15.053/8 May 9 and 11, 2013

Why is elementary probability so hard to understand?

A top 10 list

Quotes of the day

"Misunderstanding of probability may be the greatest of all impediments to scientific literacy."

-- Stephen J. Gould

"It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge."

Pierre Simon Laplace

"This branch of mathematics [probability] is the only one, I believe, in which good writers frequently get results which are entirely erroneous."

--Charles Pierce

Background

Harold Larnder Prize

“The Zen of probability”

A top 10-list: challenges and suggestions

1. Probability does not work backwards.

- Frequentists vs Bayesians

- Law of large numbers

Image by MIT OpenCourseWare.

Exercise: Suppose we select 10 random numbers between 0 and 99. What is the probability that we get the following sequence:

98 43 38 87 76 90 28 54 99 95

1. 1 in 100 billion
2. 1 in 100 trillion
3. 1 in 100 quadrillion
4. 1 in 100 quintillion

Probability does not work backwards

What is the probability that someone would ever be born with your exact DNA sequence?

Physicists have pointed out that if physical constants were slightly different, there would be no possibility of life. Is this evidence that intelligent life was a reason that the universe was created?

Probability can be used to analyze the past. But it's very subtle and slippery.

John Doe was accused of a murder in NYC. During his trial, experts stated that there was a partial fingerprint on the murder weapon that matched John Doe's fingerprint.

Suppose that there was less than a 1 in a million chance that a random person would be positive match with the partial fingerprint found on the murder weapon.

Is this extremely strong evidence that it was John Doe's fingerprint on the murder weapon?

2. What would Bayes say (WWBS)?

Assumptions most likely to lead to error (after learning information)

- **events are independent**
- **events are equally likely**

To help remember that these assumptions are often wrong: ask WWBS?

Simulating a probability of 1/3 with a coin.

Suppose you want to select A or B with equal probability.
Flip a coin.

Suppose you want to select A or B or C with equal probability.
Flip a coin twice, and repeat if necessary.

H - H  **A**

Prob(H-H | Not T-T) = 1/3

H - T  **B**

T - H  **C**

T - T  **Repeat**

2. What would Bayes say (WWBS)?

3 coin problem. Suppose you have three coins.

- coin with two heads
- coin with two tails
- coin with one head and one tail

2. What would Bayes say (WWBS)?

Suppose you choose one at random and observe one of its sides. If you observe a head, what is the probability that it is the two headed coin?

1. 1/3
2. 1/2
3. 2/3

3. Its large! it's the law!

- Take advantage of the law of large numbers, and then turn probability analysis into counting.

3-coin problem.

– Label the coins A, B, C

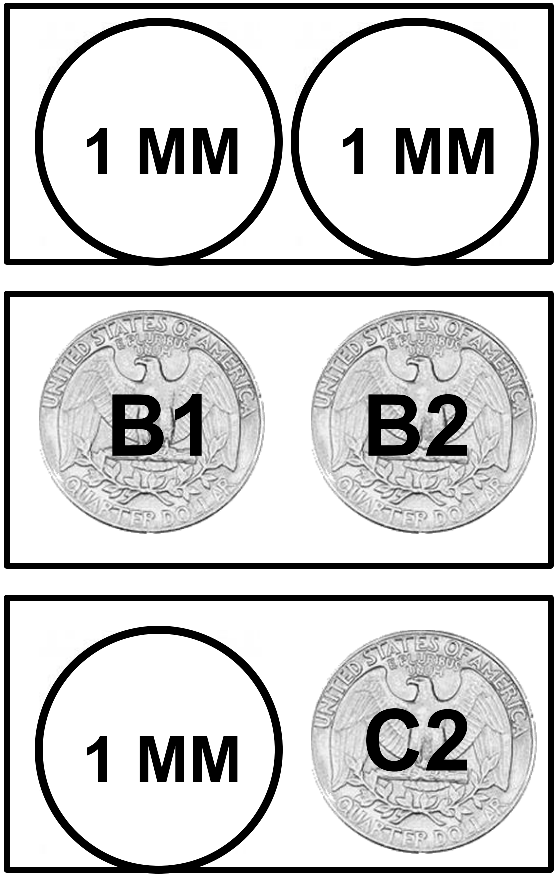
– Label the sides 1 and 2

– Select a random coin and side, 6 million times.

– How many times does a head come up?

– Of these, how many are due to the 2- headed coin?

– 2/3



4. Context and hidden assumptions

If I toss a fair coin 50 times and it comes up heads each time, what is the probability that it comes up heads the next time?

1. **1/2**
2. **Something else**

The coin example, from a different angle

Suppose that there is a 99.9999% chance that a coin is fair, and a 1 in a million chance that the coin is two-headed.

