Mathematical

By IVARS PETERSON

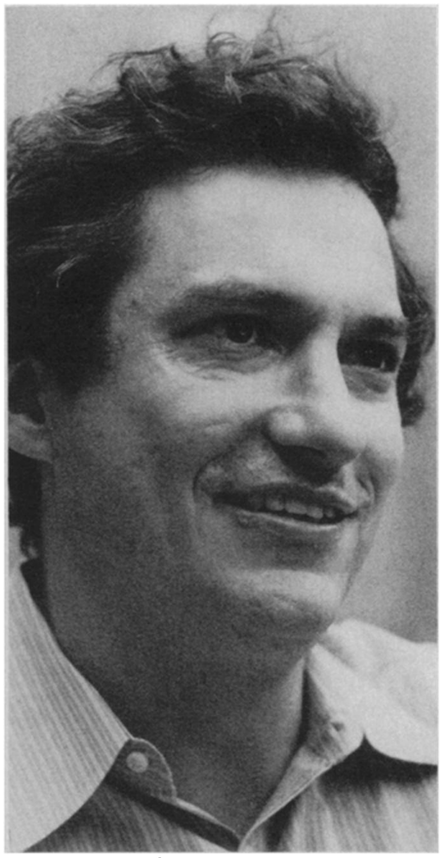
There's magic in the way Persi Diaconis shuffles playing cards. Smoothly, skillfully, he divides a deck of cards into two halves, then riffles the stacks, rapidly running the cards past his thumbs so that the two stacks interleave. He is one of perhaps two dozen people who can do a perfect shuffle, not only once but eight times, which brings a 52-card deck back into its original order. But Diaconis has taken card shuffling beyond mechanical skill and magicians' tricks. As a mathematician at Stanford University in California, he has prestidigitated a lifelong fascination with magic and a strong interest in the mathematics of gambling into a rich source of mathematical ideas.

"It's harder to do carefully than you think," says Diaconis, referring not only to card shuffling but also to its underlying mathematics. It raises age-old questions about the nature of randomness and the precise mathematical ways used to describe seemingly unpredictable events.

"Really understanding how to generate randomness is the business of the statistician," says Diaconis. "There are lots of activities in life in which you are using a little bit of randomness, and you want to know how it spreads." It comes up in lotteries and public-opinion polls, in mathematical simulations of chemical reactions and the generation of fractal patterns (SN: 1/21/84, p. 42). The spread of randomness through a repeatedly shuffled deck of cards is a worthy image for the study of all these activities, he feels.

Diaconis begins with a simple question: How many times must you riffle shuffle a deck of cards to ensure that the cards are randomly arranged? The answer isn't obvious, and for years magicians have taken advantage of this fact.

An old card trick, which depends on the two participants being in different locations, illustrates the situation. To an unsuspecting victim, the perpetrator sends a new ordered deck of playing cards and a set of precise instructions. The instructions tell the victim to cut the deck and shuffle the cards three times, using a riffle shuffle, which by nature is inexact, each time. Then the victim takes the top card, notes its identity, returns the card to somewhere in the middle of the stack, and sends back the deck. Without any difficulty, the trickster finds what was once the top card — usually to his victim's astonishment.



Stanford Univ.

Persi Diaconis

"It seems like an amazing trick," says Diaconis, but it depends on the fact that three riffle shuffles are too few to mix a previously ordered deck of cards into an arrangement that no longer shows any patterns. Cutting and shuffling an ordered deck of cards once leaves the deck with two interleaved chains, each chain having cards in the same relative order as they started. Three shuffles produce eight such chains. In this particular trick, all one needs to do to find the designated card is to deal out the cards as in a game of solitaire, putting them in chains as they come up. Eight piles will form, with one leftover card that doesn't fit any of the sequences. Presto!

"So, you can see that three shuffles are not enough to randomize a deck of cards," Diaconis says. It takes at least seven ordinary riffle shuffles before any trace of a pattern disappears. The special case of a perfect shuffle, on the other hand, carefully maintains the order of the cards during shuffling so that eight shuffles bring a deck back into its original order. "Actually, no finite number of shuffles will ever make anything exactly random," Diaconis adds, "but you soon get close enough for all practical purposes."

The mathematical proof that seven shuffles is enough to randomize a deck adequately is surprisingly recent. Using earlier mathematical ideas, including a way of representing the shuffling process mathematically, Diaconis came up with the answer and a technique that can be applied to many other situations in which rearranging something into a random order is important. For example, it's possible to calculate how many twists of Rubik's cube are needed to mix up the cube's colored squares completely.

The mathematical representation for card shuffling that Diaconis used doesn't duplicate exactly how an expert card handler shuffles a deck. But in Las Vegas, Diaconis found support for his theoretical result. By law, dealers are required to perform five riffle shuffles and two other types of shuffles before a deck is ready for play. It would take an awful lot of work to take advantage of any patterns left after that shuffling process, says Diaconis.

