

CHAPTER 22: SOME PROBLEMS ARE INTRINSICALLY DIFFICULT



The scientific method leads to faster progress on some problems than others. Like children's puzzles and games, scientific problems vary in tractability from easy to difficult. Tic-tac-toe is easy to master. Jig saw puzzles require more patience, but with persistence can be solved by nearly anyone. Rubik's cube is mind-boggling.

Introduction

In the last 75 years we have built bombs that could catastrophically alter our climate for centuries to come, yet we remain surprisingly inept at predicting the weather even a week in advance. Although we have eliminated smallpox, progress on curing cancer has been painfully slow. We can describe with some accuracy the flight of a baseball, but are inept at predicting the outcome of sporting events.

There are good reasons why ignorance persists about the weather, sports contests, and many other phenomena in spite of continual assault by the scientific approach. The very nature of these problems makes them difficult. Of course the rate of progress is in part controlled by factors other than the difficulty of the problem, including the amount of money spent and the number of people working on the problem. But these social factors will not concern us. Here we will consider a few factors that make some scientific questions intrinsically difficult:

1. Research on humans is difficult
2. Effects are difficult to measure under some circumstances
 - 2a. Intrinsically limited replication
 - 2b. Small effects amid lots of variables
 - 2c. Interactions
 - 2d. Time lags

1. Research on Humans is Difficult

There are many problems facing humans which could be ameliorated using the scientific method except that the "ideal" experiments cannot be conducted because they involve humans -- they are unethical, too expensive, or just impractical. Consider how you would react if you discovered that the government or your employer had exposed you to high doses of radiation without your knowledge, or had tested drugs on you without your approval. These kinds of manipulations are routinely performed on non-human organisms, but we do not permit them to be conducted on ourselves, even when such manipulations could be useful in solving an important problem.

Second, some manipulations with humans that are not unethical are nonetheless not feasible. Studies requiring humans to voluntarily change their behavior (for example by adopting a particular diet) pose the obvious problem that the subjects may not comply with the regimen. If the manipulation calls for an extreme change in behavior for a long period, the experiment is probably not feasible.

Experimental studies of human behavior constitute a gray area in terms of ethics. Most of us would likely frown on an experiment that involved teaching children to fear common, harmless objects. And many people would object to being "experimented upon" without being informed. Yet that is essentially what advertisers and many other businesses do when they gather data on the effectiveness of different product promotion techniques. When an advertising agency generates two versions of the same ad and compares the sales they generate, it is performing an experiment on its customers. Some ads are designed to create an aversion against a competitor's product, in the same spirit as the psychologists who taught the little girl to fear white rats. Furthermore, the customers do not know they are involved in the experiment, and quite probably do not know that such experiments are regularly done. Businesses not only attempt to discover our preferences and dislikes, but they attempt to alter our behavior in ways that benefit them --- actually teaching us to enjoy their products and dislike others. The very basis of capitalism, by which some products survive and others fail, is itself an ongoing set of experiments in human behavior.

2. Difficult detection: (2a) Intrinsically limited replication (rare events)

We have all experienced the frustration of a car, stereo, or other complicated machine failing us, only to face the embarrassment of the machine working perfectly when brought in for repair. It is usually easier to fix something that consistently fails than to fix an intermittent problem. A common solution is to simply ignore an intermittent problem until it worsens.

The difficulty of a scientific problem depends heavily on the frequency of the event being studied -- which determines the number of data that are obtained. Models of 'rare events' improve only slowly. Inconsistent or uncertain results increase the number of observations that must be made -- the number of samples that must be taken -- before we can make progress. For example, it does not require too many coin tosses to realize that we are being cheated with a 2-headed coin. But to detect whether a casino's slot machine offered "fair odds" of a win, we might need to pull the lever thousands or hundreds of thousands of times. Thus, when the event we seek is extremely rare, the problem can become physically insurmountable.

There are many kinds of rare events that confront us, many of them undesirable (Table 22.3). Any one of these events is rare enough that we are likely to ignore its possibility, but there are so many rare event possibilities that they pose a threat collectively. And from a social policy perspective, an individually rare event can still mean thousands of cases in a population the size of the U.S.

