Christoff Raath

What we can learn from Bayesian probability theory

foresight dialogues 2020

"My new working definition for profound is:
obvious only in retrospect."

\- Robin Stewart blog, 2010

If you throw a die ten times and get a six every time, is there a higher change the next one will also be a six?
Or a lower chance?

foresight dialogues 2020

A minor point of which mention will be made again later is that if the coin fell heads in one spin it was convenient to balance it head uppermost on the operator's forefinger when preparing for the next, and vice versa. This technique was rigidly adhered to throughout the sequence of experiments.

Table 1.

2000 successive spins of an ordinary coin.
0 denotes tail.
1 denotes head.

00011101001111101000110101111000100111001000001110
00101010100100001001100010000111010100010000101101
01110100001101001010000011111011111001101100101011
01010000011000111001111101101010110100110110110110
011111100001110110001010010000010100111111011101011
10001100011000110001100110100100001000011101111000
11111110000000001101011010011111011110010010101100
11101101110010000010001100101100111110100111100010
00001001101011101010110011111011001000001101011111
11010001111110010111111001110011111111010000100000
00001111100101010111100001110111001000110100001111
11000101001111111101101110110111011010010110110011
01010011011111111001011100011110111111110000010010 01
01001110111011011011111100000101010101010101001001
11101101110011100000000100110101001100100010000 1100
10111110001001101011011011100110100101010000000 10000
00001011001101011011111000101100101000011100110011
11100101011010000110001001100010010001100100001001
01000011100000011101101111001110011010101101001011
0100000011101101000100011100100111000010100000000010
10010001011000010010100011111101101111010101010000
011000101000000100000000010000001100100011011101010
11011000110111010110010010111000101101101010110110
00001011011101010101000011100111000110100111011101
1000110111000000100111100011101000010100000111110100
00111111111111101010100100110001011110010101000 1111
11000110101010011010010111110000111011110110011001
11111010000011101010111101101011100001000101101001
10011010000101111101111010110011011110000010110010
00110110101111101011100101001101100100011000011000
01010011000110100111010000011001100011101011100001
11010111011110101101101111001111011100011011010000
01011110100111011001001110001111011000011110011111
01101011101110011011100011001111001011101010010010
10100011010111011000111110000010000000100111010 11
1000101110100010111111011100000011111110110000000010
101111111011100010000110000110001111101001110110000
00001111011100011101010001011000110111010001110111
10000010000110100000101000010101000101100010111100
00101110010111010010110010110100011000001110000111

Two schools of thought on probability

John Edmund Kerrich (1903-1985) spent his time as prisoner during the second world war conducting experiments in probability using coins and ping-pong balls. Excerpt from his coin-toss experiments on the left.

"Frequentists"

The *chance* that something can happen
Probabilities are *objective* things that tell us something about the world

You must imagine you can do something an **infinite amount of times**. The proportions of outcomes will then indicate the probabilities.

If I flip a coin an infinite number of times, I will get 50/50 heads and tails.

The problem with frequentism is that many events happen only once.

What does a probability mean in these instances?

SA won the world cup, but what was the probability that we would win? What does it even mean to ask this now, in hindsight?

Two schools of thought on probability

Frequentists think like mathematicians, see the world as *exact* and *objective* and capable of being clearly defined.

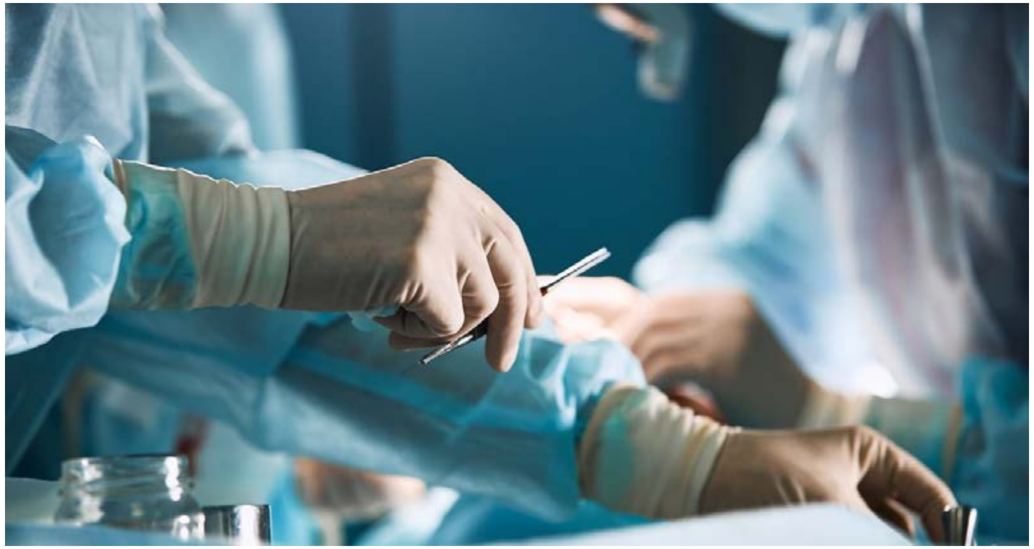
Bayesians are in the other camp…

Probabilities are expressions of your **states of belief** in cases of uncertainty or incomplete information

Subjective.
Personal.
About beliefs.

Bayesians keep on asking:
how well do we know what we think we know?

