The Problem of Thinking Too Much

Persi Diaconis, Mary V. Sunseri Professor of Mathematics and Statistics, Stanford University

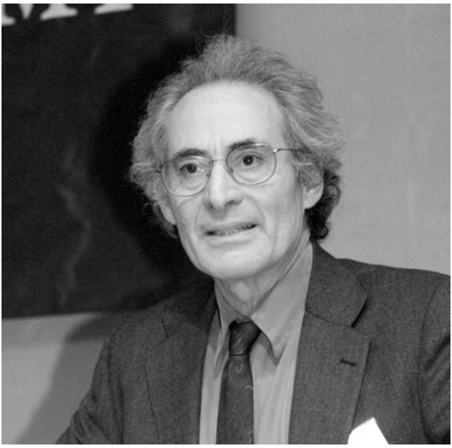
Introduction: ***Barry C. Mazur,*** Gerhard Gade University Professor, Harvard University

Barry C. Mazur

Persi Diaconis is a pal of mine. He's also someone who, by his work and interests, demonstrates the unity of intellectual life—that you can have the broadest range and still engage in the deepest projects. Persi is a leading researcher in statistics, probability theory, and Bayesian inference. He's done wonderful work in pure math as well, most notably in group representation theory. He has the gift of being able to ask the simplest of questions. Those are the questions that educate you about a subject just because they're asked. And Persi's research is always illuminated by a story, as he calls it—that is, a thread that ties the pure intellectual question to a wider world.

Persi's world is indeed wide. It includes discovering beautiful connections among group-representation theory, algebraic geometry, card-shuffling procedures, and Monte Carlo algorithms; studying random-number generators, both theoretical and very practical; analyzing and interpreting real-world applications of statistics, as in voting procedures; critiquing misrepresentations of science and mathematics, in particular the protocols of experiments

This presentation was given at the 1865th Stated Meeting, held at the House of the Academy on December 11, 2002.



Barry C. Mazur (Harvard University)

regarding extrasensory perception; writing on the general concept of coincidence; and working on historical treatises about probability and magic. As is well known, Persi is also a magician, credited with, as Martin Gardner once wrote, "inventing and performing some of the best magic tricks ever."

As for honors, there's a long list. He was, for example, one of the earliest recipients of the MacArthur Fellowship. He's a member of the National Academy of Sciences and was president of the Institute of Mathematical Statistics. On top of all this, Persi has an exemplary gift for explaining things, so I should let him do just that.

Persi Diaconis

Consider the predicament of a centipede who starts thinking about which leg to move and winds up going nowhere. It is a familiar problem: Any action we take has so many unforeseen consequences, how can we possibly choose?

Here is a less grand example: I don't like moving the knives, forks, and spoons from dishwasher to drawer. There seems no sensible way to proceed. I frequently catch myself staring at the configuration, hoping for insight. Should I take the tallest things first, or just grab a handful and sort them at the drawer? Perhaps I should stop thinking and do what comes naturally. Before giving in to "thinking too little," I recall a friend's suggestion: you can speed things up by sorting the silverware as you put it into the dishwasher. On reflection, though, this might lead to nested spoons not getting clean. And so it goes.

I'm not brazen enough to attempt a careful definition of "thinking" in the face of a reasonably well-posed problem. I would certainly include mental computation (e.g., running scenarios, doing back-of-the-envelope calculations), gathering information (e.g., searching memory or the Web, calling friends), searching for parallels (e.g., recognizing that the problem seems roughly like another problem one knows how to solve, or thinking of an easier special case), and, finally, trying to maneuver one's mind into places where one is in tune with the problem and can have a leap of insight.

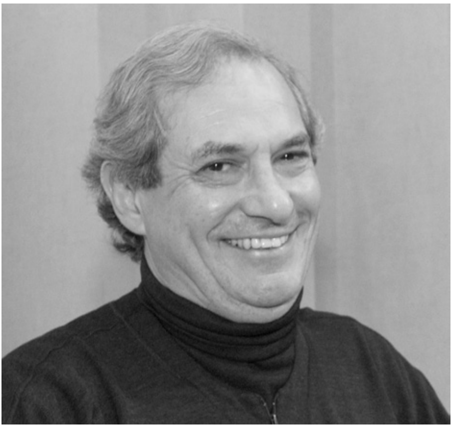
The problem is this: We can spend endless time thinking and wind up doing nothing—or, worse, getting involved in the minutiae of a partially baked idea and believing that pursuing it is the same as making progress on the original problem.

The study of what to do given limited resources has many tendrils. I will review work in economics, psychology, search theory, computer science, and my own field, mathematical statistics. These aren't of much help, but at the end I will note a few rules of thumb that seem useful.