TABLE 22.3 RARE EVENTS IN OUR LIVES
Adverse reactions to common drugs and vaccines
Side effects of food additives
Transportation accidents
Equipment failures on airplanes and space shuttles
Cardiac arrest under anesthesia
Large liability awards against insurance companies
Floods, tornadoes, lightning, and hurricanes
Leukemia
Winning the lottery

Measuring a rare event may require huge samples.

Childhood leukemia is one of the few cancers that occurs at appreciable levels in children. The disease is fatal unless halted with an extremely radical and difficult treatment. The odds in the U.S. are that about 1 in every 20,000 children will develop leukemia before becoming an adult. This number is a baseline, or average rate. We would obviously like to reduce the number of cases below 1 in 20,000, but we also want to ensure against environmental changes that increase it.

Studies now outdated suggested that the childhood leukemia rate may nearly double due to exposure to intense electromagnetic fields --- the sort of everyday radiation emitted from electric appliances, power lines, and transformers atop telephone poles. Even though a doubling of this rate still means that each individual has an

excellent chance of avoiding leukemia, the doubling would constitute a serious increase in the number of childhood leukemia cases in a country the size of the U.S.

With a rate of 1 in 20,000, we expect only 5 cases in 100,000, or 10 cases if the rate is doubled. Yet, if we indeed observed 5 cases out of 100,000 for one group and 10 out of 100,000 for another group, the difference between 5 and 10 is not large enough to convince us that pure chance isn't responsible for the discrepancy. Even larger numbers of individuals would need to be sampled. Herein lies the problem: a sample of 200,000 children is not adequate for detecting even a doubling of the leukemia rate. When considering that a variety of data must be collected on each child, the enormity and cost of the problem becomes staggering.

The rare event problem can also impact business. Suppose that a product is tested with 1000 subjects and found to be satisfactory and safe. If it is hazardous to 1 out of 10,000 people, then a test of 1,000 people is likely to miss the hazardous effect. Yet when the product is marketed, it will come in contact with possibly millions of people, and its drawbacks will become obvious from the hundreds of people who suffer from it. Liability costs for even a few of those afflicted could easily wipe out all profits. This problem applies to manufacturers of drugs and food additives, obviously enough, but also to manufacturers of fabrics, household chemicals, equipment, toys, and an innumerable list of other items with which physical accidents may occur.

An adequate sample may be impossible

The rapid urbanization of the last century notwithstanding, much of the U.S. is populated by small communities of a few hundred to a few thousand people. Importantly, an environmental hazard may increase the incidence of cancer, birth defect, or miscarriage yet the entire community may be so small that there is no statistical basis for demonstrating an ill effect of the hazard.

Furthermore, a corporation exposing a small village to a toxic chemical, for example, may be virtually immune from legal accountability (provided that people are not killed or hospitalized en masse), because too few cases will ever come to pass. Disputes between small communities and large corporations spraying herbicides have in fact occurred over this very point. Similar debates have arisen over whether the emissions from chemical manufacturers have increased the number of cases of anencephaly (babies born with essentially no brain) in small communities along the Texas-Mexico border.

Even if a suspected hazard such as a toxic waste site or gasoline tank farm occurs in a large city, there is no guarantee that enough people will be affected to produce convincing scientific evidence that the suspected hazard really is bad. If the toxic waste site only increases cancer rates in residents who live within several blocks of the site, then it is likely that only a very small number will contract cancer because of the toxic waste site. Even though the toxic waste site is in a large city, the scientific issues are very similar to those encountered in understanding environmental hazards in small communities.

