It entails logic and judgment applied to the array of evidence available.

Once the diagnosis is determined, however, the situation changes. The focus shifts to the selection of a treatment approach. The ambiguity about what is happening has been largely resolved. The remaining task is to choose from among the different therapeutic candidates. Putting aside the issue of side effects, the therapy offering the best chance of a cure will be selected. Not that long ago, this too was settled mainly by appealing to the presumed clinical expertise of the clinician (doctor, psychiatrist, social worker, teacher, etc.). Not any longer.

Since the 1950s, research to evaluate alternative treatment modalities has become increasingly standardized and objective. So-called evidence-based medicine depends heavily on statistical theory for the design, conduct, and analysis of research. This technology appears to generate knowledge that is demonstrably reliable because human subjectivity and fallibility have been eliminated from the process. Central to the modern research enterprise is probability theory. Probability defines the terms within which questions and answers are framed. Moreover, rather than merely advising the clinician, evidence-based recommendations based on statistics are intended to represent the "optimal" decision.[15]

When these new statistical methods were originally introduced, they promised to ameliorate serious problems that were then widespread, such as exaggerated claims of efficacy and outright quackery. However, it could not be imagined to what extent these safeguards would eventually come to *define* our standard of what constitutes respectable science. Statistical methods are now virtually the *only* way to conduct research in many fields, especially those that study human beings. What has resulted is a profound *disconnect* between clinical and statistical perceptions in many instances.

Research focuses on what is likely to happen "on the average" in certain specified circumstances. What, for example, is the effect on the mortality rate for middle-aged men who adopt a low-dose aspirin regimen? However, the clinician's concern is her particular patient. What will happen to

Sam Smith if he starts on an aspirin regimen tomorrow? So, she may balk at mechanically following some general guidelines that are alleged to be statistically optimal:

> Each of us is unique in the interplay of genetic makeup and environment. The path to maintaining or regaining health is not the same for everyone. Choices in this gray zone are frequently not simple or obvious. For that reason, medicine involves personalized and nuanced decision making by both the patient and doctor. … Although presented as scientific, formulas that reduce the experience of illness to numbers are flawed and artificial. Yet insurers and government officials are pressuring physicians and hospitals to standardize care using such formulas. Policy planners and even some doctors have declared that the art of medicine is passé, that care should be delivered in an industrialized fashion with nurses and doctors following operating manuals.[16]

In a real sense, clinicians and researchers tend to inhabit different conceptual worlds. The clinician is sensitive to the ambiguities of the "gray zone" in which difficult decisions must be made. She is in a land where the uncertainty is mainly of the "what is really going on here?" kind. For the researcher, on the other hand, the world must look black and white, so that the rules of probability math can be applied. This ambiguity blindness has become absolutely necessary. Without it, as we will see, the elaborate machinery of statistical methodology would come to a grinding halt. Consequently, there is no middle road between the clinical and statistical perspectives.

To be clearer on this point, let us hark back to our hypothetical problem of medical treatment. Suppose you have discovered the cause of the patient's symptoms, a rare type of virulent bacterial infection. Your problem now is to select which antibiotic to try first. There are three possibilities, each of which you have prescribed in the past many times. Your decision will hinge primarily on the probability of achieving a cure *for this patient*. We are accustomed to thinking that there actually exists, in some objective sense, a true probability that applies to this patient. In fact, there is no such probability out there!

A probability is a mental construct. In this sense, it is entirely subjective, or personal, in nature. However, probability must also have something to do with observations in the outside world. Indeed, an important (perhaps the only) relevant source of evidence may be a statistical rate of cure that you can find in the medical literature. Surely, these rates (percentages) can be interpreted as probabilities, or at least as approximations to them. Moreover, because these statistics are objective and precise, they are ordinarily expected to trump any subjective considerations.

The problem is that the "objective" probability may not be applicable to your particular patient. You may have specific knowledge and insight that influence your level of ambiguity or of doubt. For example, you might know that Sam Smith tends to comply poorly with complicated instructions for taking medicine properly. So, the statistically indicated treatment modality might not work as well for him as for the typical subject in the clinical studies. Ideally, you would possess some system for rationally taking account of all factors, both qualitative and quantitative, that seem relevant. However, the statistically based probability is not open to debate or refinement in any way. That is because probability by its very nature entails *willful ignorance*.

My term willful ignorance refers to the inescapable fact that probabilities are not geared directly to individuals. An assessment of probability can of course be *applied* to any particular individual, but that is a matter of judgment. By choosing a statistically based probability, you effectively regard this individual as a random member of the population upon which the statistics were derived. In other words, you *ignore* any distinguishing features of the individual or his circumstances that might modify the probability.

TOWARD A NEW SCIENCE

Relying uncritically on statistics for answers has become so second-nature to us that we have forgotten how recent and revolutionary this way of thinking really is. That is the crux of the problems we now face. Fortunately, there is a path out of the stagnation that plagues our research currently, and it is surprisingly simple, *in theory*. Unfortunately, practical implementation of this idea will require a seismic shift in behavior to achieve. In a nutshell, we must learn to become more *mindful* in applying probability-based statistical methods and criteria.