If the first 50 tosses of the coin are heads, what is the probability that it is the 2-headed coin.

$$\frac{\frac{1}{1{,}000{,}000}}{\frac{1}{1{,}000{,}000}+\frac{1}{2^{50}}} = 99.99999991\%$$

4. Context and hidden assumptions

A parent with two children reveals that at least one of them is a boy. What is the probability that both are boys?

1. **1/3**
2. **1/2**
3. **something else**

A mental experiment

4 million parents of two children are selected at random.

We list the oldest child first.

- **1 million B-B**
- **1 million B-G**
- **1 million G-B**
- **1 million G-G**

Prob(B-B | at least one boy) = 1 / 3

But, what was hidden?

A parent with two children reveals that at least one of them is a boy.

- **1 million B-B**
- **1 million B-G**
- **1 million G-B**
- **1 million G-G**

In the first three cases, the parent reveals that at least one child is a boy.

What does a parent of two girls reveal?

5. The sound of the dog not barking

Bob was bragging about his high school football team. He claimed that it was ranked as one of the top 25 high school football teams in Texas.

What is the probability that it was one of the top 10 high school football teams?

Examples of advertising claims

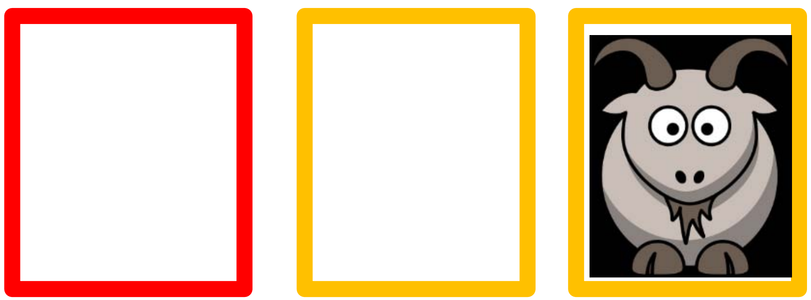
"Easy-Off has 33% more cleaning power than another popular brand."

"Special Morning--33% more nutrition."

"Wonder Bread helps build strong bodies 12 ways."

The Monte Hall Problem

1. You pick door number 1.
2. Monte Hall reveals a goat behind door number 3.
3. Should you switch to door number 2?



- About the past
- Conditional probabilities
- Hidden assumptions

- Depends on what Monte Hall says and doesn’t say.

Goat image: public domain

Simulating Monte Hall 6 MM times.

You pick door number 1.

2 million times

2 million times

2 million times

Goat and car images: public domain

6. Simple Heuristics

We employ heuristics to develop answers quickly.

These are rules of thumb.

As a rule of thumb, these heuristics often work quite well.

Albert and the room lottery

In college, we had a lottery for choosing rooms.

- **There were 75 persons and 75 rooms**
- **#s 1 to 75 were put into a hat.**
- **Each person drew a number**
- **Persons selected rooms in the order of the number they drew, with “1” going first.**

Why is elementary probability so hard?

1. Probability does not work backwards?
2. What would Bayes say (WWBS)?
3. It's large, it's the law?
4. Context and hidden assumptions
5. The sound of the dog not barking
6. Simple heuristics

6. The birthday paradox

Suppose that there are 12 persons in a room.
Which is more likely:

1. **At least two of the twelve have the same birthday?**
2. **All twelve were born in different months.**

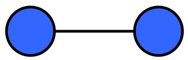
Consider the first three persons.

- **Prob(birthday match) < 1%**
- **Prob(3 different months) > 75%**

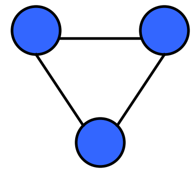
Consider all twelve persons.

- **Prob(birthday match) ~ 1/6**
- **Prob(12 different months) <** $\dfrac{1}{18,000}$