Related problems: Dispersed events and unique events

There are some obvious generalizations and extensions of rare event problems. One is dispersed effects: a large number of people are affected, but they are not clustered in any obvious way. Dispersion is a common problem in the detection of infectious diseases and is an acknowledged problem in bioterrorism awareness. For any one type of food item, there are relatively few food processing centers in the country. For example, most lettuce used in restaurants and fast-food chains is chopped up in a few sites. Suppose one of those sites was contaminated with an infectious bacterium that caused 400 consumers to get sick: the effect would be a distributed outbreak of the illness, but only 2-4 per major city. If the sickness was nothing out of the ordinary (e.g., diarrhea, with recovery in 3 days), the contamination would go undetected. If all 400 illnesses happened in one city, it might well be detected. Indeed, the clustering of illnesses was critical to the detection of an *E. coli* O157 outbreak in Seattle a few years ago (known as the “Jack-in-the-Box” episode) – there had been a similar outbreak in Nevada years earlier that had gone unnoticed. Likewise, the discovery of hantavirus infections in the U.S. was accidental, and found only because of a geographic cluster of illnesses in the four-corners area.

A second difficulty in applying the scientific method is that some events cannot be replicated. Historical events are the most obvious, and some controversy and angst in our society revolves around past events that weren't sufficiently documented and don't have satisfactory answers (Kennedy assassination, supposed aliens near Roswell). Some types of large-scale events have the same problem – they can only happen once, because the whole population is affected. (The mass polio vaccination with the live Sabin vaccine around 1960 comes to mind, but this problem affects the implementation of many government programs. Likewise, the HIV epidemic is unique for the size of its impact on the world.) In many large-scale events, there will be components that are replicated (e.g., in the HIV epidemic, infections are replicated millions of times) but will also be components that are unique (e.g., world-wide economic impacts of the massive toll).

Overcoming Rarity

If a large sample can't be obtained, there are several alternatives that enable us to side-step the problem posed by rare events. The general solution is to turn to alternative models that facilitate the observation of large numbers of cases.

Models of surrogates:

Although cancer is a common affliction of humans, the development of specific cancers in response to specific factors is rare (e.g., the leukemia risk from increased levels of radiation is not very high). To assess cancer risk, some studies instead look at abnormalities other than cancer, such as chromosome aberrations in blood cells or precancerous growths, and other studies assess mutation rates in bacteria, which can be analyzed by the billions. These cancer surrogates are chosen because they are thought to accurately reflect the likelihood of developing cancer and because they occur at higher frequencies than the cancers themselves. In the same vein, one could use the near-collisions of aircraft to study the factors influencing actual collisions, which are themselves exceedingly rare.

Inflating the rates:

The world is heterogeneous, and when science studies a rare phenomenon, there may be special circumstances in which the phenomenon is common or can be rendered common. To test equipment failure, it is often a simple matter to stress equipment under laboratory conditions to increase its rate of failure, thereby obtaining information about its failure under more normal circumstances. In medicine, rats are often subjected to extremely high doses of substances, to increase the frequency of any ill effects that might be felt by a tiny minority of human consumers. And medical models with inflated rates are not always rats. People who for whatever reason receive higher-than-normal doses of radiation, alcohol, and other drugs are sometimes studied specifically for the purpose of determining risks of lower doses. Airline cockpit simulators can mimic unusual combinations of events, thereby increasing a pilot's ability to survive adverse conditions.

Tracing causal chains:

We have implicitly assumed in this chapter that in order to demonstrate that, say, a toxic waste dump causes cancer in nearby residents, you must establish a correlation, or association, between the waste dump and cancer. This assumption is not completely valid. If we can understand why a rare event is occurring, we may be able to draw reliable conclusions with even small sample sizes. In the beginning of the chapter, we pointed out that it would likely take many thousands of pulls of a slot machine lever to determine the odds of winning. But there is a more direct approach: Simply open the machine up, and look at how the odds have been set (However, we don't recommend trying this on a casino floor.) As we discussed in the chapter on correlation and causation, many problems can be attacked similarly. Examples include scientists identifying the particular genes that an environmental hazard causes to mutate, and secret service agents looking at the details of past presidential assassinations to understand the psychological profile of assassins and the circumstances in which a threat is likely to develop.