In trying to solve shuffling problems, Diaconis has pioneered the application to statistics of a branch of mathematics called group theory. Originally invented more than 100 years ago to solve problems in number theory and with no practical applications in mind, group theory now provides insights into such diverse pursuits as crystallography, particle physics and, because of the work of Diaconis, statistics.

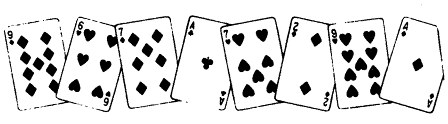
"The nice thing in working on this stuff is that the new tools you develop turn out

Shuffling

The mathematics of card shuffling deals out new methods for handling statistical data and sorting files

to be useful for lots of other things," Diaconis says. At first glance, making sense of ranked data seems unrelated to shuffling cards. Suppose 500 people taste five different types of cookies and rank them from one to five according to how much they like each one. A statistician ends up having to search for trends and patterns in a confusing collection of 500 rankings.

Diaconis says, "Each ranking is, in fact, a rearrangement of five things, which is like a possible shuffle of a deck of five cards." Just as five different cards can take on 120 different arrangements, rankings of five items can appear in 120 different orders. The Diaconis technique involves assigning a number to each of the 120 possible rankings, then listing the number of people who choose each ranking. For example, 26 people may have put cookies A, B, C, D and E in the order 4, 1, 3, 5 and 2, and so on. Group theory provides a way of expressing all this information as the sum of numerical terms, just as the color in a beam of light can be resolved into a weighted sum of the basic colors in the spectrum.



The remarkable thing, Diaconis says, is that the numerical terms seem to correspond to intuitive explanations that make sense. The first term gives an idea of the overall popularity of individual items, while the second term may suggest a "pair effect." That is, after adjusting for the popularity of a single type of cookie, the two chocolate chip cookies do better than the two oatmeal cookies. "Before, people didn't know how to adjust for the popularity of individual cookies," Diaconis says. "The group theorists actually figured out a way to disentangle the effect of a single cookie from the effect of pairs and the effect of triples and so forth."

With this new technique, statisticians can make sense out of all kinds of complex ranked data. Diaconis says, "It gives you ways of thinking about the data that seem useful in applied situations." In an election in which voters must rank the candidates in order of preference, the technique not only provides popularity ratings but may also separate out information about the effect of political party or ethnic background on the voting.

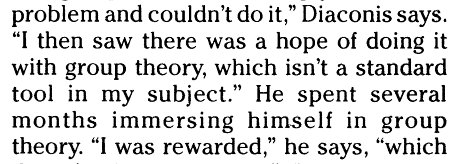
However, the mathematical results tell you only that certain ranked items — whether cookies or candidates — seem to clump together. A statistician must ask, "Can you see a way in which these particular 'clumped' items are similar?" Diaconis says, "The business of statistics is making relevant comparisons. This method points you toward possible significant factors, and your interpretation 'in English' has to fit the individual problem."

"There are other ways of looking at ranked data, but this gives you a way to do what people always wanted to do with these data but didn't know how to do," Diaconis notes. Already, food scientists are among those showing interest in this new method of statistical analysis.

Librarians are becoming interested in a slightly different outgrowth of the mathematics of card shuffling. In this case, it's a mathematical model for how quickly an ordered set of books gradually becomes disordered as people remove books from a shelf and then replace them in the wrong spot. "How long do you have to wait until it becomes chaotic?" Diaconis asks. "Is it a week, a month, a year?" The combination of group theory and statistics can answer these questions, says Diaconis.

The answers matter to library science specialists who have developed ways of sorting books that depend on how mixed up the books are. A stack of books left for only a week may not have a random order unless mixing was very rapid. "It might be," Diaconis says, "that the sorting algorithm that was designed for use on the random permutation wouldn't take advantage of the patterns that remain because things were only mixed a little." Diaconis' results also have potential applications in many other fields where sorting and mixing are important.

Diaconis likes working on practical problems because they force him to learn new things in order to solve a particular problem. "I can't just sit down and learn stuff. I try, and it goes in one ear, and I forget it," he says. "But if I need to learn it to do something, I can learn anything and really learn to use it well."



It was the intractability of a particular shuffling problem that originally forced Diaconis to learn group theory. "The best probabilists on the West Coast had thought quite hard and deeply about that problem and couldn't do it," Diaconis says. "I then saw there was a hope of doing it with group theory, which isn't a standard tool in my subject." He spent several months immersing himself in group theory. "I was rewarded," he says, "which doesn't always happen." Looking back, Diaconis is delighted that he had found another use for something that was originally invented because it was mathematically beautiful. □