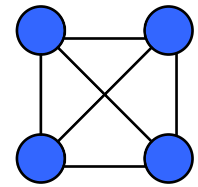
Birthday problem



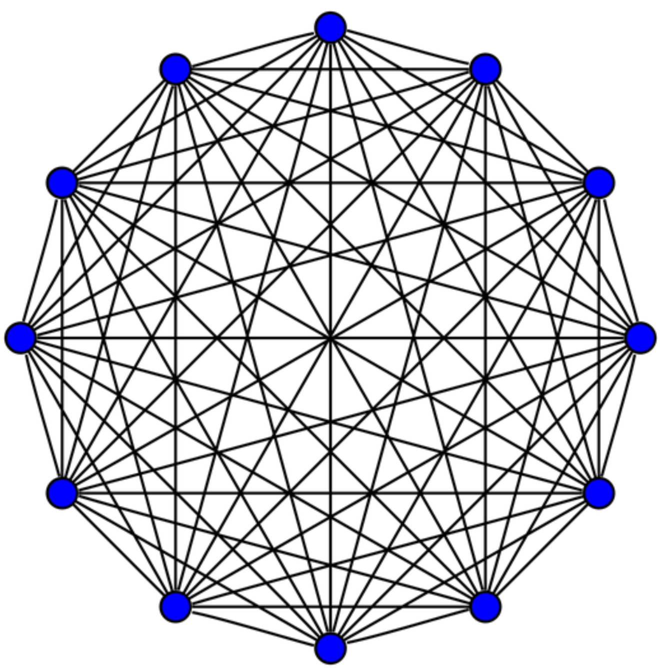
2 persons
1 link



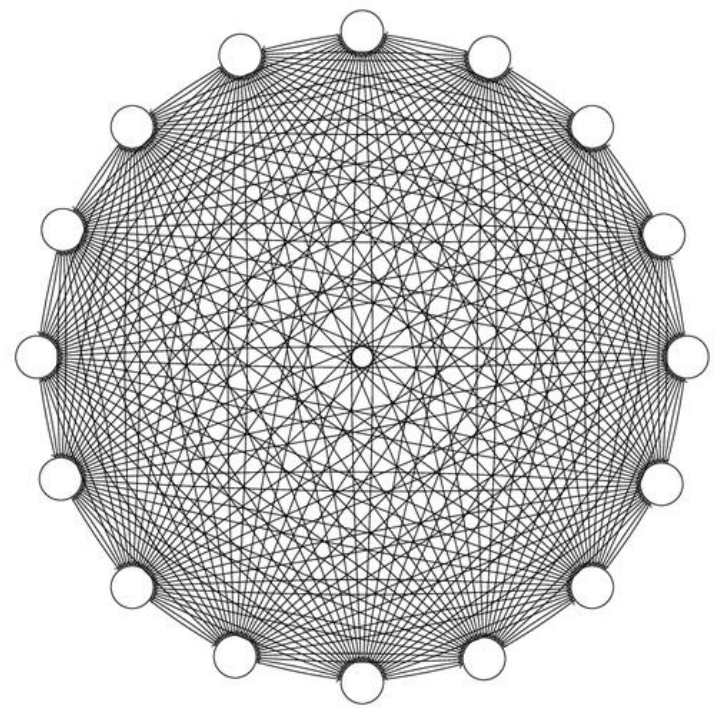
3 persons
3 links



4 persons
6 links



12 persons
66 links



16 persons
120 links

6. The law of small numbers

Kidney cancer incidence study

- **3,141 counties in U.S.**

Counties with lowest incidence of kidney cancer

- **rural**
- **sparsely populated**
- **traditionally Republican states**

What is the explanation?

6. Simple heuristics: the law of small numbers

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

1. **The larger hospital**
2. **The smaller hospital**
3. **About the same (that is, within 5% of each other)**

6. The law of VERY small numbers
(not in Kahneman's book)

- **Firing decisions are often based on a single mistake**
- **We often like or dislike someone based on their physical similarity to someone whom we like or dislike**
- **We make conclusions about groups of people based on the actions of very few of them.**

