

Search: CSICOP On-line

◀ [Home](#) : [Skeptical Inquirer magazine](#) : [September/October 1998](#) : [Buy this back issue](#)

Coincidences: Remarkable or Random?

Most improbable coincidences likely result from play of random events. The very nature of randomness assures that combing random data will yield some pattern.



[Bruce Martin](#)

"You don't believe in telepathy?" My friend, a sober professional, looked askance. "Do you?" I replied. "Of course. So many times I've been out for the evening and suddenly became worried about the kids. Upon calling home, I've learned one is sick, hurt himself, or having nightmares. How else can you explain it?"

Such episodes have happened to us all and it's common to hear the words, "It couldn't be just coincidence." Today the explanation many people reach for involves mental telepathy or psychic stirrings. But should we leap so readily into the arms of a mystic realm? Could such events result from coincidence after all?

There are two features of coincidences not well known among the public. First, we tend to overlook the powerful reinforcement of coincidences, both waking and in dreams, in our memories. Non-coincidental events do not register in our memories with nearly the same intensity. Second, we fail to realize the extent to which highly improbable events occur daily to everyone. It is not possible to estimate all the probabilities of many paired events that occur in our daily lives. We often tend to assign coincidences a lesser probability than they deserve.

However, it is possible to calculate the probabilities of some seemingly improbable events with precision. These examples provide clues as to how our expectations fail to agree with reality.

Coincident Birthdates

In a random selection of twenty-three persons there is a 50 percent chance that at least two of them celebrate the same birthdate. Who has not been surprised at learning this for the first time? The calculation is straightforward. First find the probability that everyone in a group of people have different birthdates (X) and then subtract this fraction from one to obtain the probability of at least one common birthdate in the group (P), $P = 1 - X$. Probabilities range from 0 to 1, or may be expressed as 0 to 100%. For no coincident birthdates a second person has a choice of 364 days, a third person 363 days, and the nth person $366 - n$ days. So the probability for all different birthdates becomes

$$\text{For two people: } X_2 = \frac{365 \cdot 364}{365 \cdot 365}$$

$$\text{For three people: } X_3 = \frac{365 \cdot 364 \cdot 363}{365 \cdot 365 \cdot 365}$$

For n people:

$$X_n = \frac{365 \cdot 364 \cdot 363 \dots 366-n}{365 \cdot 365 \cdot 365 \dots 365} = \frac{365!}{(365-n)! 365^n}$$

With its factorials the last equality is not especially useful unless one possesses the capability of handling very large numbers. It is

instructive to use a spreadsheet or a loop in a computer language to calculate X_n from the first equality for successive values of n . When $n = 23$, one finds $X = 0.493$ and $P = 0.507$. A plot of the probability of at least one common birthdate, P , versus the number of people, n , appears as the right hand curve of circles in Figure 1. The curve shows that the probability of at least two people sharing a common birthdate rises slowly, at first passing just less than 12% probability with ten people, rising through 50% probability at the open circle corresponding to twenty-three people, then flattening out and reaching 90% probability in a group of forty-one people. This means that on the average, out of ten random groups of forty-one persons, in nine of them at least two persons will celebrate identical birthdates. No mysterious forces are needed to explain this coincidence.

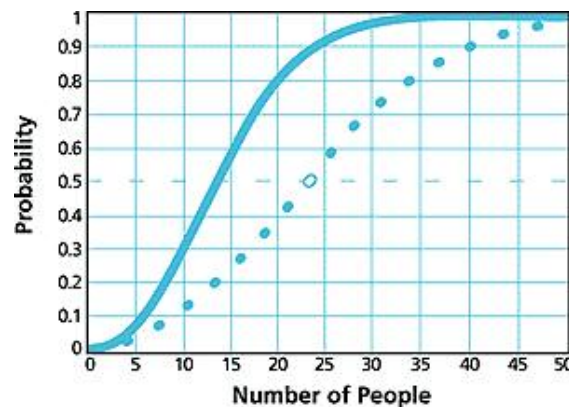


Figure 1. Probabilities of Coincident Birthdates: The right-hand curve of circles represents the probability that in a random group of people at least two celebrate the same birthdate. As indicated by the open circle just above the horizontal line at 0.50 probability, an even 50% chance is achieved at just 23 people. The left-hand curve represents the probability that in a random group of people at least two share a birthdate within one day of each other. The 50% chance for this three-day coincidence occurs with just 14 people.