The outcome of an election



The outcome of a surgical procedure

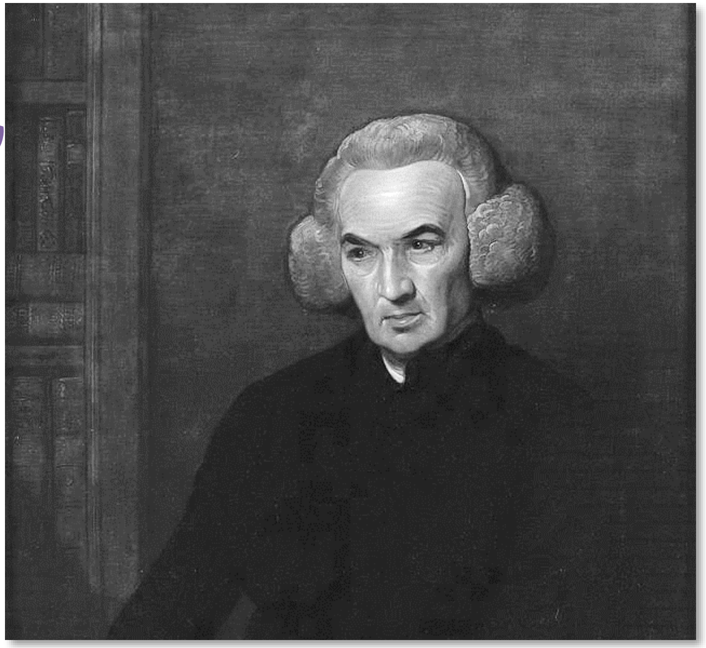
Thomas Bayes

1701 – 1761
Statistician, philosopher and Presbyterian minister

Published only two works in his academic life

…the infamous **Bayes Theorem was not one of them**

Richard Price



1723 – 1791
Discovered Bayes Theorem after Bayes' death in unfinished works

Realised the significance of the theorem.
Believed the theory could assist in proving the existence of God.

Presented to the Royal Society on 23 Dec 1763

foresight dialogues 2020

A familiar application – pathology

Sensitivity (true negative rate TNR) = probability the test will give a negative result if patient is indeed negative

Specificity (true positive rate TPR) = probability the test will give a positive result if patient is indeed negative

Now imagine you tested positive for a really bad but rare disease

Specificity of test is 99%

Prevalence of the disease is 1 out of 1000 (0.1%)

Hypothesis (I have the disease)

Probability of testing positive (E) if you have the disease (H) = **specificity**

Prior probability of having the disease (before test result)

$$P(H|E) = \frac{\quad}{\quad}$$

Evidence (I have tested positive for the disease)

Probability of testing positive

Correct positive result plus false positive result

$$P(H)*P(E|H)+P(H')*P(E|H')$$

A familiar application – pathology

Now imagine you tested positive for a really bad but rare disease

Specificity of test is 99%

Prevalence of the disease is 1 out of 1000 (0.1%)

9.016%

Hypothesis (I have the disease)

Probability of testing positive (E) if you have the disease (H) = **specificity**

Prior probability of having the disease (before test result)

Often the hardest part – and most of the time just a guess

0.99 **0.001**

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

0.01098

Evidence (I have tested positive for the disease)