The cancer study again

- **One would expect small counties to have larger variation.**

- **The counties with the lowest incidence will be small counties.**

- **The counties with the highest incidence will be small counties**

7. Segal’s law

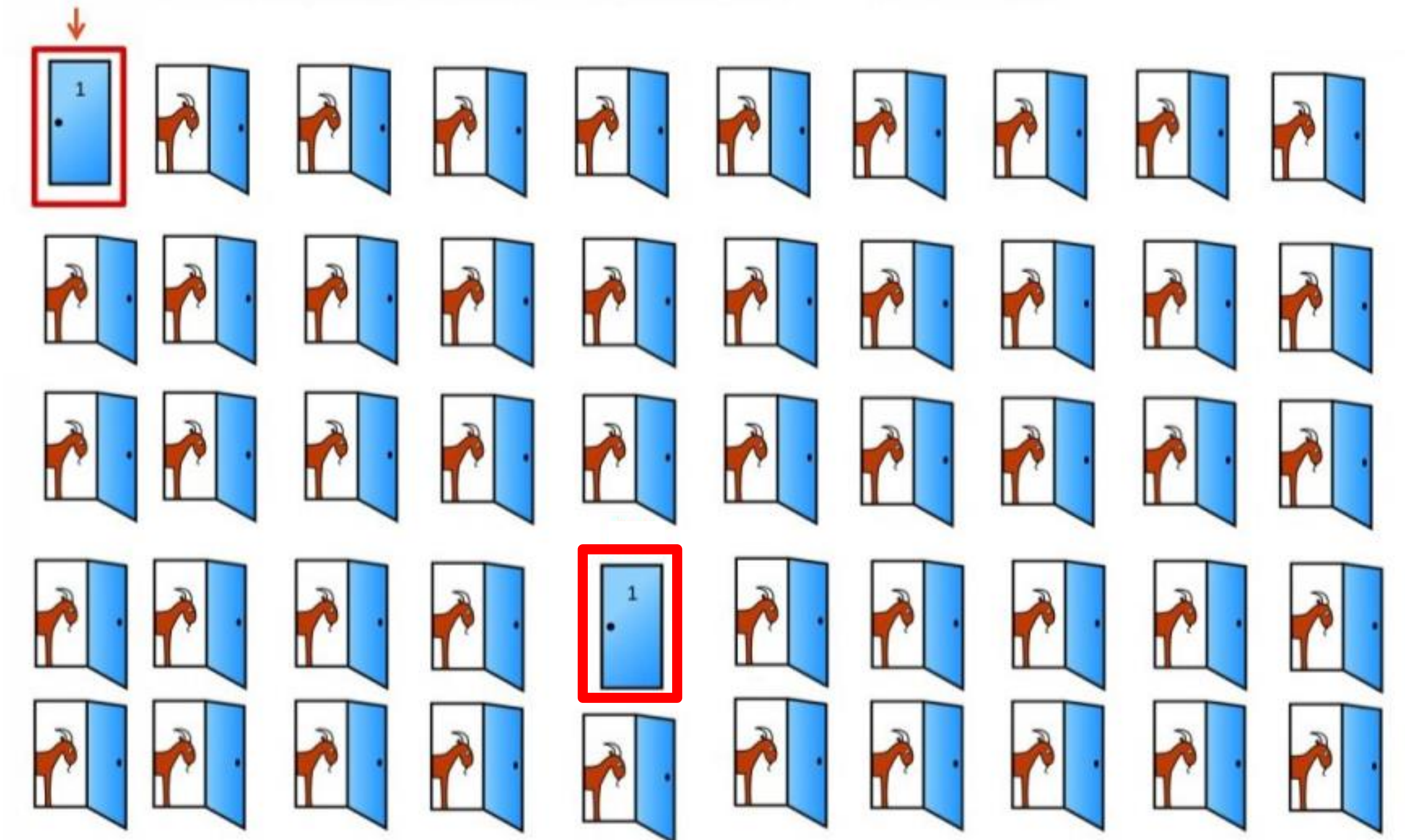
"A man with a watch knows what time it is. A man with two watches is never sure.”

Corollary. A man with a watch refuses to believe the time given by a different watch.

WYSIATI: What you see is all there is.

- **General concept (in Thinking Fast and Slow)**
- **Hindsight bias**
- **Bias towards illusion of understanding**
- **Premonitions and intuitions about the past**

The Monte Hall Problem again



Public domain image

Suppose Monte Hall revealed all but door number 35 (out of 50). Would you want to switch?