An Example

One of the most satisfying parts of the subjective approach to statistics is Bruno de Finetti's solution of common inferential problems through exchangeability. Some of us think de Finetti has solved Hume's Problem: When is it reasonable to think that the future will be like the past? I want to present the simplest example and show how thinking too much can make a mess of something beautiful.

Consider observing repeated flips of a coin. The outcomes will be called heads (H) and tails (T). In



Speaker Persi Diaconis (Stanford University)

a subjective treatment of such problems, one attempts to quantify prior knowledge into a probability distribution for the outcomes. For example, your best guess that the next three tosses yield HHT is the number P(HHT). In many situations, the order of the outcomes is judged irrelevant. Then P(HHT) equals P(HTH) equals P(THH). Such probability assignments are called "exchangeable."

Bruno de Finetti proved that an exchangeable probability assignment for a long series of outcomes can be represented as a mixture of coin tossing: For any sequence a, b, . . . , z of potential outcomes,

$$\mathrm{P}(\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}) = \int_0^1 p^A (1-p)^B \mu(dp)$$

with A the number of heads and B the number of tails among a, b, . . . , z. The right side of this formula has been used since Thomas Bayes (1764) and Pierre-Simon Laplace (1774) introduced Bayesian statistics. Modern Bayesians call $p^A\ (1-p)^B$ the likelihood and μ the a priori probability. Subjectivists such as de Finetti, Ramsey, and Savage (as well as Diaconis) prefer not to speak about nonobservable things such as "*p,* the long-term frequency of heads." They are willing to assign probabilities to potentially observable things such as "one head in the next ten tosses." As de Finetti's Theorem shows, in the presence of exchangeability, the two formulations are equivalent.

The mathematical development goes further. After observing A heads and B tails, predictions about future trials have the same type of representation, with the prior μ replaced by a posterior distribution given by Bayes's formula. Laplace and many followers proved that as the number of trials increases, the posterior distribution becomes tightly focused on the observed proportion of heads—that is, $A/(A+B)$ if A heads and B tails are observed. Predictions of the future, then, essentially use this frequency; the prior μ is washed away. Of course, with a small number of trials, the prior μ can matter. If the prior μ is tightly focused, the number of trials required to wash it away may be very large. The mathematics makes perfect sense of this; fifty trials are often enough. The whole package gives a natural, elegant account of proper inference. I will stick to flipping coins, but all of this works for any inferential task, from factory inspection of defective parts to evaluation of a novel medical procedure.

Enter Physics

Our analysis of coin tossing thus far has made no contact with the physical act of tossing a coin. We now put in a bit of physics and stir; I promise, a mess will emerge. When a coin is flipped and leaves the hand, it has a definite velocity in the upward direction and a rate of spin (revolutions per second). If we know these parameters, Newton's Laws allow us to calculate how long the coin will take before returning to its starting height and, thus, how many times it will turn over. If the coin is caught without bouncing, we can predict whether it will land heads or tails.

A neat analysis by Joe Keller appeared in a 1986 issue of *American Mathematical Monthly.* The sketch in Figure 1 shows the velocity/spin plane. A flip of the coin is represented by a dot on the figure, corresponding to the velocity and rate of spin. For a dot far to the right and close to the axis, the velocity is high, but spin is low. The coin goes up

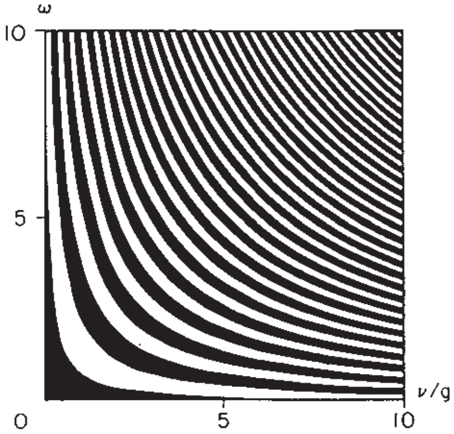


Figure 1. Partition of phase space induced by heads and tails

like a pizza and doesn't turn over at all. All the points below the curve correspond to flips in which the coin doesn't turn over. The adjacent region contains points at which the coin turns over exactly once. It is bounded by a similar curve. Beyond this, the coin turns over exactly twice, and so on.

As the figure shows, moving away from the origin, the curves get closer together. Thus, for vigorous flips, small changes in the initial conditions make for the difference between heads and tails.