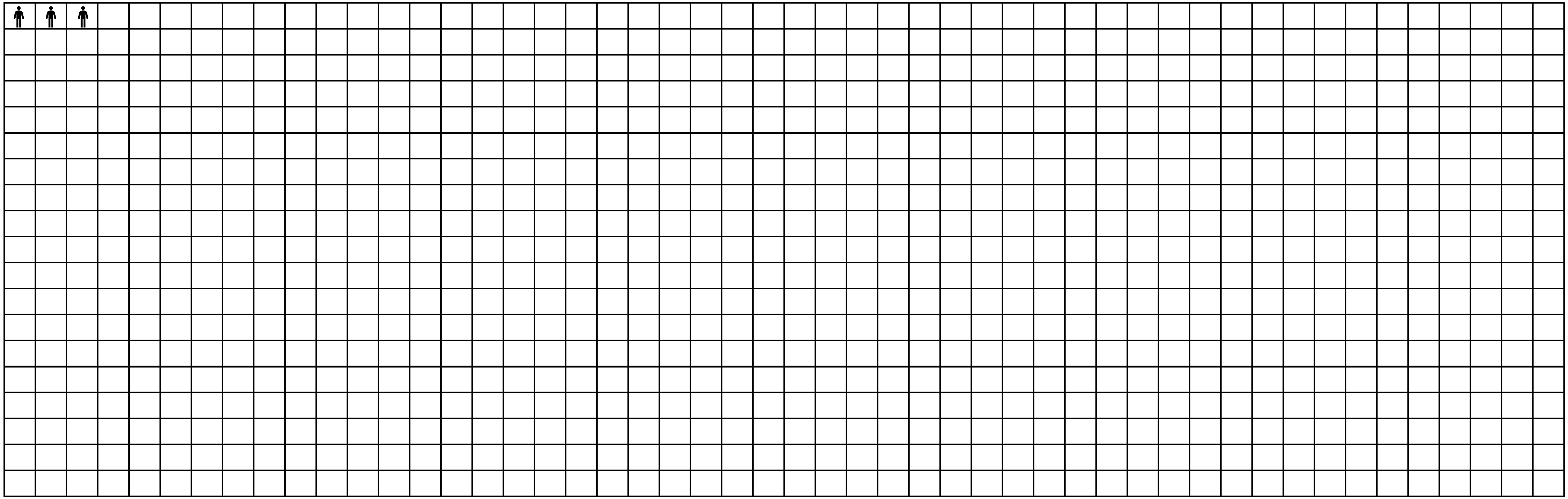
Probability of testing positive

Correct positive result plus false positive result

$$P(H)*P(E|H)+P(H')*P(E|H')$$

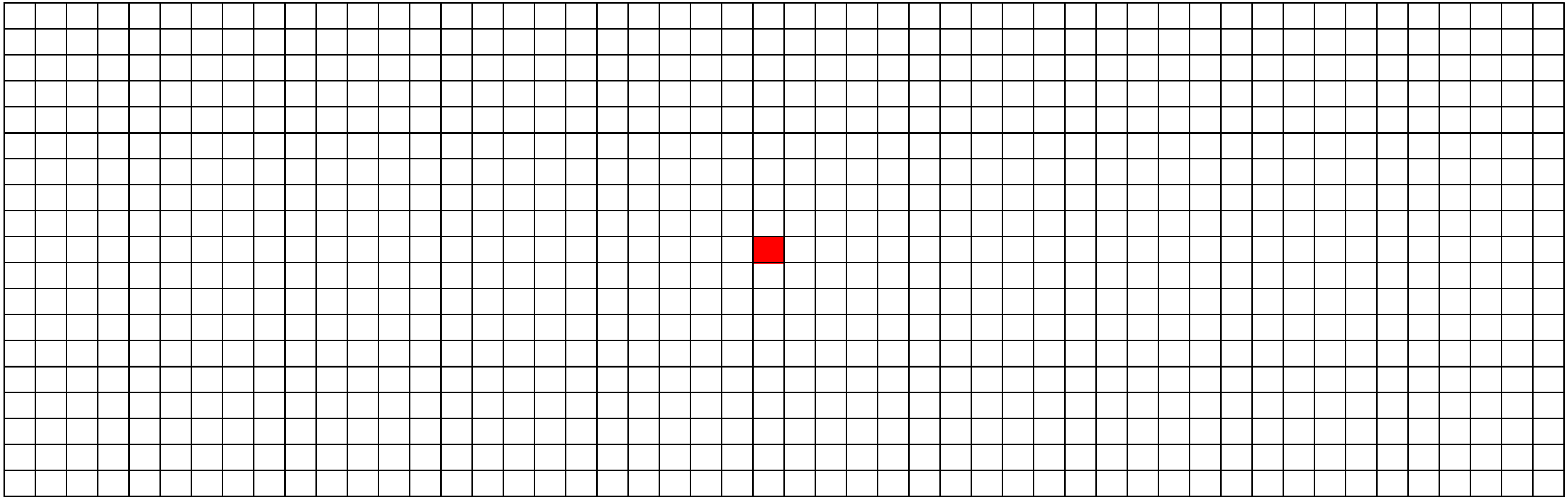
0.001 * 0.99 + 0.999 * 0.01 = 0.01098

Out of a thousand people…



Out of a thousand people…

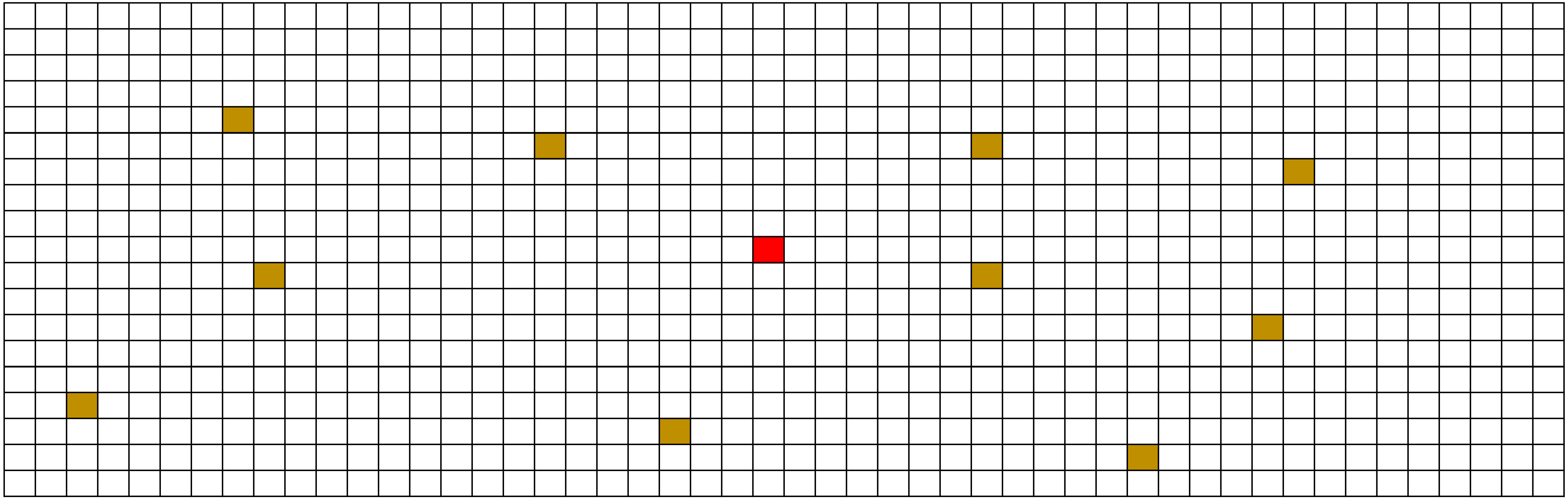
…one is likely to have the disease (0.001 x 1000)



Out of a thousand people…

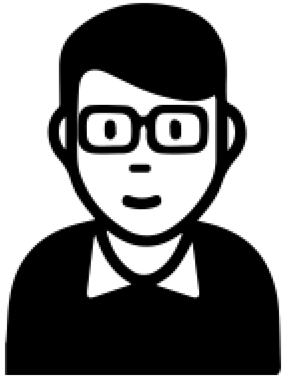
…one is likely to have the disease (0.001 x 1000)

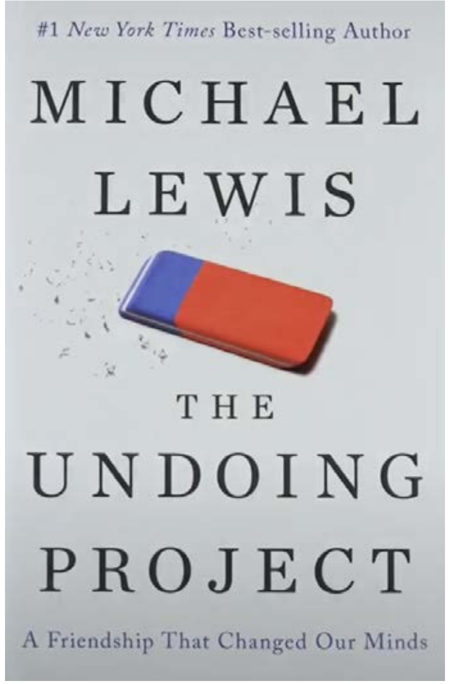
…while ten are likely to have false positives (0.01 x 1000)