8. Patterns or randomness?

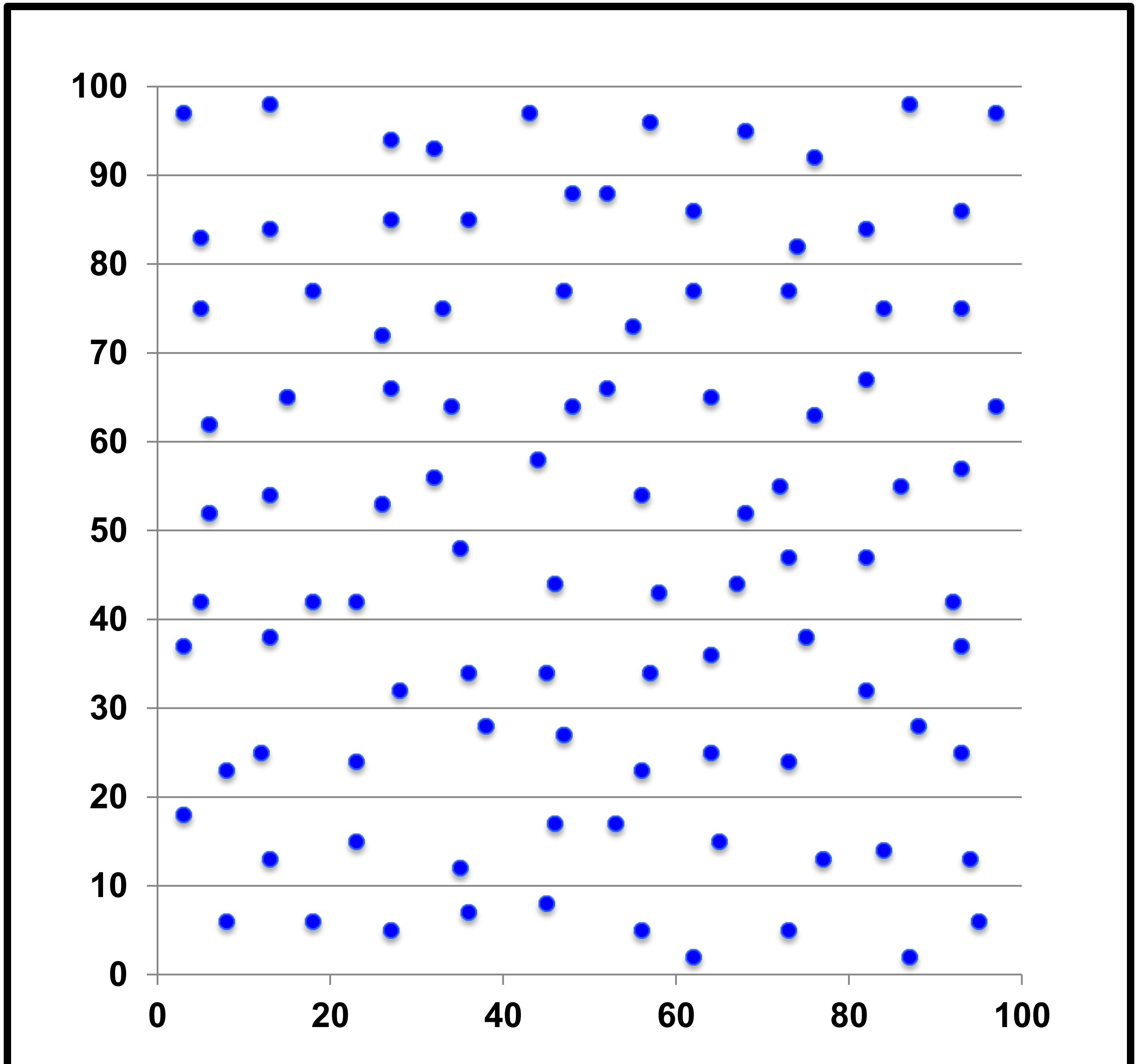
- **Recognizing randomness**
- **Recognizing patterns**

The gender of 6 babies will be observed in the hospital. Which of the following three are more likely to be observed.

1. **B B B G G G**
2. **G G G G G G**
3. **B G B B G B**

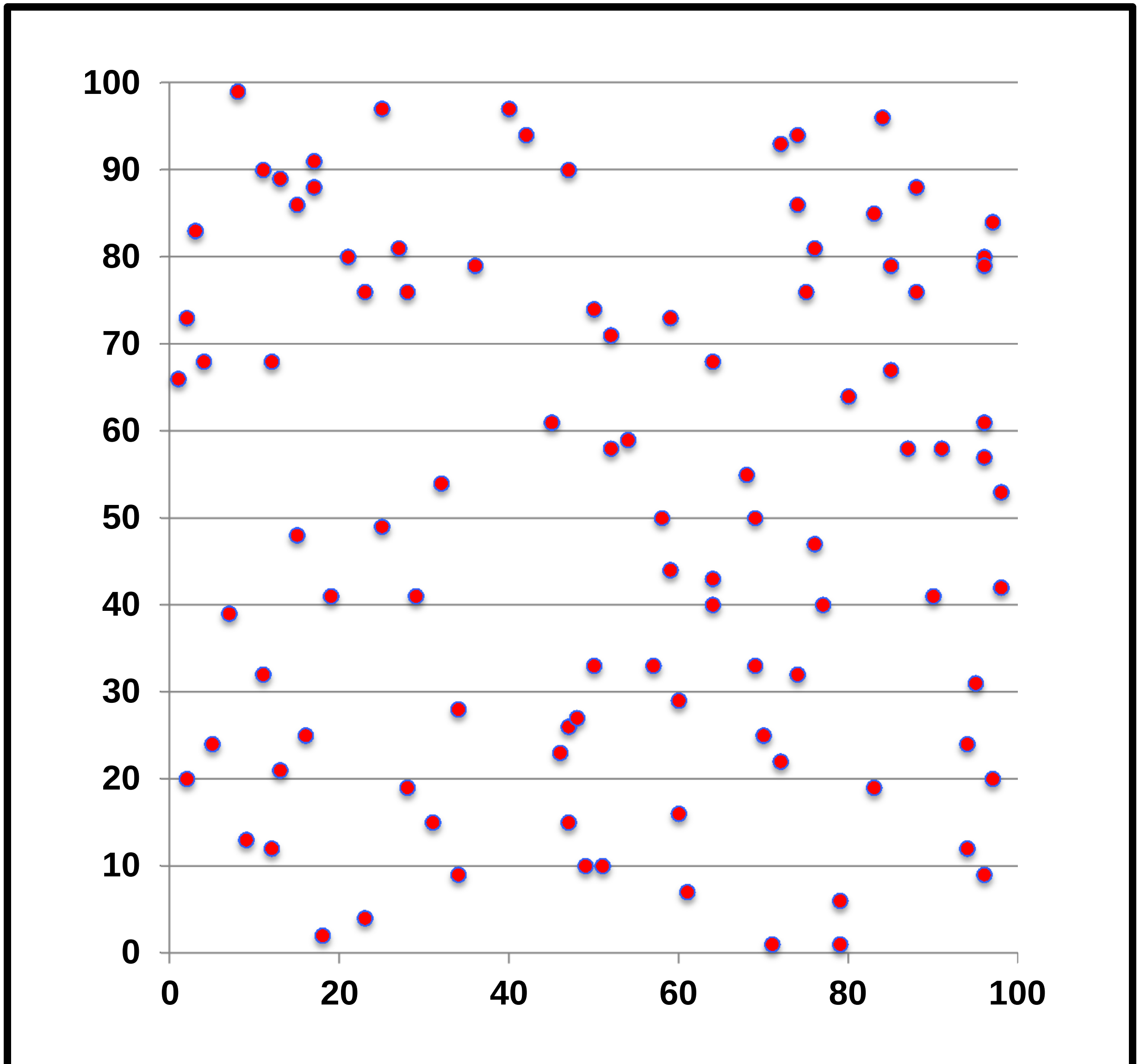
Does uniform randomness look this?