2. Difficult detection: (2b) Small effects amid lots of variables

When many variables are present, identifying the effect of one can be a profound challenge. If data are correlational, which they usually are, then the 3rd variable problem will overwhelm our ability to identify a cause of the small effects -- or the third variable problem can be assumed to overwhelm, so we cannot be sure of an observed effect. This is not a sample size problem: the overall effect may well be statistically detectable (e.g., an overall modest increase in a cancer rate). The challenge created by the many variables is *attributing* the effect to the correct variable -- there could be too many to consider. Ultimately, if the right variable is guessed, it may be detectable with appropriate studies. The larger the overall effect of one variable, the easier attribution becomes. But as the overall effect becomes small, the intrinsically limited replication problem confounds attribution.

This realization poses a problem for the detection of environmental hazards. Industry dumps thousands of chemicals into our environment, and we are subjected to hundreds of chemicals in our food, cars, and homes (often at very low levels). If some of these cause harm at low levels, we are unlikely to attribute those harms to the correct cause, even though hundreds of people may be affected annually across the U.S. There is a group of scientists who fear that our environment is polluted with 'endocrine disruptors' that affect behavior and reproduction. There are so many of those pollutants, that even though we have evidence of systematic changes in our reproductive health (e.g., declining sperm counts in much of Europe), we cannot attribute those changes in health to any single or few compounds.

One solution is to set exposure limits based on animal studies and hope that those animal models are suitable for humans. But we can only do that when we are lucky enough to know what effects to assay for in the animals.

2. Difficult detection: (2c) Interactions

The arch-villain Joker in the 1989 movie *Batman* devised a plan to poison the citizens of Gotham City. Rather than simply put a single poison into one product, Joker used a poison which required the combined effects of multiple ingredients. No single product was by itself toxic. Batman discovered the formula to Joker's toxic scheme, and the public was advised accordingly: "avoid the following combinations: deodorants with baby powder, hair spray, and lipstick."

The sinister dimension to Joker's plan is readily apparent to us, because we can all appreciate how difficult it would have been to discover that a combination of products was deadly. Several years ago, when a real villain was lacing bottles of Tylenol with cyanide in the U.S., the problem was simple enough to trace, because a single product was the source of the poison. But imagine the difficulty of tracing the problem if a combination of three products was toxic, and that each of these products by itself was innocuous. Likely, many people would die before anyone determined that a particular combination of products was fatal.

The phenomenon that underlies this example is an interaction among many factors: we cannot discern the whole from a sum of the parts. This is a problem because science typically functions in the same way that we construct a jigsaw puzzle. That is, although the problem involves many pieces and is overwhelmingly complex, progress is made one piece at a time, building on previous successes. Most improved models are relatively minor modifications of their predecessors. But suppose that the puzzle consisted of many pieces, each of which could fit with several other pieces, yet only one combination enabled all pieces to fit together. In this case, we could make many starts, only to find that they invariably led nowhere.

Interactions are ubiquitous in our lives at one level or another (Table 22.4). Many events from the non-scientific and non-industrial side of our lives involve interactions at one level or another: a joke without its punchline is not half funny.

Interactions of common experience:

EXAMPLE	INGREDIENTS	RESULT	BASIS OF INTERACTION
Flash powder	mixture of magnesium powder and potassium nitrate, plus energy	explosion	neither ingredient alone generates a reaction
Lethal gas	mixing household bleach and ammonia cleansers	chlorine gas	each cleanser is safe when used alone
Atomic bomb	critical mass of plutonium or uranium	chain reaction of atomic disintegrations	Half of a critical mass does not release half the energy of a critical mass
Cooking recipes	various spices and food items	prepared meals	eating the prepared meal has a greater appeal than eating each ingredient separately
Drug complications	different drugs designed for different purposes	drug-induced death or illness	when used separately, drugs produce positive health effects

Why Interactions are a Problem:

The problem posed by interactions is simply that the effect of one factor depends on which other factors are present. Thus factor X may behave one way in the absence of Y but behave quite differently when it is present. It is often difficult for science to anticipate the effects of interactions because science often works with single effects. For someone cleaning around the house, it seems perfectly logical to mix different cleansers to reduce the number of times a surface needs to be cleaned. Indeed, many household products and over-the-counter drugs actively advertise a multiplicity of components --- the all-in-one principle. But occasionally, the combination of two or more safe ingredients holds a surprise, such as deadly chlorine gas.