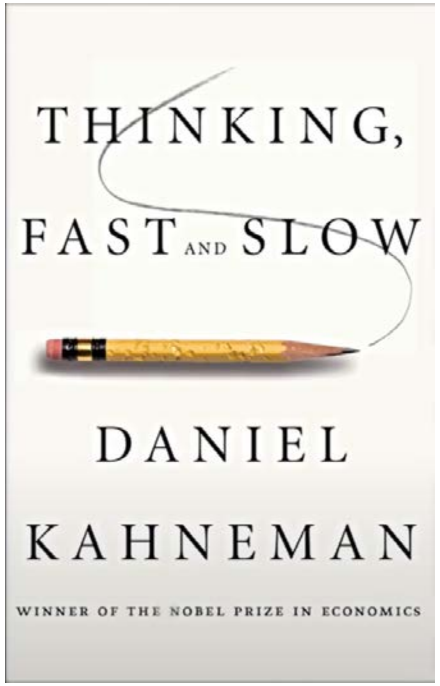


One out of eleven = 9%

"Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."







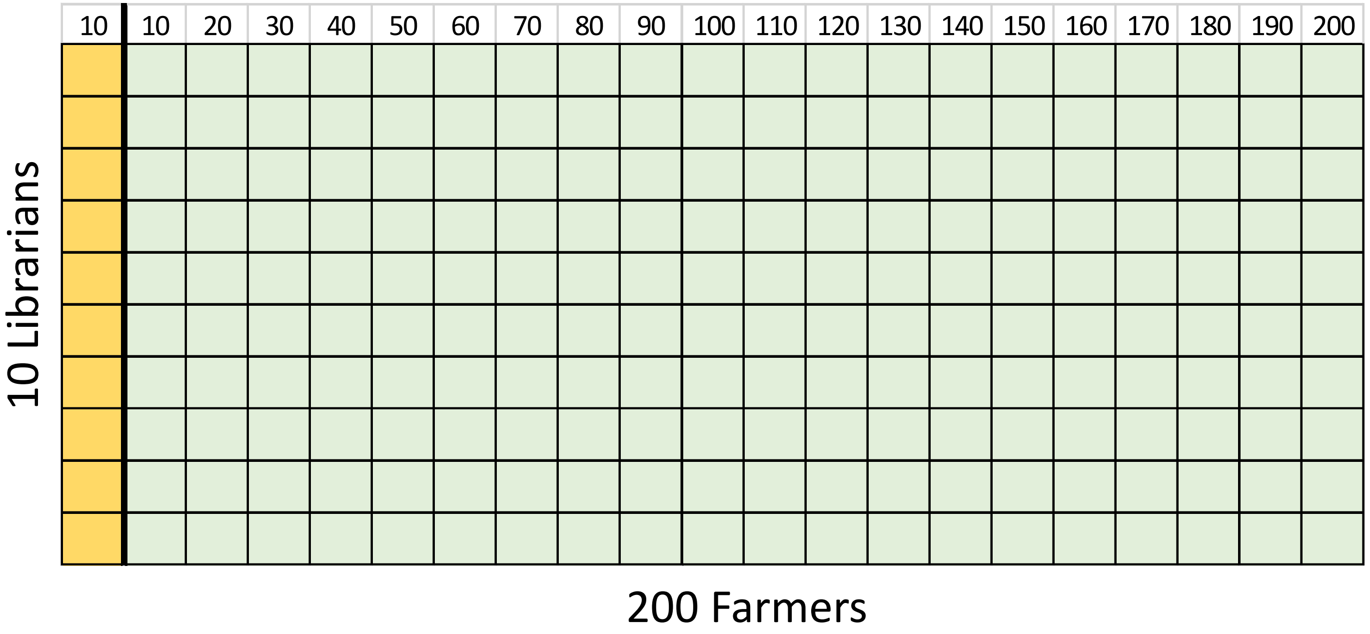
Is Steve more likely to be a librarian or a farmer?

Research by **Kahneman** and **Tversky** found the vast majority of respondents suggest Steve is more likely a **librarian**

…Almost no one appeared to incorporate the ratio of farmers to librarians in their responses