A



Or this? or neither?

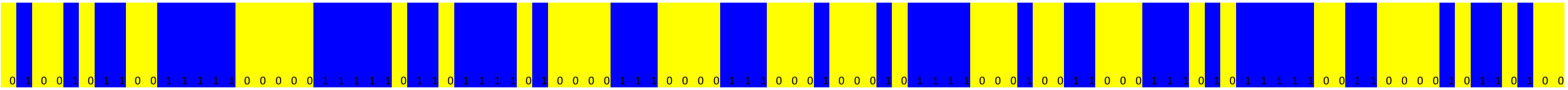
B



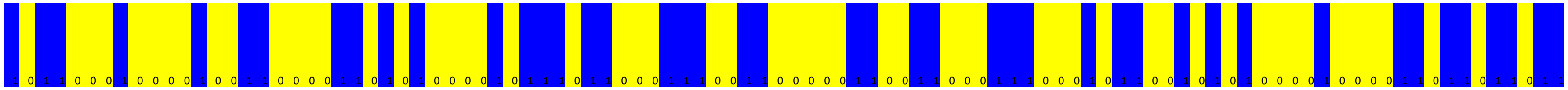
1. A
2. B
3. Neither

Which is generated uniformly at random?

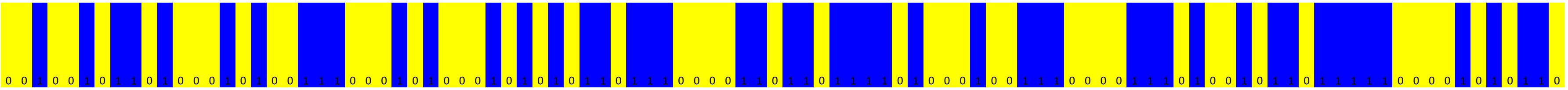
A



B



C



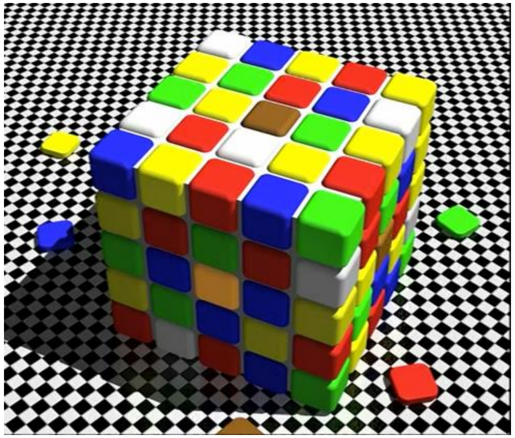
The yellow segments are sequences of 0's.

The blue segments are sequences of 1's.