Note that the probability of coincident birthdays for $23 \times 23 = 46$ people is not 100%, as some might suppose, but 95% as shown by the right-hand curve in Figure 1. Extension of the curve beyond the limit of Figure 1 reveals that fifty-seven people produce a 99% probability of coincident birthdays.

The same principle may be used to calculate the probability that at least two people in a random group possess birthdates within one day (same and two adjacent days). This condition is less restrictive than the former, and 50% probability is passed with just fourteen people. The left-hand curve in Figure 1 shows a plot for the probabilities of within-one-day birthdates.

Delving a little deeper into some aspects of the probabilities of identical birthdates provides additional insight. Note that we said several times "at least two people" sharing a common birthdate. As the group size increases the chances for multiple coincidences also increase. The descending curve at the left of Figure 2 represents the probability of no coincidences (NC) of birthdates, identical to the X_n values calculated above. The first curve with a maximum plots the probability of only one pair (1P) sharing an identical birthdate. The maximum occurs at twenty-eight people with a probability of almost 0.39. As the group becomes larger the probability of other coincidences increases as well. The second curve with a maximum represents the probability of exactly two pairs (2P) sharing an identical birthdate. Its maximum occurs at thirty-nine people with a probability of 0.28. The last, rising curve

in Figure 2 plots the total probabilities of all remaining coincidences ($>2P$), consisting of three pairs, triplets, etc. For all numbers of people, the probabilities of all four curves total 1.00.

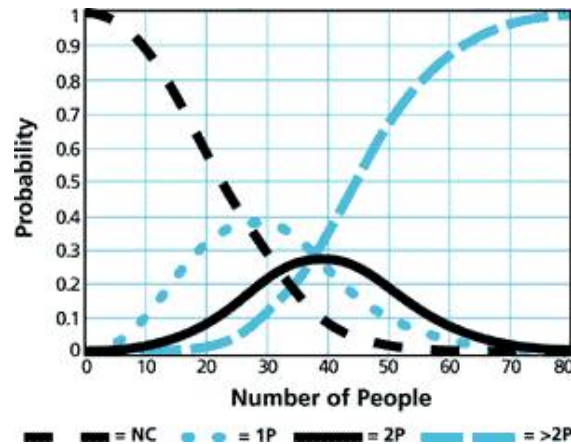


Figure 2. Probabilities of Multiple Coincident Birthdates: The descending curve at the left represents the probability for no shared birthdates, no coincidences (NC). The first curve with a maximum plots the probability of only one pair (1P) with an identical birthdate. The second curve with a maximum represents the probability of exactly two pairs (2P) sharing identical birthdates (different date for each pair). The ascending curve at the right plots the probability of all other coincidences ($>2P$), three pairs, triplet, etc. For any number of people the probabilities of the four curves total 1.00.

Figure 2 shows that for twenty-three people the probabilities are 0.36 for one pair, 0.11 for two pairs, and 0.03 for the total of all other coincidences for a probability sum of 0.50. We have broken down the 0.50 probability for at least one coincidence discussed above for twenty-three people into component contributions. For twenty-three people the probability of no coincidences is also 0.50, as shown in the descending curve (NC) of Figure 2. There is an almost triple intersection at thirty-eight people where the chance of 1 identical pair, 2 identical pairs, and the total of all other coincidences is 28-29%. For thirty-eight or more people the total of all other coincidences becomes greater than the exactly one and two pair possibilities, and passes through 50% chance at forty-five people. In a random group of more than forty-five people there is a better than even chance that there are more than two coincidental birthdates.