The question arises: When normal people flip real coins, where are we on this picture? I became fascinated by this problem and have carried out a series of experiments. It is not hard to determine typical velocity. Get a friend with a stopwatch, practice a bit, and time how long the coin takes in its rise and fall. A typical one-foot toss takes about half a second (this corresponds to an upward velocity of about 5½ miles per hour). Determining rate of spin is trickier. I got a tunable strobe, painted the coin black on one side and white on the other, and tuned the strobe until the coin "froze," showing only white. All of this took many hours. The coin never perfectly froze, and there was variation from flip to flip. In the course of experimenting, I had a good idea. I tied a strand of dental floss about three feet long to the coin. This was flattened, the coin flipped, the flip timed, and then we unwrapped the floss to see how often the coin had turned over. On the basis of these experiments, we determined that a typical coin turns at a rate of 35 to 40 revolutions per second (rps). A flip lasts half a second, so a flipped coin rotates between 17 and 20 times.

There is not very much variability in coin flips, and practiced magicians (including myself) can control them pretty precisely. My colleagues at the Harvard Physics Department built me a perfect coin flipper that comes up heads every time. Most human flippers do not have this kind of control and are in the range of 5½ mph and 35 to 40 rps. Where is this on Figure 1? In the units of Figure 1, the velocity is about ⅕—very close to the zero. However, the spin coordinate is about 40—way off the graph. Thus, the picture says nothing about real flips. However, the math behind the picture determines how close the regions are in the appropriate zone. Using this and the observed spread of the measured data allows us to conclude that coin tossing is fair to two decimals but not to three. That is, typical flips show biases such as .495 or .503.

Blending Subjective Probability and Physics

Our refined analysis can be blended into the probability specification. Now, instead of observing heads and tails at each flip, we observe velocity/spin pairs. If these are judged exchangeable, a version of de Finetti's Theorem applies to show that any coherent probability assignment must be a mixture of independent and identically distributed assignments:

$$P((v, w) \text{ in } A, \ldots, (v', w') \text{ in } B) = \int F(A) \ldots F(B)\, \mu(dF)$$

The meaning of these symbols is slightly frightening, even to a mathematical grownup. On the right, F is a probability distribution on the velocity/spin plane. Thus μ is a probability on the space of all probabilities. Here, de Finetti's Theorem tells us that thinking about successive flips is the same as thinking about measures for measures. There is a set of tools for doing this, but at the present state of development it is a difficult task. It is even dangerous. The space of all probability measures is infinite-dimensional. Our finite-dimensional intuitions break down, and hardened professionals have suggested prior distributions with the following property: as more and more data come in, we become surer and surer of the wrong answer.

This occurs in the age-old problem of estimating the size of an object based on a series of repeated measurements. Classically, everyone uses the average. This is based on assuming that the measurement errors follow the bell-shaped curve. Owning up to not knowing the distribution of the errors, some statisticians put a prior distribution on this unknown distribution. The corresponding posterior distribution can become more and more tightly peaked about the wrong answer as more and more data come in. A survey of these problems and available remedies can be found in my joint work with David Freedman in the *Annals of Statistics* (1986).

What's the Point?

This has been a lengthy example aimed at making the following point. Starting with the simple problem of predicting binary outcomes and then thinking about the underlying physics and dynamics, we

Mitchell Rabkin (Beth Israel Deaconess Medical Center), Ruth Rabkin, Leon Eisenberg (Harvard Medical School), and Carola Eisenberg

Robert Alberty and Stephen Crandall (both, MIT) with George Hatsopoulos (Pharos L.L.C., Waltham, MA)

were led from de Finetti's original, satisfactory solution to talking close to nonsense. The analysis led to introspection about opinions on which we have small hold and to a focus on technical issues far from the original problem. I hope the details of the example do not obscure what I regard as its nearly universal quality. In every area of academic and more practical study, we can find simple examples that on introspection grow into unspeakable "creatures." The technical details take over, and practitioners are fooled into thinking they are doing serious work. Contact with the original problem is lost.

I am really troubled by the coin-tossing example. It shouldn't be that thinking carefully about a problem and adding carefully collected outside data, Newtonian mechanics, and some detailed calculations should make a mess of things.

Thinking About Thinking Too Much

The problem of thinking too much has a prominent place in the age-old debate between theory and practice. Galen's second-century attempts to balance between rationalist and empiricist physicians ring true today. In his *Three Treatises on the Nature of Science* (trans. R. Walzer and M. Frode), Galen noted that an opponent of the new theories claimed "there was a simple way in which mankind actually had made enormous progress in medicine. Over the ages men had learned from dire experience, by trial and error, what was conducive and what was detrimental to health. Not only did he claim that one should not abandon this simple method in favor of fanciful philosophical theories, which do not lead anywhere; he also argued that good doctors in practice relied on this experience anyway, since their theories were too vague and too general to guide their practice." In my own field of statistics, the rationalists are called decision theorists and the empiricists are called exploratory data analysts. The modern debaters make many of the same rhetorical moves that Galen chronicled.