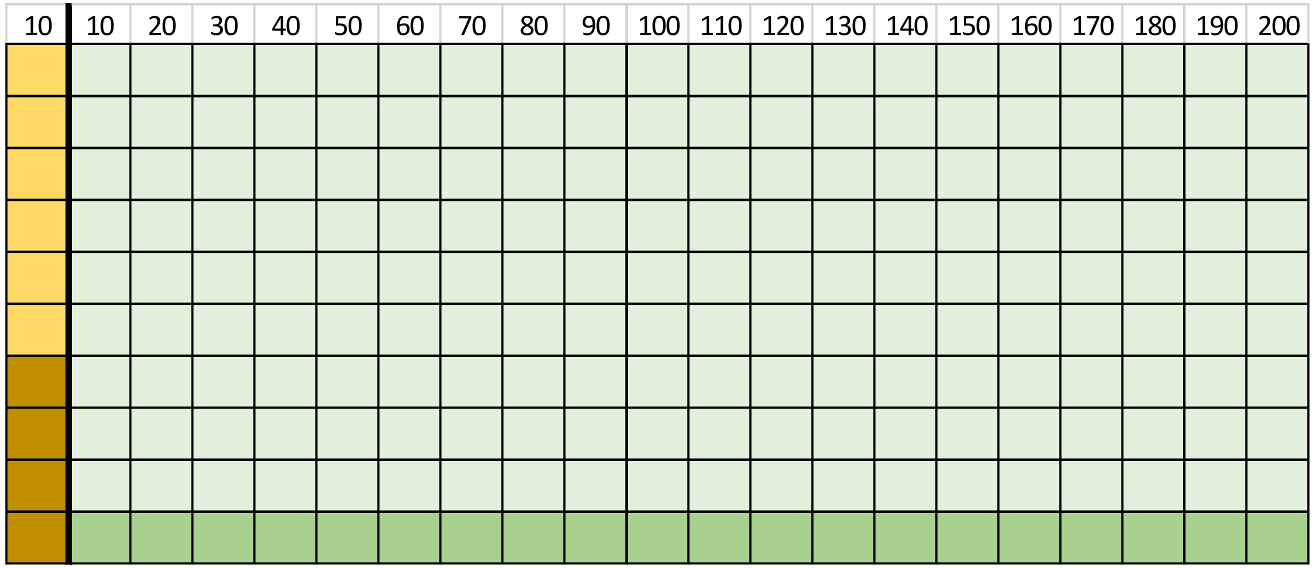
About 1 librarian for every 20 farmers

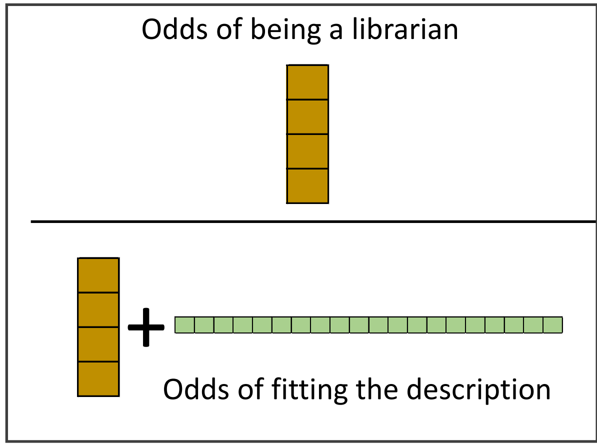
If it is true that there are 20 farmers for every librarian, let us imagine a polulation like this:



$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)} = \frac{40\% * 10}{4+20} = \frac{4}{24} = 16.7\%$$

40% chance a librarian fits this description
=4 librarians





Only a 10% chance a farmer fits this description
=20 farmers

So about 4 librarians and 20 farmers fit this description.

Therefore, probability that a random person fitting this description is a librarian is 4 / 24 = **16.7%**
Despite the fact that a librarian is four times as likely to fit the description

New evidence does not determine your beliefs in a vacuum. **It should *update* your prior beliefs.**

Prior belief: 1 librarian for every 20 farmers
(4.7% chance of being a librarian)

Evidence: Steve is a nerd

Updated belief:
16.7% chance of being a librarian
("posterior" belief)

P(H) = probability a hypothesis is true (before any evidence)

P(E|H) = probability of seeing the evidence if the hypothesis is true

P(E) = probability of seeing the evidence

P(H|E) = probability a hypothesis is true given some evidence

New evidence does not determine your beliefs in a vacuum. **It should *update* your prior beliefs.**

Central to Bayesian theory is that **we can never know the world perfectly**. We must continue to update our beliefs of the world as evidence emerges.

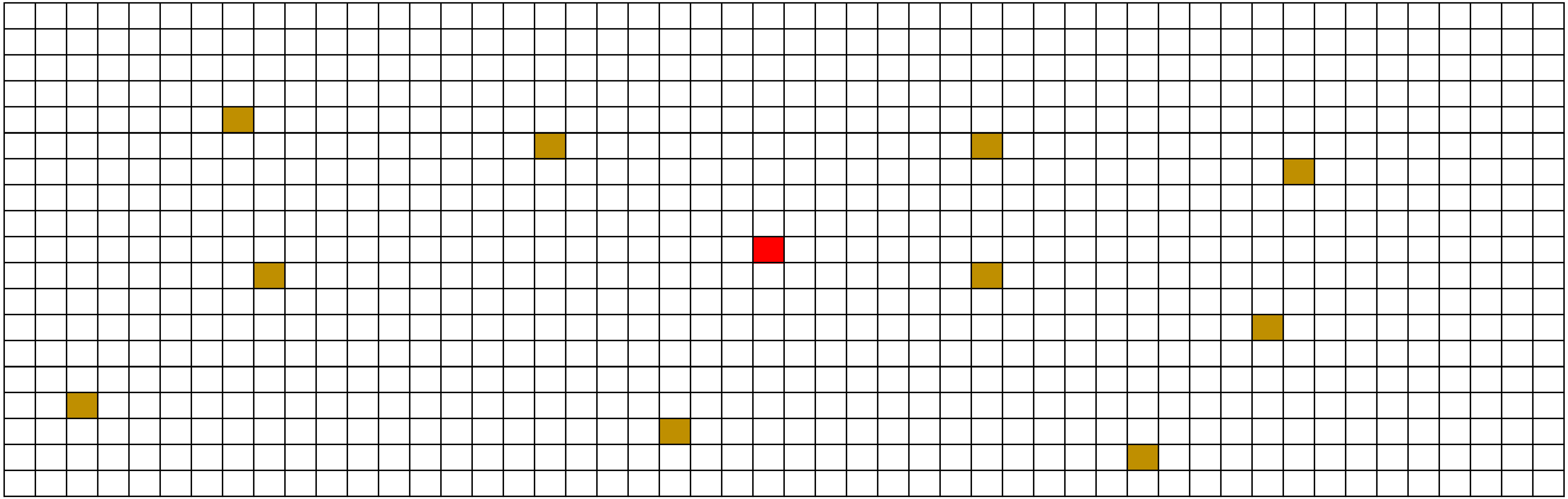
Bayes' theorem was meant to be used iteratively.


Will the sun rise again tomorrow?