What this series of calculations boils down to is this: If coincident birthdates are so much more common than we would have guessed, isn't it likely that many of those other striking coincidences in our lives are the outcome of probability as well? We should not multiply hypotheses: the principle of Occam's Razor states that the simplest explanation is to be preferred.

Presidential Coincidences

Consider the birth and death dates of American presidents to see how this reasoning works in real cases. There have been forty-one presidential births, and Figure 1 indicates a 90% probability that at least

Table 1
Presidential Coincidences
Births, 41 cases (90%)

November 2 James K. Polk
Warren G. Harding
Deaths, 36 cases (83%)

| | | |
|---|---------|--|
| two presidents should have been born on the same day. There is one such coincidence: James K. Polk and Warren G. Harding were both born on November 2. The result appears in Table 1. With forty-one cases there is a 66% chance of a second coincidence, but none has yet occurred. [The result may be obtained by adding the probabilities for forty-one persons for the 2P (0.28) and >2P (0.38) curves of Figure 2.] Perhaps the next president's birthdate will coincide with one of the previous forty-one. (The birthdates of neither Albert Gore nor Colin Powell do so.) | March 8 | Millard Fillmore William Howard Taft |
| | July 4 | John Adams Thomas Jefferson James Monroe |

Of the thirty-six dead presidents Figure 1 indicates an 83% probability that at least two should have died on the same date. The results also appear in Table 1. Both Millard Fillmore and William Howard Taft died on March 8. With 36 cases there is a 51% chance of a second coincidence.

In what seems an astounding coincidence, three early presidents died on July 4, as listed in Table 1. Both John Adams and Thomas Jefferson died in the same year, 1826, on the fiftieth anniversary of their signing the Declaration of Independence. Adams's final words, that his long-time rival and correspondent Jefferson "still lives," were mistaken, as Jefferson had died earlier that same day. James Monroe died on the same date five years later. Presidential scholars suggest that the former early presidents made an effort to hang on till July 4. James Madison rejected stimulants that might have prolonged his life, and he died six days earlier on June 28 (in 1836). It seems evident that for the deaths of several presidents July 4 is not a random date. Only one president, Calvin Coolidge, was born on July 4.

| | |
|--|---|
| An important point in all of the preceding is that no birthdate was specified in advance. Table 2 lists the crowd sizes for 50% probabilities. The first entry restates what we already know: a group of twenty-three suffices for at least two to possess the same unspecified birthdate. If we specify a particular birthdate, such as today, a crowd of 253 people is required to have an even chance for one person with that birthdate. For at least two persons to possess a specified birthdate, the 50% probability is not reached until there is a mob of 613 people. This huge difference of twenty-three versus 613 for 50% probability of at least two persons with a common birthdate is due to the fact that the date is unspecified for the group and specified for the mob. That some improbable event will occur is likely; that a particular one will occur is unlikely. If we look at our personal coincidences, we see that they were rarely predicted in advance. | Table 2 Crowd Size for 50% Probabilities |
| | At least two have the same, unspecified birthday 23 |
| | At least one has the specified birthday 253 |
| | At least two have the same, specified birthday 613 |

Abraham Lincoln and John Kennedy

It is always possible to comb random data to find some regularities. A well-known qualitative example is the comparison of coincidences in the lives of Abraham Lincoln and John Kennedy, two presidents with seven letters in their last names, and elected to office 100 years apart, 1860 and 1960. Both were assassinated on Friday in the presence of their wives, Lincoln in Ford's theater and Kennedy in an automobile made by the Ford motor company. Both assassins went by three names: John Wilkes Booth and Lee Harvey Oswald, with fifteen letters in each complete name. Oswald shot Kennedy from a warehouse and fled to a theater, and Booth

shot Lincoln in a theater and fled to a barn (a kind of warehouse). Both succeeding vice-presidents were southern Democrats and former senators named Johnson (Andrew and Lyndon), with thirteen letters in their names and born 100 years apart, 1808 and 1908.

