

March 30, 2008
Op-Ed Contributors

A Journey to Baseball's Alternate Universe

By SAMUEL ARBESMAN and STEVEN STROGATZ

Ithaca, N.Y.

WITH the baseball season under way and the memory of scandal in the sport so fresh, many fans yearn for an earlier era, a time when mythology mingled with baseball. The sport's most mythic achievement is Joe DiMaggio's 56-game hitting streak, a feat that has never come even close to being matched. Fans and scientists alike, including Edward M. Purcell, a Nobel laureate in physics, and Stephen Jay Gould, the evolutionary biologist, have described the streak as well-nigh impossible.

In a fit of scientific skepticism, we decided to calculate how unlikely Joltin' Joe's achievement really was. Using a comprehensive collection of baseball statistics from 1871 to 2005, we simulated the entire history of baseball 10,000 times in a computer. In essence, we programmed the computer to construct an enormous set of parallel baseball universes, all with the same players but subject to the vagaries of chance in each one.

Here's how it works. Think of baseball players' performances at bat as being like coin tosses. Hitting streaks are like runs of many heads in a row. Suppose a hypothetical player named Joe Coin had a 50-50 chance of getting at least one hit per game, and suppose that he played 154 games during the 1941 season. We could learn something about Coin's chances of having a 56-game hitting streak in 1941 by flipping a real coin 154 times, recording the series of heads and tails, and observing what his longest streak of heads happened to be.

Our simulations did something very much like this, except instead of a coin, we used random numbers generated by a computer. Also, instead of assuming that a player has a 50 percent chance of hitting successfully in each game, we used baseball statistics to calculate each player's odds, as determined by his actual batting performance in a given year.

For example, in 1941 Joe DiMaggio had an 81 percent chance of getting at least one hit in each game (this statistic can be calculated using his total number of hits in the season, the number of games he played and his number of plate appearances). We simulated a mock version of his 1941 season, using the computer equivalent of a trick coin that comes up heads 81 percent of the time.

But the right question is not how likely it was for DiMaggio to have a 56-game hitting streak in 1941. The question is: How likely was it that anyone in the history of baseball would have achieved a streak that long or longer?

To answer this, our simulation repeated the coin-flipping experiments for every player in the history of the game, for every season in which he played. This is what we mean by a simulation of the entire history of baseball.

To tease out the meaningful lessons from random effects (fluky streaks that happen by luck), we redid the whole thing 10,000 times. In each of these simulated histories, somebody holds the record for the longest hitting streak. We tabulated who that player was, when he did it, and how long his streak was.

And suddenly the unlikely becomes likely: we get a very long streak each time we run baseball history. These results are shown in Figure 1. The streaks ranged from 39 games at the shortest, to a freakish baseball universe where the record was a remarkable (and remarkably rare) 109 games.

More than half the time, or in 5,295 baseball universes, the record for the longest hitting streak exceeded 53 games. Two-thirds of the time, the best streak was between 50 and 64 games.

In other words, streaks of 56 games or longer are not at all an unusual occurrence. Forty-two percent of the simulated baseball histories have a streak of DiMaggio's length or longer. You shouldn't be too surprised that someone, at some time in the history of the game, accomplished what DiMaggio did.

The real surprise is when the record was set. Our analysis reveals that 1941 was one of the least likely seasons for such an epic streak to occur.

Figure 2 shows the number of times, out of 10,000 simulations, that the longest streak occurred in a particular year. The likeliest time for the longest streak to have occurred was in the 19th century, back in the misty beginnings of baseball. Or maybe in the 1920s or '30s.

But not in 1941, or afterward. That season was the miracle year in only 19 of our alternate major-league histories. By comparison, in 1,290 of our baseball universes, or more than a tenth, the record was set in a single year: 1894.

And Joe DiMaggio is nowhere near the likeliest player to hold the record for longest hitting streak in baseball history. He is No. 56 on the list. (Fifty-six? Cue "The Twilight Zone" music.) Two old-timers, Hugh Duffy and Willie Keeler, are the most probable record holders. Between them, they set the record in more than a thousand of the parallel baseball universes. Ty Cobb did it nearly 300 times.

DiMaggio held the record 28 times. Plus once more, when it counted.

Samuel Arbesman is a graduate student at Cornell. Steven Strogatz is a professor of applied mathematics there.