The extension of this principle from ordinary problems to scientific ones is simple, as is a realization of the difficulty it poses. Time and again, science fails to give us advance warning of dangerous interactions, and people are injured or die before we are able to arrive at an adequate model to explain the phenomenon. For example, the deadly combination afforded by sedatives and alcohol was discovered by trial and error. The death of a few celebrities in the 1950's and 1960's made this interaction well known. The history of new discoveries in applied chemistry is replete with examples of botched protocols that led to completely unexpected results.

Doing science amid interactions

To a large extent, science is simply saddled with this problem. The Noble Prize awarded for the discovery of bizarre combinations of metals that have superconducting properties reflects the difficulty of such problems. Two kinds of approaches help overcome this general problem, but neither is a completely satisfactory solution: models of mechanisms (or, equivalently, causal chains), and models of single components.

Tracing causal chains: This is exactly the same principle we have already discussed in reference to rare events and in the chapters on causation and correlation. The atomic bomb, for example, was not discovered by accident, rather it was predicted from knowledge about radioactive disintegration products and energies for specific Uranium and Plutonium isotopes. In this case, an explosive chain reaction results when the sum of many individual fissions reaches a critical threshold that was predicted from first principles.

Models of single components: In many other cases, complex interactions can be anticipated by first looking at one or more of the ingredients separately. The driving force in gunpowder and flash powder is an oxidizing chemical. Although explosions will result from specific combinations of ingredients, the oxidizing agent is capable of sustaining combustion with a much wider range of ingredients, so it becomes a simple matter to explore different combinations to optimize the rate of combustion.

2. Difficult detection: (2d) Time Lags

For some problems, the data needed to test models can only be gathered slowly. Consider the procedure we use in adjusting the temperature of a shower and the dial of a radio. Both involve a simple application of the scientific method: We start at some initial setting, evaluate the setting, readjust the setting, evaluate the new setting, and repeat the process until the desired goal is achieved.

Because we obtain the needed data faster, progress in setting the radio dial occurs more quickly than progress on adjusting the shower temperature. The difference is due to time lags. Shower handles are typically several feet removed from the shower head, and it may take up to a minute before an adjustment at the handle translates into a change in the temperature of water on our skin. By contrast, turning the radio dial results in a nearly instantaneous change in the radio's output. Because gathering data about shower temperature involves a longer time lag than gathering data about a radio frequency, it takes longer to adjust a shower.

Time lags abound in our world. Some spectacular and renowned cases involve the orbits of heavenly bodies. Those of us who did not observe Haley's comet this century are not likely to have another chance, because the time lag is 77 years. No one knew to expect the Hale-Bopp comet in 1997 because it had not been observed for nearly 2000 years. We all fear overexposure to radiation because it increases the risk of cancer. However, the onset of cancer typically follows exposure to radiation by many years (5 years for leukemia and 20 years for other cancers). In economics, the Federal Reserve Bank Board (the "Fed") attempts to influence the U.S. economy by adjusting interest rates; the effect of a change in interest rates takes months to be translated into an impact on the economy. And couples wishing to become parents must wait at least 9 months if they do the job by the classical method, and often much longer if they wish to adopt.

Not only do time lags increase the cycle time of the scientific method, but they also are sometimes so long as to escape detection altogether. The first point is evident from our shower analogy. The longer the delay between faucet adjustment and temperature change, the longer it takes to find an acceptable setting. It may take the same number of adjustments to adjust both the shower and radio dial, but it simply takes more total time when the time lag is long.

The second point --- the possible failure to recognize a time lag --- is more subtle and more sinister. If a time lag is extraordinarily long, we may be unable even to determine that it is present. For example, if the cancer rates we experience today are determined by chemical exposures to our grandfathers when they were 10, it would be nearly impossible to discover the effect.