But if we compare other relevant attributes we fail to find coincidences. Lincoln and Kennedy were born and died in different months, dates, and states, and neither date is 100 years apart. Their ages at death were different, as were the names of their wives. Of course, had any of these features corresponded for the two presidents, it would have been included in the list of "mysterious" coincidences. For any two people with reasonably eventful lives it is possible to find coincidences between them. Two people meeting at a party often find some striking coincidence between them, but what it is -- birthdate, hometown, etc. -- is not predicted in advance

Bridge Hands

In the card game bridge there are a possible 635,013,559,600 different thirteen-card hands. This number of hands could be realized if all the people in the world played bridge for a day. For an individual it would take several million years of continuous playing to be dealt each of these hands. Yet any given hand held by a player is equally probable, or rather, equally improbable, as its probability is $1/635,013,559,600$ or a little better than one part in a million million. Any hand is just as improbable as thirteen spades. Bridge hands are an example of the daily occurrence of very improbable events, but of course, the hands are not specified in advance.

Consider a group of just 10 or more students in a classroom of a college that draws students from several states. During school session, numerous such classrooms exist each day. Yet the odds against predicting the exact make up of any classroom ten years in advance (all the students and teacher born by then) are truly astronomical. This is another example of the daily occurrence of highly improbable events.

Runs of Heads and Tails

What sequence of head(H) and tails(T) might you expect in random tossing of a coin? Not all heads nor all tails, nor even the alternating sequence (HTHTHTHT), as this series is obviously regular and not random. In a random sequence we expect runs of both heads and tails. We can simulate progressions of coin tosses from a random sequence of numbers.

So far as is known, the decimal digits of the irrational number p , which multiplies the diameter of a circle to obtain the circumference, are random. This does not mean that every time p is calculated a different result is obtained, but rather that the value of any single digit is not predictable from preceding digits. An example of a pattern leading to predictability is the sequence of decimal digits in the fraction $1/7 = 0.142857142857142857. . .$, where there is an obvious repeat every six digits.

The decimal digits of p have been calculated to hundreds of millions of digits by high-speed computers, but we list only the first 100 digits in four rows of 25 digits.

3.141592653589793238462643
 38327950288419716939937510
 582097494459230781**64062862**
089986280348253421170679

There are fifty-one even digits and forty-nine odd digits. There is an almost an even distribution when the first 100 decimal digits are divided in another way: forty-nine digits from 0 to 4 and

fifty-one digits from 5 to 9.

Since the decimal digits of p are random, we may simulate a random sequence of heads and tails in coin tossing by assigning even digits to heads and odd digits to tails. The sequence of heads and tails in 100 tosses with 25 tosses per line becomes

```
THTTTHHTTTHTTTTHTHHHHHT
THTHTTTHHHHTTTTHTTTTTT
THTHHHTTHTHHTTHTHTHHHHH
HHHHTTTHHHHTHHHTTHTTTHTT
```

Combing the random sequence we find some regularities, such as the alternating sequence of eight tosses from 62-69 (underlined). The probability of an alternating sequence of 8 tosses is once in $27 = 128$ tosses. There are some long runs of all heads and all tails. There are two runs of 5 heads, one run of 6 heads, one run of 8 tails, and a surprising run of 10 heads. The p decimal digits 69-78 are all even (refer to underlined digits). A run of ten even digits should occur only once in $2^{10} = 1,024$ digits. Yet such a run occurs within the first eighty digits.

So what have we here? A proof that the decimal digits of p are not random? No, what we have instead is a demonstration of how it is always possible to comb random data and find regularities not specified in advance. Since ten even digits occur within the first 100 decimal digits of p , we might (mistakenly) think we are on to something, and that such a run might occur frequently. In fact a run of ten even digits does not occur again in the first 1,000 decimal digits of p . In the first 1,000 digits a single run of ten odd digits occurs from 411-420.

