*Ockham's Razors*: 102-128

- Unifying theories seem simpler than disunifying theories
- For observations $O_1$ & $O_2$, unifying theory U seems simpler than disunifying theory $T_1$&$T_2$
- Is the difference in simplicity epistemically relevant?

Is $Pr(U|O_1 \& O_2) > Pr(T_1\&T_2|O_1 \& O_2)$ ?

Is Ockham's razor related to probabilistic likelihood?

Common causes as unifying hypotheses


Two students turn in identical papers. Consider the following common-cause (CC) and separate-cause (SC) hypotheses (courtesy of Wesley Salmon):

- (CC) The two students searched the Internet together and found a file that they agreed to plagiarize.
- (SC) The two students worked separately and independently.

Intuitively,

Pr(the papers match|CC) >> Pr(the papers match|SC)

> "This suggests that Ockham's razor may sometimes have a likelihood justification."

Hans Reichenbach:

*The Principle of the Common Cause*: If an improbable coincidence has occurred,

there ~~must exist~~ *probably exists* a common cause.

Singular events:

- Both lamps in a room go out suddenly
- Several actors in a stage play fall ill showing symptoms of food poisoning

Repeated events:

- Two geysers which are not far apart spout irregularly, but throw up their columns of water always at the same time.
- Barometers always show the same indication if they are not far apart

*Associated geysers*

Suppose each geyser spouts at a frequency of one hour per week (1/168)

But both geysers always spout at the same time

freq(geyser 1 spouts & geyser 2 spouts) > freq(geyser 1 spouts)freq(geyser 2 spouts)

i.e.

1/168 > (1/168)(1/168)

The geysers spout simultaneously more often than one would expect if they were probabilistically independent

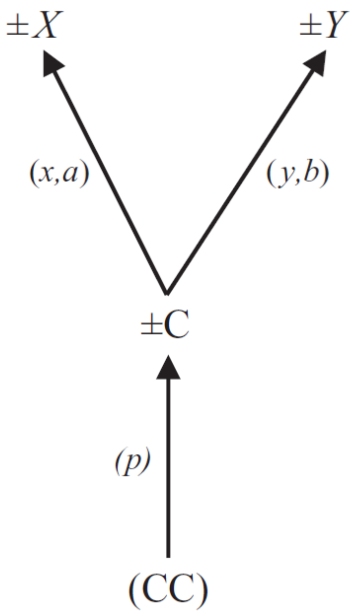



Figure 2.3

$Pr_{CC}(X|C) = x$

$Pr_{CC}(X|notC) = a$

$Pr_{CC}(Y|C) = y$

$Pr_{CC}(Y|notC) = b$

$Pr_{CC}(C) = p$

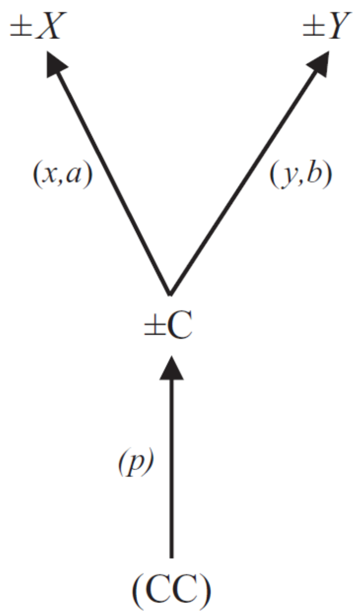



Figure 2.3

$Pr_{CC}(X|C) = x$

$Pr_{CC}(X|notC) = a$

$Pr_{CC}(Y|C) = y$

$Pr_{CC}(Y|notC) = b$

$Pr_{CC}(C) = p$

(Screening-off, nonzero, positive correlation of cause and effect)

*With the right assumptions in place,*

Effects of a common cause will be correlated.

If the geysers are correlated, does that mean they have a common cause?

Beware the fallacy of affirming the consequent!

But isn't the correlation *evidence* for a common cause?