1. **Sequence A**
2. **Sequence B**
3. **Sequence C**
4. **A, B, and C.**

Streaks in sports

- Hot hand in sport
  - nearly universally believed
  - basketball players have streaks in which they are more likely to make baskets
  - statistical analysis indicates that this is almost entirely an illusion.

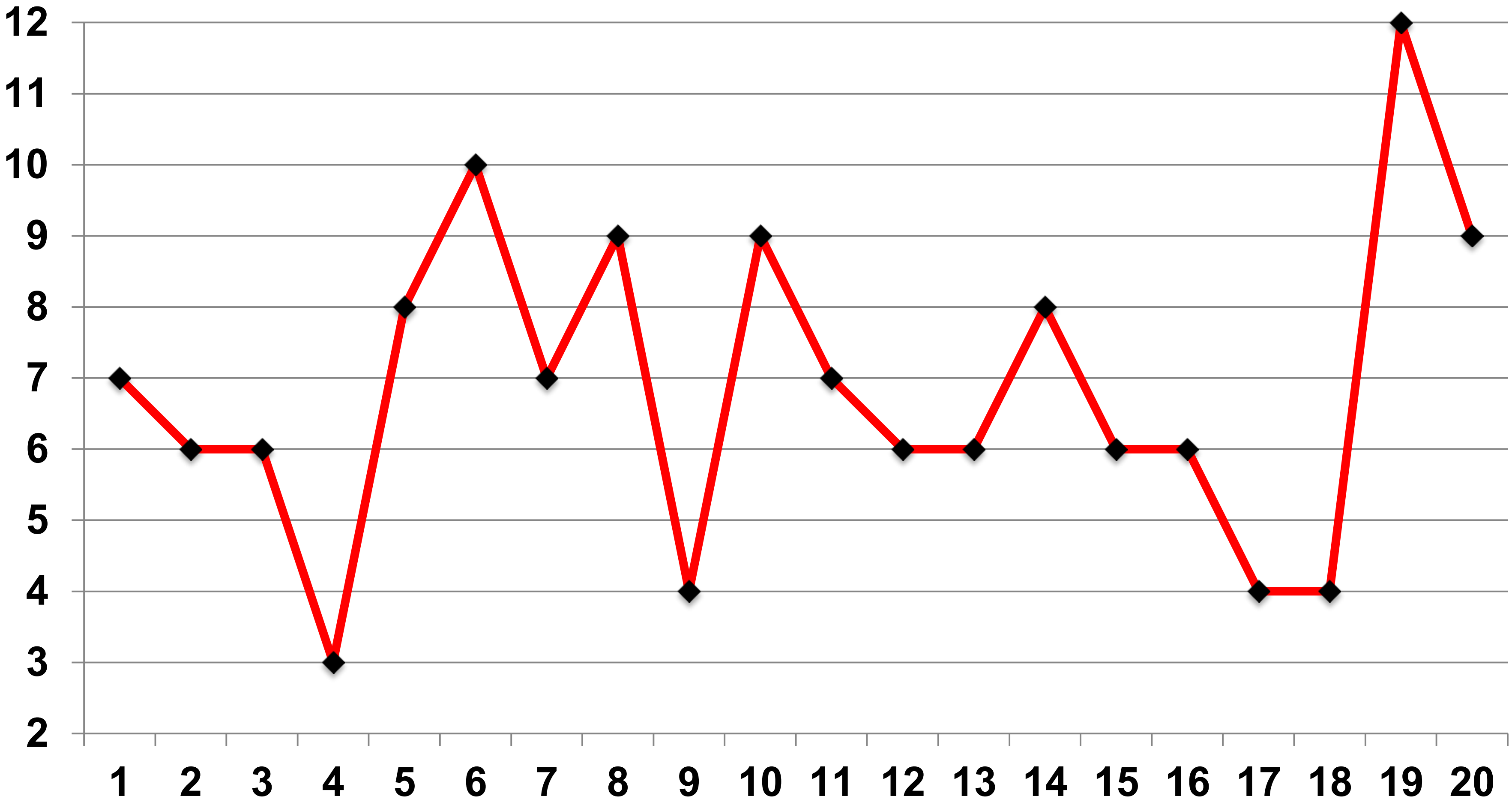


Courtesy of R. Beau Lotto. Used with permission.