The point is that the very nature of randomness assures that combing random data will yield some pattern. But what that pattern is cannot be specified in advance. If someone finds a pattern combing random data, he or she may use it as a hypothesis for investigation of more data but should never make a general conclusion from it. In our example we discovered (but did not predict) ten even digits within the first 100 digits but not again in the next 900 digits. For confirmation of a trend, the target data must be stated in advance of data inspection. If an unexpected pattern does emerge during inspection after the data is obtained, the pattern can be used as a hypothesis for obtaining and inspecting an entirely new set of data.

The heads and tails sequence may be applied in other ways. Consider a football quarterback who completes 50% of his passes or a basketball player who makes 50% of his or her free throws. Assign heads (H) to a pass completion or made free throw and tails (T) to a miss, and then one expects long runs of completions and misses as shown in the HT sequence above. Most hot and cold streaks in sports are just the consequence of randomness. The "hot hand" is most often an illusion of significance that appears in data sets that are random.

We may utilize the random sequence of p decimal digits to find likely streaks for a .300 hitter in baseball. For example, assign the digits 0, 2, and 4 to hits and the other seven digits to outs. Then, out of the first 100 decimal digits there are 30 hits and 70 outs. If we divide the sequence of 100 digits into successive groups of four, a representative number of bats per game, we obtain the results for twenty-five games. Our .300 hitter then goes hitless in four games (three in succession for a "slump"), strokes one hit in thirteen games, two hits in seven games, three hits in one game, and has no game in which he gets four hits. Astonishingly, the batter gets at least one hit in the last thirteen games, considered enough to be a real "streak." But this "streak" arises out of the random sequence of p decimal digits. A batter's slump or hitting streak is likely just the result of randomness in play.

Clearly, unspecified improbable coincidences occur daily to

everyone, and these coincidences are most likely the result of randomness. If the data set is large enough, coincidences are sure to appear, as demonstrated with the first 100 decimal digits of p . The chance of tossing five straight heads is only 3 percent, but for 100 tosses the chance becomes 96 percent. Though applied in a different context, Ramsey theory (Scientific American, July 1990) states that "Every large set of numbers, points, or objects necessarily contains a highly regular pattern." It is not necessary to posit mysterious forces to explain coincidences.

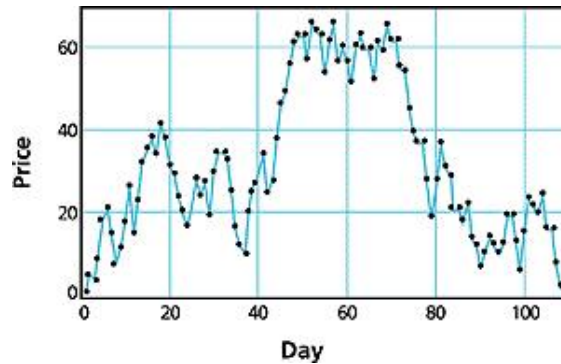


Figure 3. Stock Market Simulation: Daily stock market action presented as price for 109 trading days generated from the random decimal digits of p . The representative "head and shoulders top" is shown to be consistent with random play of the market. Of course, the number of days is flexible, one decimal digit may represent any fraction or number of days. [See Note](#) for a description of how the price action was generated.