Out of a thousand people…

…one is likely to have the disease (0.001 x 1000)

…while ten are likely to have false positives (0.01 x 1000)



One out of eleven = 9%

prior belief

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

posterior belief

posterior belief becomes my new prior

new prior belief

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

new posterior belief

and so on...

0.99 0.1%

9.016% $$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

0.01098

Suppose I redo this test to get a second opinion.
Suppose this result is *also* positive.

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

0.99 9.016%

90.7% $$= \frac{P(E|H) * P(H)}{(H)*P(E|H)+P(H')*P(E|H')}$$

Correct positive
9.016% * 0.99

False positive
90.984% * 0.01

prior belief

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

posterior belief

posterior belief becomes my new prior

new prior belief

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

new posterior belief

and so on...

evidence

0.1%

Prior view

9.016%

Updated belief after one positive test

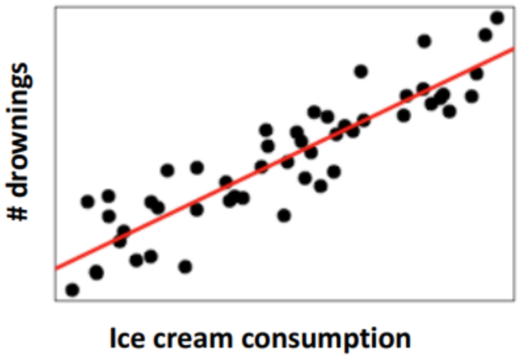
evidence

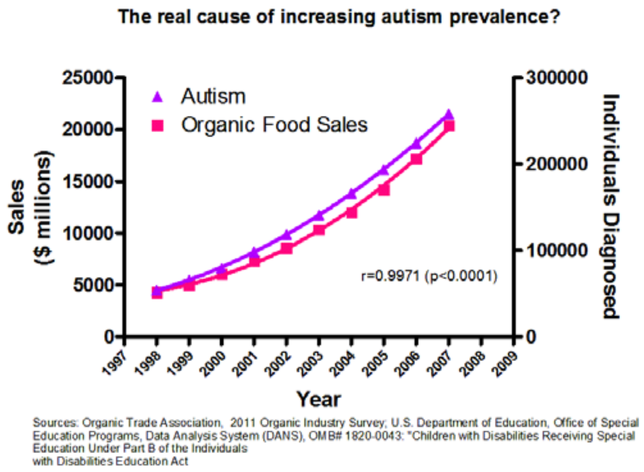
90.7%

Updated belief after two positive tests


UM, SO WHAT IS YOUR POINT?