Mindfulness can be described as a way of perceiving and behaving that is characterized by openness, creativity, and flexibility. Psychologist Ellen

Langer has suggested several qualities that tend to characterize a mindful person.

- The ability to create new categories
- Openness to new information
- Awareness of multiple perspectives
- A focus on process more than outcome
- A basic respect for intuition.

These reflect precisely the attitude of a scientist who is motivated primarily by potential *opportunities* to advance human knowledge. Such an individual thrives on ambiguity, because it offers a wealth of *possibilities* to be explored.

In contrast, statistical methodology as it is applied today does not encourage these attributes. Rather, it has become *mindless* in its mechanical emphasis on prespecified hypotheses about average effects and formal testing procedures. It is no wonder that clinicians and qualitatively oriented researchers are uncomfortable with such unnatural modes of thinking:

> Just as mindlessness is the rigid reliance on old categories, mindfulness means the continual creation of new ones. Categorizing and recategorizing, labeling and relabeling as one masters the world are processes natural to children. They are an adaptive and necessary part of surviving in the world.

These dynamic processes are equally essential in scientific research to cope with and resolve ambiguity. By relying so heavily on statistical procedures based on probability theory, ambiguity is effectively swept under the rug. This is a *fundamental* problem, because the essence of probability is quantification of doubt, which requires ambiguity about categories and labels to be willfully ignored. So, the problem of ambiguity cannot truly be evaded within the framework of probability, only sidestepped. A better approach is to broaden our understanding of uncertainty in order to resolve ambiguity more productively.

Am I arguing that mathematical probability and statistical methods should be avoided? Far from it. We will need these tools to address the problems of a much more data-rich future. However, our research methodology needs somehow to make more room for mindfulness, even though that will entail confronting ambiguity as well as doubt. Doing so may require us to cultivate a greater degree of tolerance for error, but this is unavoidable. Being wrong, as Kathryn Schulz reminds us, is normal; the problem is how we deal with this ever-present possibility.[17]

We must avoid worrying about mistakes to the point of stifling creativity. After all, it took Einstein 10 years and countless false leads before coming up with the general theory of relativity.[18] It is OK to be wrong, as long as you are in the mode of continually testing and revising your theories in the light of evidence. In research that relies on the analysis of statistical data, that means placing more emphasis on successful *replication* of findings. By maintaining a balance between theoretical speculation and empirical evidence, we can increase the chances of generating knowledge that will make sense to both the researcher and the clinician.

Accomplishing this necessary evolution of methodology will entail both technical challenges and a major alteration of our scientific culture and incentives. Mathematical probability and statistical analysis will continue to play important roles in the future of research. But these tools must continue to develop in ways that take fuller advantage of the emerging opportunities. To conclude, I offer a personal anecdote that I have often used to exemplify the kind of mindful statistical analysis that will be necessary:

> The first legal case in which I provided statistical expertise was an employment discrimination lawsuit against a Boston-based Fortune 500 company. The plaintiffs were convinced that black workers were being systematically prevented from rising to higher-level positions within the manufacturing division of the company. … I dutifully subjected the data to various standard analyses, searching for an effect of race on promotion rates, but came up empty. Despite repeated failures, I harbored a nagging suspicion that something important had been overlooked.
>
> I began to scrutinize listings of the data, trying to discern some hidden pattern behind the numbers. Preliminary ideas led to further questions and to discussions with some of the plaintiffs. This interactive process yielded a more refined understanding of personnel decision-making at the company. Eventually, it became clear to me what was "really" going on.
>
> Up to a certain level in the hierarchy of positions, there was virtually no relationship between race and promotion. But for a particular level mid-way up the organizational ladder, very few workers were being promoted from within the company when openings arose. Rather, these particular jobs were being filled

JWST445-c01 JWST445-Weisberg May 1, 2014 13:26 Printer Name: Trim: 6.125in × 9.25in

> primarily through outside hires, and almost always by white applicants. Moreover, these external candidates were sometimes less qualified than the internally available workers. We came to call this peculiar dynamic "the bottleneck."
>
> This subtle pattern, once recognized, was supported anecdotally in several ways. The statistical data, coupled with qualitative supporting information, was eventually presented to the defendant company's attorneys. The response to our demonstration of the bottleneck phenomenon was dramatic: a sudden interest in negotiation after many months of intransigence. Within weeks, a settlement of the case was forged.[19]

This unorthodox approach did not rely on any of the traditional statistical methods I had been taught, which made me somewhat uncomfortable. Throughout the subsequent 30 years, I have had a number of similar "out-of-the-box" experiences. Consequently, I have become much more confident that my approach in such cases was based on some kind of logic not encompassed in traditional methods. This book has resulted in part from my desire to understand and articulate what this logic might be.