Random Prices in the Stock Market

Given the current fascination with the long bull market in stocks, we can generate an even more interesting result from the random decimal digits of p . Let us plot on the x-axis the number of the decimal digit and on the y-axis a price value that is generated from the decimal digits as described in the Figure 3 caption and [Note](#) so that there is an arbitrary and equal balance between the up and down directions for price. For the first 108 decimal digits of p the entire plot is in positive territory. Starting at zero the plot works its way haltingly to increasingly positive values, attaining a plateau from the 48-71 decimal digits before it begins to work its way down, almost returning to zero on the 99th digit, and crossing into negative territory after the 108th decimal digit. To a stock market technician this plot represents a head and shoulders top in a plot of a stock price or stock market price versus time. It is all there in Figure 3, a top and shoulders on both sides of the top. Yet this plot was generated from the first 109 random decimal digits of p ! The maximum value of 65 on the y-axis is reached three times in the plateau region and is more than 7 times greater than the maximum single move of 9. Therefore, we may conclude that a head and shoulders top in stock or commodity prices may represent nothing more than random play in the markets. (Over the longer term there is a rising trend in stock market averages.)

A recent sweepstakes received in the mail offered a grand prize of \$5,000,000. The fine print stated the chances of winning this prize as one in 200,000,000. Out of this large population some one person will win the sweepstakes. With such incredibly unfavorable odds each person must decide for him or herself whether it is worth the time and the first class postage to return the entry. The sure big winner appears to be the postal service, which garners more than ten times the grand prize amount in postage.

So, the next time you hear, "It couldn't be just coincidence," you

will be fully justified in answering, "Why not?"

CSICOP Presidential Coincidences Contest

Back in 1992, the Skeptical Inquirer held a Spooky Presidential Coincidences Contest, in response to Ann Landers printing "for the zillionth time" a list of chilling parallels between John F. Kennedy and Abraham Lincoln. The task was for readers to come up with their own list of coincidences between other pairs of presidents. There were two contest winners, Arturo Magidin of Mexico City, and Chris Fishel, a student at the University of Virginia. Magidin came up with sixteen stunning coincidences between Kennedy and former Mexican President Alvaro Obregón, while Fishel managed to come up with lists of coincidences between no fewer than twenty-one different pairs of U.S. presidents.

A few examples from Magidin's list: Both "Kennedy" and "Obregón" have seven letters each; each was assassinated; both their assassins had three names and died shortly after killing the president; Kennedy and Obregón were both married in years ending in 3, each had a son who died shortly after birth, and both came from large families and died in their forties.

Fishel came up with dozens of coincidences; here are a few between Thomas Jefferson and Andrew Jackson. Both men served two full terms; both their wives died before they became president; each had six-letter first names; both were in debt at the time of their deaths; each had a state capital named after him, and both their predecessors refused to attend their inaugurations. [For more information and the full lists, see SI Spring 1992, 16(3); and Winter 1993, 17(2).]

Acknowledgement

I am indebted to Professor Russell N. Grimes of the University of Virginia for discussions of expressions leading to Figure 2 and Table 2.

Note

The price action was generated in a positive direction when the preceding digit is odd (except for 5) and in a negative direction when the preceding digit is even, with the magnitude of the direction given by the value of the digit. Thus preceding odd digits $1 + 3 + 7 + 9 = 20$ generate a positive direction and preceding even digits $0 + 2 + 4 + 6 + 8 = 20$ a negative direction. The sum in the two directions is the same. For the left over digit 5 the direction is up or down depending upon whether the previous digit is odd or even, respectively. In the first 108 decimal digits there are eight 5s, half each generating positive and negative directions. Therefore we have a perfectly arbitrary and equal balance between the positive and negative directions.

References

- Epstein, Richard A. 1967. *The Theory of Gambling and Statistical Logic*. New York: Academic Press.
- Falk, Ruma. 1981. On Coincidences. *Skeptical Inquirer* 6(2): Winter: 18-31.
- Graham, Ronald L., and Joel H. Spencer. 1990. Ramsey Theory. *Scientific American* 263 (1): 112-117 (July).
- Paulos, John Allen. 1988. *Innumeracy*. New York: Random House.
- Weaver, Warren. 1963. *Lady Luck*. Garden City, NY: Anchor Books.

About the Author

Bruce Martin is Professor Emeritus of Chemistry at the University of Virginia, Charlottesville, Virginia.