Economists use Herbert Simon's ideas of "satisficing" and "bounded rationality," along with more theoretical tools associated with John Harsanyi's "value of information." Psychologists such as Daniel Kahneman and Amos Tversky accept the value of the heuristics that we use when we abandon calculation and go with our gut. They have created theories of framing and support that allow adjustment for the inevitable biases. These give a framework for balancing the decision to keep thinking versus getting on with deciding.

Computer science explicitly recognizes the limits of thinking through ideas like complexity theory. For some tasks, computationally feasible algorithms can be proved to do reasonably well. Here is a simple example. Suppose you want to pack two suitcases with objects of weight a, b, . . . , z. You

Daniel Bell (Harvard University) and Martin Cohn (MIT) want to pack them as close to evenly as you can. It can be shown that this is a virtually impossible problem. Despite fifty years of effort, we don't know how to find the best method of packing, save for trying all of the exponentially many possibilities. Any progress would give solution to thousands of other intractable problems. Most of us conclude that the optimal solution is impossible to find.

Undeterred, my friend Ron Graham proposed the following: sort the objects from heaviest to lightest (this is quick to do). Then fill the two suitcases by beginning with the heaviest item, and each time placing the next thing into the lighter suitcase. Here is an example with five things of weight 3, 3, 2, 2, 2. The algorithm builds up two groups as follows:

$$3,\quad 3/3,\quad \begin{matrix}2\\3\end{matrix}/3,\quad \begin{matrix}2\\3\end{matrix}/\begin{matrix}2\\3\end{matrix},\quad \begin{matrix}2\\2\\3\end{matrix}/\begin{matrix}2\\3\end{matrix}$$

This misses the perfect solution, which puts 3, 3 in one pile and 2, 2, 2 in the other. One measure of the goodness of a proposed solution is the ratio of the size of the larger pile to the size of the larger pile in the optimal solution. This is 7/6 in the example. Graham proved that in any problem, no matter what the size of the numbers, this "greedy" heuristic always does at worst 7/6 compared to the optimal. We would be lucky to do as well in more realistic problems.

An agglomeration of economics, psychology, decision theory, and a bit of complexity theory is the current dominant paradigm. It advises roughly quantifying our uncertainty, costs, and benefits (utility) and then choosing the course that maximizes expected utility per unit of time. A lively account can be found in I. J. Good's book *Good Thinking* (don't miss his essay on "How Rational Should a Manager Be?").

To be honest, the academic discussion doesn't shed much light on the practical problem. Here's an illustration: Some years ago I was trying to decide

Helen Pounds, William F. Pounds (MIT), and Paul Doty (Harvard University)

whether or not to move to Harvard from Stanford. I had bored my friends silly with endless discussion. Finally, one of them said, "You're one of our leading decision theorists. Maybe you should make a list of the costs and benefits and try to roughly calculate your expected utility." Without thinking, I blurted out, "Come on, Sandy, this is serious."

Some Rules of Thumb

One of the most useful things to come out of my study is a collection of the rules of thumb my friends use in their decision making. For example, one of my Ph.D. advisers, Fred Mosteller, told me, "Other things being equal, finish the job that is nearest done." A famous physicist offered this advice: "Don't waste time on obscure fine points that rarely occur." I've been told that Albert Einstein displayed the following aphorism in his office: "Things that are difficult to do are being done from the wrong centers and are not worth doing." Decision theorist I. J. Good writes, "The older we become, the more important it is to use what we know rather than learn more." Galen offered this: "If a lot of smart people have thought about a problem [e.g., God's existence, life on other planets] and disagree, then it can't be decided."

There are many ways we avoid thinking. I've often been offered the algorithm "Ask your wife to decide" (but never "Ask your husband"). One of my most endearing memories of the great psychologist of decision making under uncertainty, Amos

Tversky, recalls his way of ordering in restaurants: "Barbara? What do I want?"

Clearly, we have a wealth of experience, gathered over millennia, coded into our gut responses. Surely, we all hope to call on this. A rule of thumb in this direction is "Trust your gut reaction when dealing with natural tasks such as raising children."

It's a fascinating insight into the problem of thinking too much that these rules of thumb seem more useful than the conclusions drawn from more theoretical attacks.

In retrospect, I think I should have followed my friend's advice and made a list of costs and benefits—if only so that I could tap into what I was really after, along the lines of the following "grook" by Piet Hein:

A Psychological Tip

Whenever you're called on to make up your mind,
 and you're hampered by not having any,
the best way to solve the dilemma, you'll find,
 is simply by spinning a penny.
No—not so that chance shall decide the affair
 while you're passively standing there moping;
but the moment the penny is up in the air,
 you suddenly know what you're hoping.