Examples of time lags:

LONG TIME LAGS	
Greenhouse warming	There is a lag of decades between industrial activities that increase atmospheric carbon dioxide, and any resulting change in climate.
Weight loss diets	Only after weeks or months on a diet do you achieve significant weight loss.
Diet and heart disease	Many people eat a high fat diet for decades before suffering a heart attack.
SHORT TIME LAGS	
Computers	You can obtain data from many computers almost instantaneously; that is, with essentially no time lag. The computer responds almost instantly when you type in a command.
Steering a ship	There is an obvious time lag between turning the steering wheel of an ocean liner and the actual change in direction of the ship. Pilots also face time lags in landings and take-offs because the momentum of the plane does not change quickly in response to the controls.

Social problems created by long lags

Consider the difficulties posed by time lags in drug manufacturing. If you are testing a new drug, how long should the participants in the study be followed before you can conclude that the drug is safe? Is one year sufficient? Five years? It strains the limits of credulity to imagine how our drug-based health treatment would be affected if trials needed to be followed even for 10 years before a product could be approved. New companies would have to find sources of revenue for at least 10 years before they could begin marketing their first product. The shelves are full of drugs that would not be available under such rules, and of course, improvements in those drugs would be even longer in coming.

The U.S. drug marketplace has witnessed such a problem. From the late 1940's until 1970, the drug DES (diethylstilbestrol) was administered to many women in early pregnancy to suppress miscarriage. It was only later discovered that its use results in an increased cancer rate in their offspring --- 20 to 40 years after exposure to the drug. If the drug had caused cancer immediately (e.g., in the pregnant women or their newborn), then it would have been contraindicated long before 1970, and many fewer people would have developed DES-caused cancer. Similarly, a 1993 trial of a hepatitis B drug killed 5 of the 15 volunteers. Part of the reason so many died was that the lethal effect of the drug was somewhat delayed --- a time lag had not been anticipated.

There is no definitive solution to this dilemma. Countless drugs that we are taking now could be having a delayed effect. Some compromise must be struck between the conflicting goals of adequate testing to ensure safety and maximizing the number of effective drugs available. Actions that the government takes to increase the safety of drugs available (by requiring tests run for longer periods of time), will often prevent or delay some safe and effective drugs from coming to market, because the drugs can't be sold while experiments determining their safety are undertaken. Moreover, regardless of the number of tests undertaken, there is no way to be absolutely certain that a given drug is safe.

Another set of problems arises because long time lags make it difficult to determine who is to blame for poor performance. Is the current recession a consequence of the policies of the current president, or a predecessor? Are the company's earnings the first year after a new CEO is hired a consequence of his/her actions, or the actions of a predecessor? In both these cases, uncertainty about the duration of a time lag obscures the answer.

Time lags are common problems, and scientists have discovered ways to lessen their impact. A common approach is to study alternative models that incorporate a shorter time lag (see Table 22.2). The utility of an alternative model with short time lags depends on its similarity to the main model in question. Viruses (bacteriophages) and fruitflies yielded major insights to the study of human genetics because they have a vastly shorter generation time than do humans.

Perhaps the biggest difficulty is posed by unexpected time lags, as with DES-caused cancer. If you aren't expecting a time lag, there's not much you can do about it until you stumble on it.

Models that reduce time lags:

Genetics of viruses. Their short life cycle led to rapid understanding of genetic principles that apply to nearly all life.
Rodents are used in cancer research because their short life, relative to humans, enables testing for otherwise long-term effects
Flow charts enable coordination of different dimensions of complicated construction and other social projects, so that excessive delays are avoided.
Political polls provide politicians with rapid feedback about public perception of their performance. The politician can change their positions and their behavior in response to the poll, so they don't have to wait for an election to discover their popularity.
Sneak previews enable marketing agencies to anticipate public reaction to a product before it is made widely available. Changes in packaging and marketing strategy can occur much more quickly if they only affect a small market. Once the bugs are ironed out in a small market, the resulting marketing strategy is then used nationally.
Early reviews and advance advertising. A company may speed public awareness of a product prior to or coincident with its availability in the marketplace.