Correlation vs causation

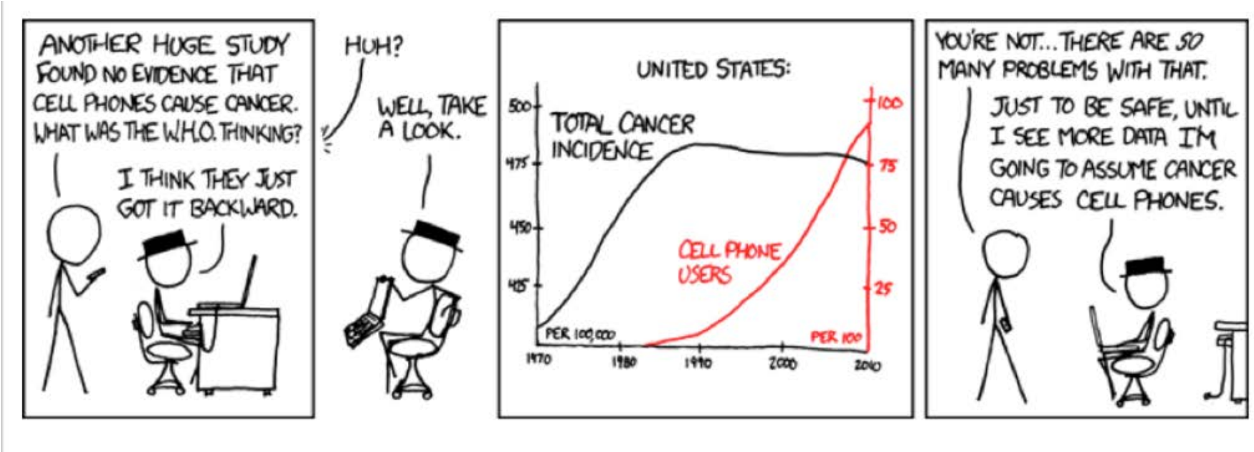


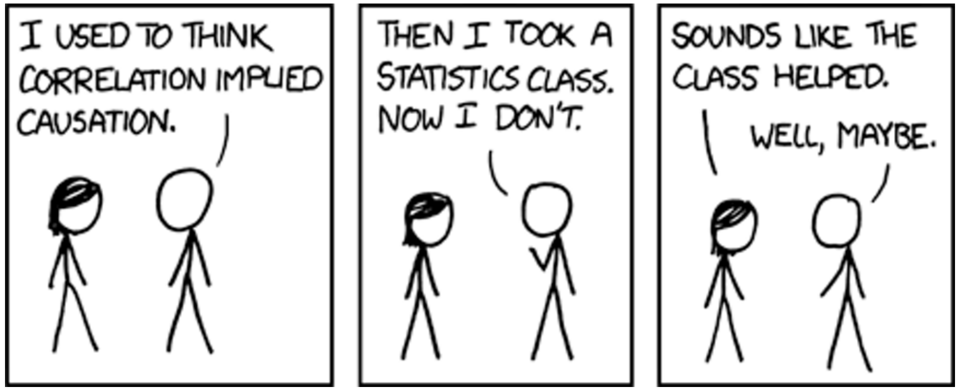


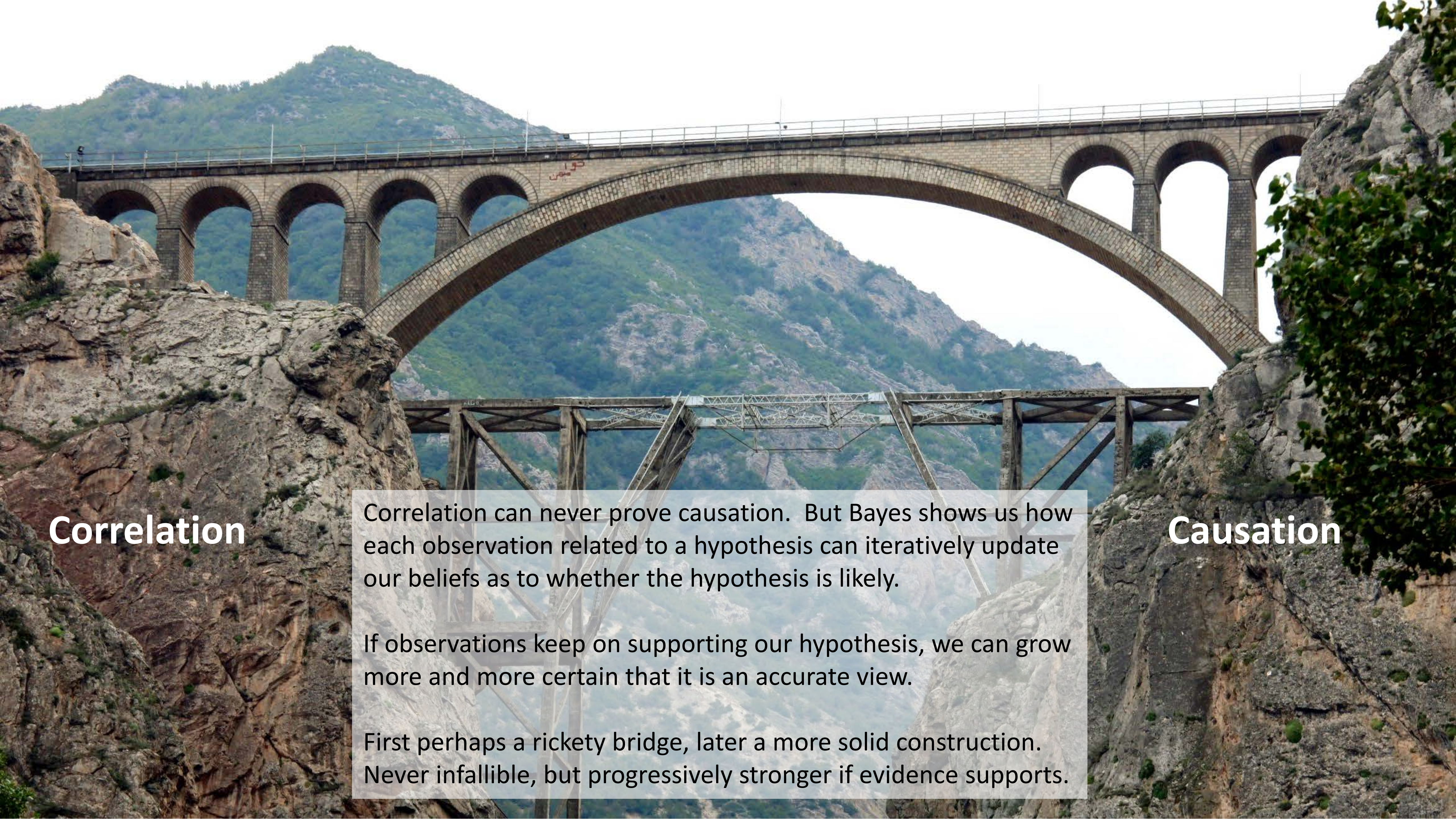
Third (common causal) variable
"obesity and CO_2 levels both increased historically, so if we all pick up weight we could solve global warming"

Reverse causation
"firemen must be the cause of fires because whenever I see a fire, there are firemen around"