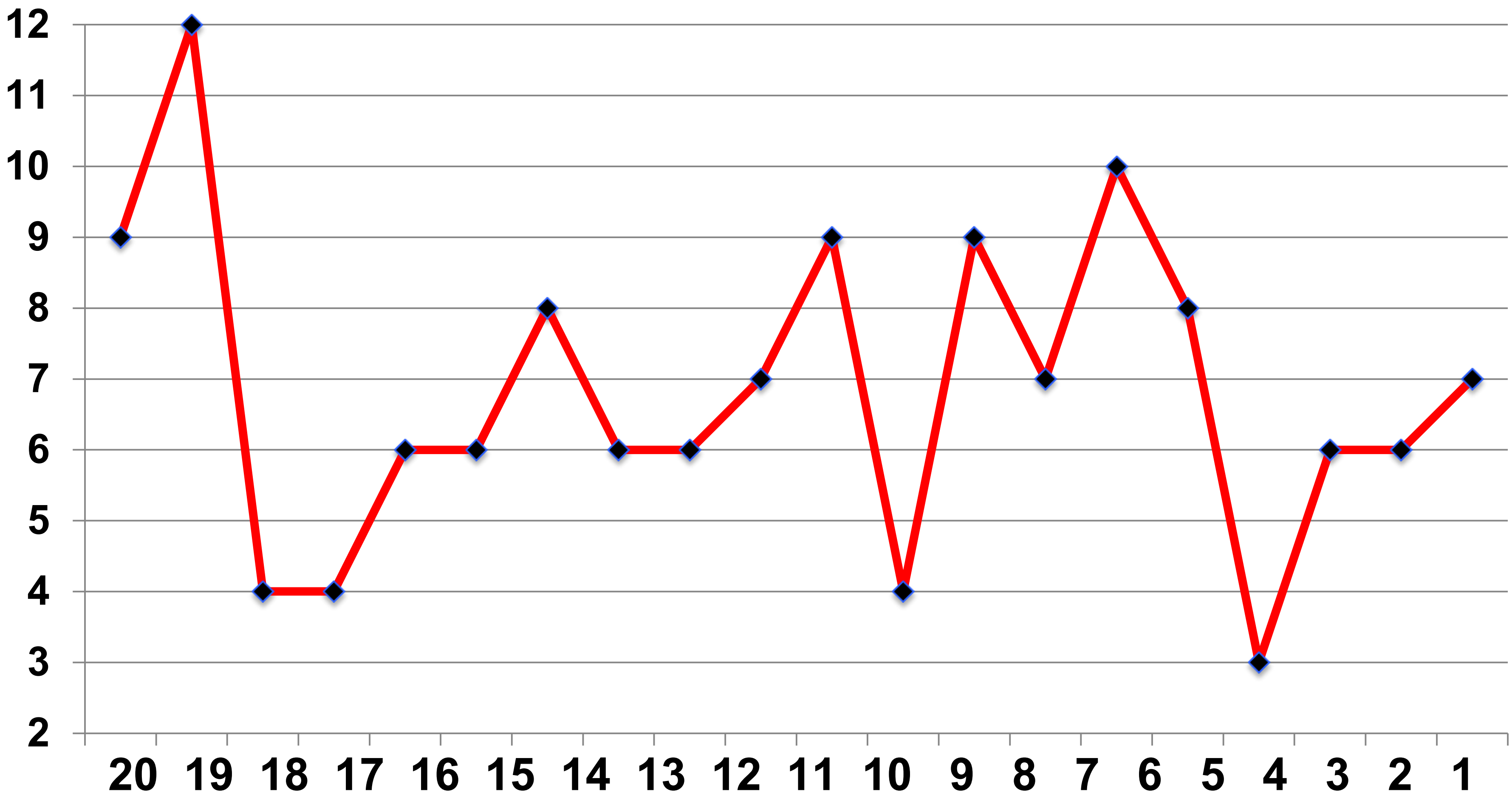
Optical illusions:
start at 8:00
Colors on a 2-dimensional picture or video..

http://www.lottolab.org/illusiondemos/Demo%2014.html

Israeli airforce flight instructors



The dice example, but in reverse order



Regression to the mean

9. Prospect theory and risk assessment

Problem 1. Which do you choose?

1. **$900 for sure or**
2. **90% chance to get $1000.**

Problem 2. Which do you choose?

1. **Lose $900 for sure or**
2. **90% chance to lose $1000.**

10. Interesting connections help.

- **Birthday paradox and coincidences.**

10. Interesting connections help.

- **Birthday paradox and coincidences.**

- **Views of randomness, streaks in sports, and gambler’s fallacy**

10. Interesting connections help.

- **Birthday paradox and coincidences.**
- **Views of randomness, streaks in sports, and gambler's fallacy**
- **Seeing patterns and superstitions.**

Summary

Elementary probability for many reasons

- We use it to analyze past events
- Conditional probability is hard to intuit
- Hidden assumptions
- “Reasonable” perspectives lead to different conclusions
- Our natural mental heuristics sometimes lead us astray

Ways to help out

- Use the law of large numbers to give guidance on conditional probability
- Visualizations
- Interesting connections