Coincidental relationship
Washington Redskins home game vs election results



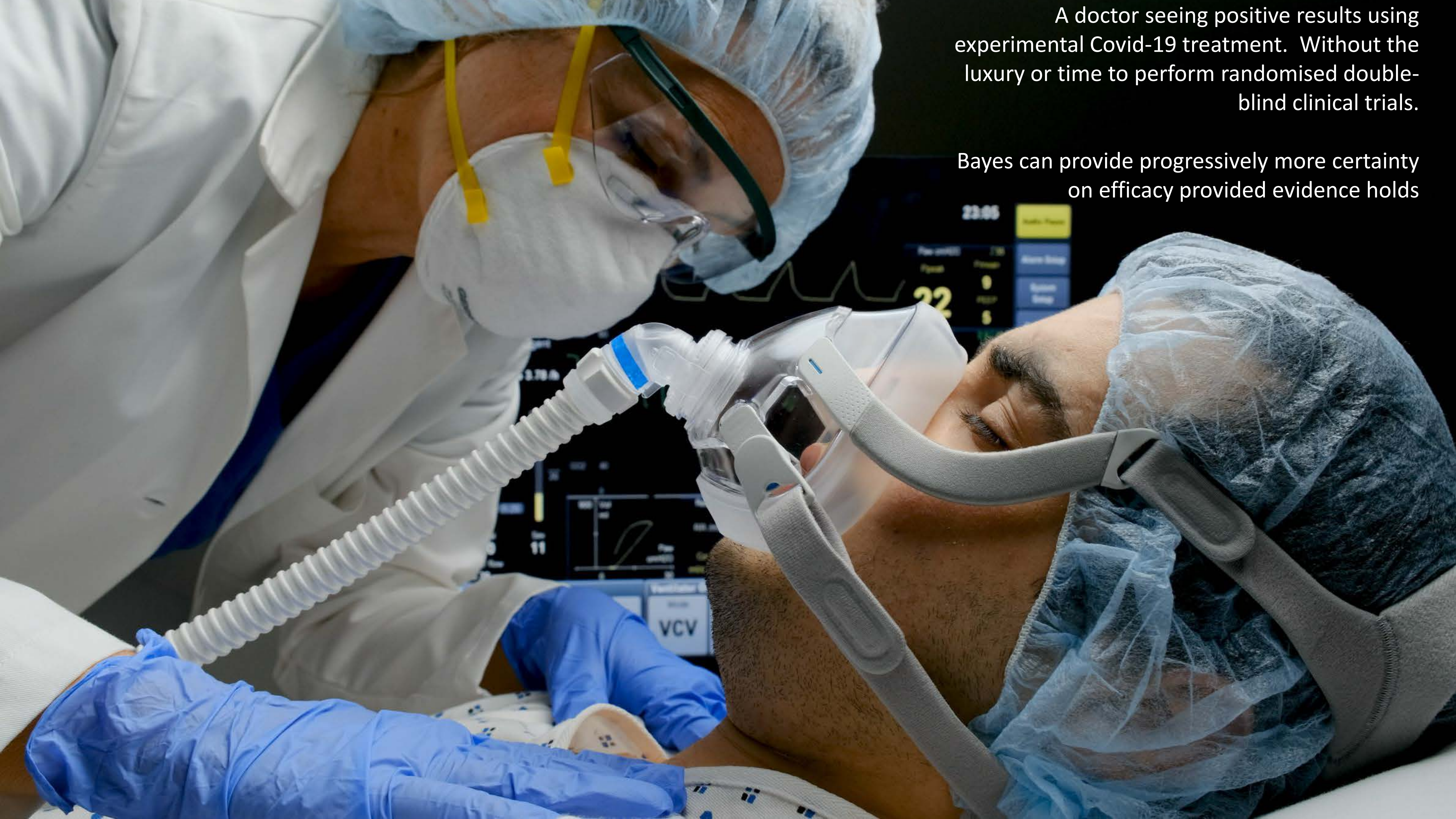




Correlation can never prove causation. But Bayes shows us how each observation related to a hypothesis can iteratively update our beliefs as to whether the hypothesis is likely.

If observations keep on supporting our hypothesis, we can grow more and more certain that it is an accurate view.

First perhaps a rickety bridge, later a more solid construction. Never infallible, but progressively stronger if evidence supports.


A doctor seeing positive results using experimental Covid-19 treatment. Without the luxury or time to perform randomised double-blind clinical trials.
Bayes can provide progressively more certainty on efficacy provided evidence holds

Bayes usefulness in bridging the correlation-causation gap comes with one serious caveat:

We should NEVER be 100% or 0% certain about any belief

0.1% | 0% | 100%

Absolute belief with 0% or 100% certainty

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

9.1% | 0% | 100%

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

92% | 0% | 100%

Absolute beliefs can never change. 0% times anything remains 0%.

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

97% | 0% | 100%

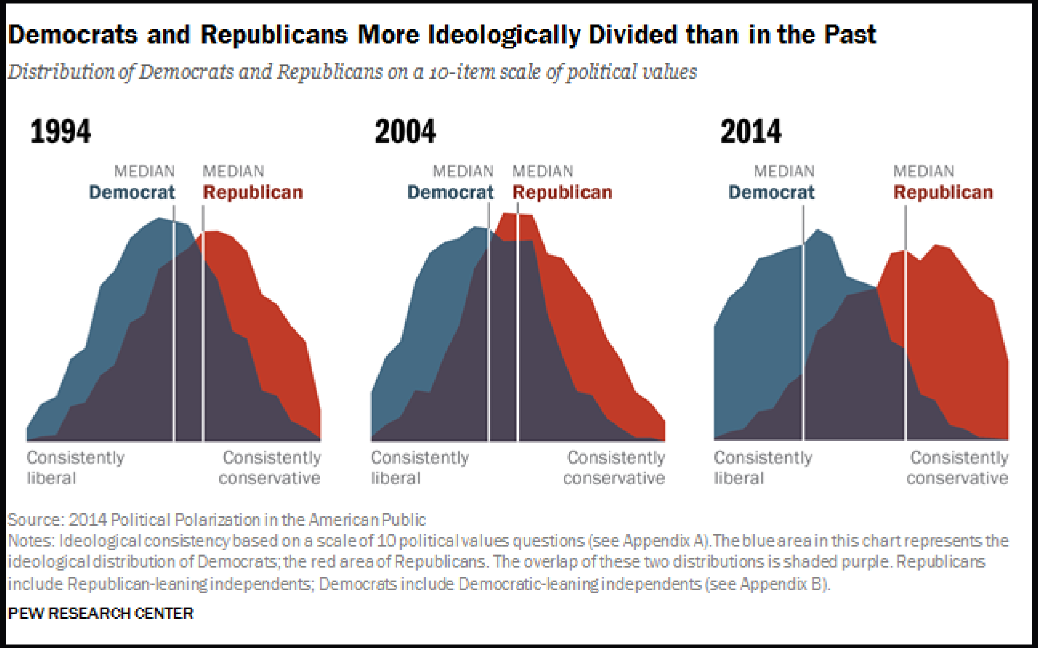
$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Healthy Bayesian process…

Absolute beliefs cannot be updated by considering more evidence

It is vital to expose oneself to contrary evidence and weigh this carefully through the Bayesian process, testing the validity of your own beliefs.

Is the earth round? Yes. But I'm only 99.9% sure of that…




It always seems impossible
until it is done

Summary

Different ways to think about probability

Some probabilities are personal, result in beliefs

- Machine learning
- Medicine
- Science

Evidence should never **determine** your belief, it should merely **UPDATE**

Iterative updating of prior beliefs

Bayes can help bridge the gap between correlation and causation

The danger of holding absolute beliefs

foresight dialogues 2020

foresight

dialogues 2020

Q&A