A fly in the ointment:

we only observe *association*, not correlation

(Because we observe frequencies, not probabilities)

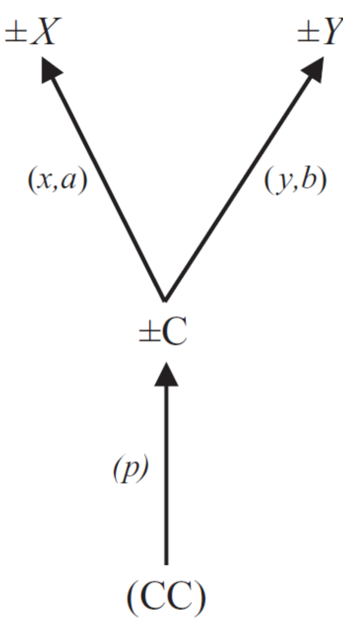



Figure 2.3

vs.

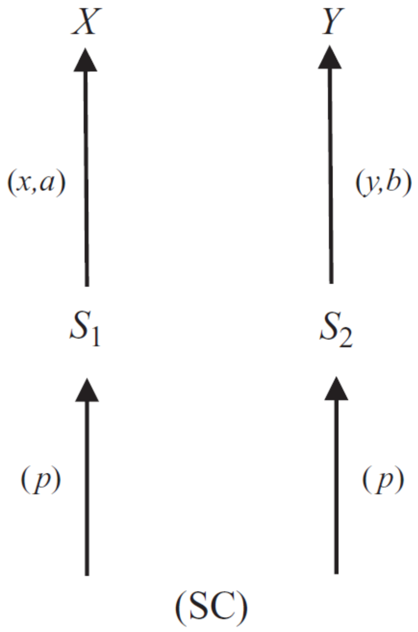



Figure 2.4

Bayesian analysis tells us that an association is *probable* under the common-cause hypothesis, not certain.

So, we need to show that it is *more* probable than a competing, separate-cause hypothesis

*With the right assumptions in place*, we can show that

Pr(X and Y are positively associated|CC) > Pr(X and Y are positively associated|SC)

i.e. the positive association evidentially favors CC over SC

Both models have assumed the processes generating data for the geyser example are *i.i.d.*:

*Independent and identically distributed.*

- Probabilistically independent
- The same probability distribution applies to each time interval

The plagiarism example is *not* i.i.d.:

- A word's probability of appearing at a given place in a paper depends on the place.
  ("Or" is less likely to be the first word than it is to be the 30th word)
- Words that appear once are more likely to appear again.
  (*Chekov's gun*)

Another non-i.i.d. example:

Do two species have a common ancestor?

We score each species for 8 dichotomous (+/-) traits $T_1 - T_8$:

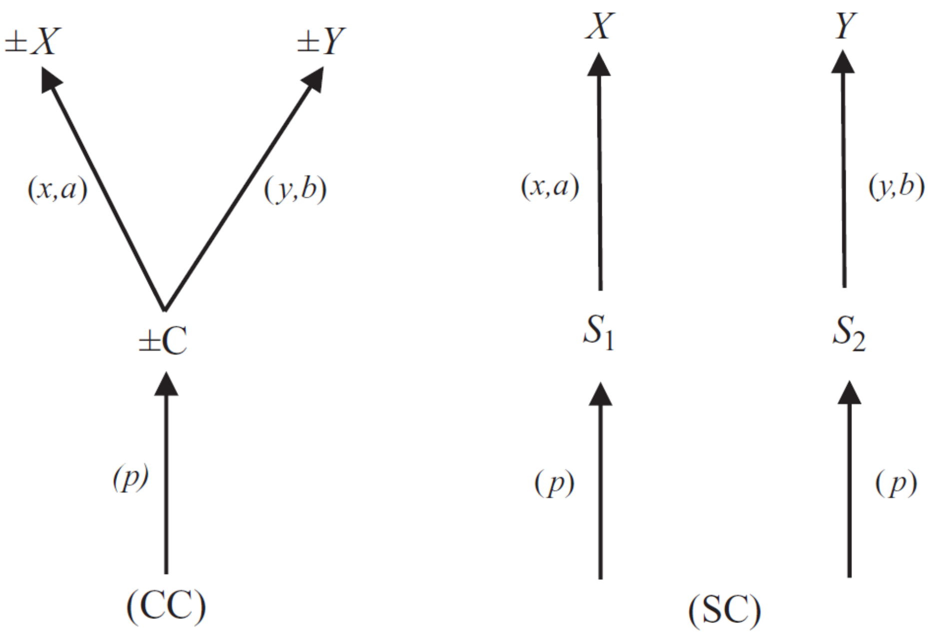
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| species 1 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |
| species 2 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |

Since both species are positive for all 8 traits, the association is **zero**:

freq(species 1 is + and species 2 is +) = 8/8 = freq(species 1 is +)freq(species 2 is +) = (8/8)(8/8)

We can get around this by treating each trait individually, so we have 8 observations of the form:

Species 1 has trait $T_i$ and Species 2 has trait $T_i$.




CA SA

$Pr_{CA}(X|C) = x = Pr_{SC}(X|S_1)$

$Pr_{CA}(X|notC) = a = Pr_{SC}(X|notS_1)$

$Pr_{CA}(Y|C) = y = Pr_{SC}(Y|S_2)$

$Pr_{CA}(Y|notC) = b = Pr_{SC}(Y|notS_2)$

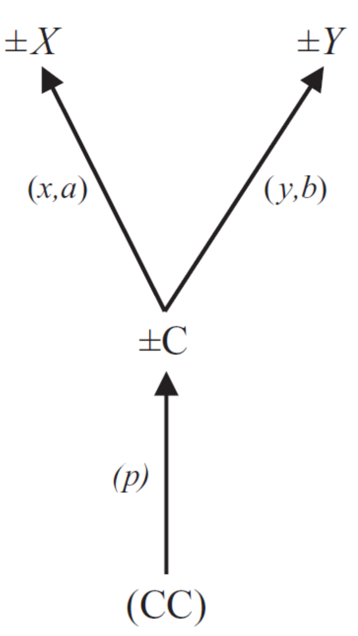
$Pr_{CA}(C) = p = Pr_{SC}(S_1) = Pr_{SC}(S_2)$

The common ancestry hypothesis says that there exists a most recent common ancestor of species X and Y and that it has some state or other for each of the 8 traits.

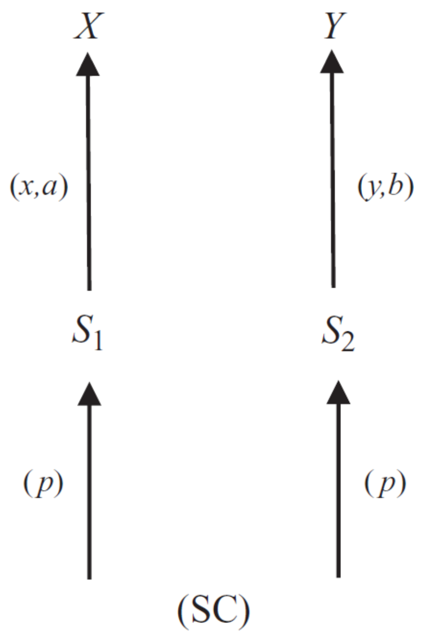
To make our old models work for the species example, we need to assume ***cross-model homogeneity***:

*The parameters used in the common cause model have the same values as the counterpart parameters that are used in the separate cause model.*

Parameters are *x, a, y, b, p*




CA




SA

$\Pr_{CA}(X|C) = x = \Pr_{SC}(X|S_1)$

$\Pr_{CA}(X|\mathit{not}C) = a = \Pr_{SC}(X|\mathit{not}S_1)$

$\Pr_{CA}(Y|C) = y = \Pr_{SC}(Y|S_2)$

$\Pr_{CA}(Y|\mathit{not}C) = b = \Pr_{SC}(Y|\mathit{not}S_2)$

$\Pr_{CA}(C) = p = \Pr_{SC}(S_1) = \Pr_{SC}(S_2)$

Assuming cross-model homogeneity (as well as the original assumptions from earlier), we can prove:

Pr(the two species have + $T_i$|CA) > Pr(the two species have +$T_i$|SA)

i.e.

$xyp + ab(1 - p) > [xp + a(1 - p)]\ [yp + b(1 - p)]$

With this result we can now say, *qualitatively*, there are 8 observations in favor of common ancestry and zero against.

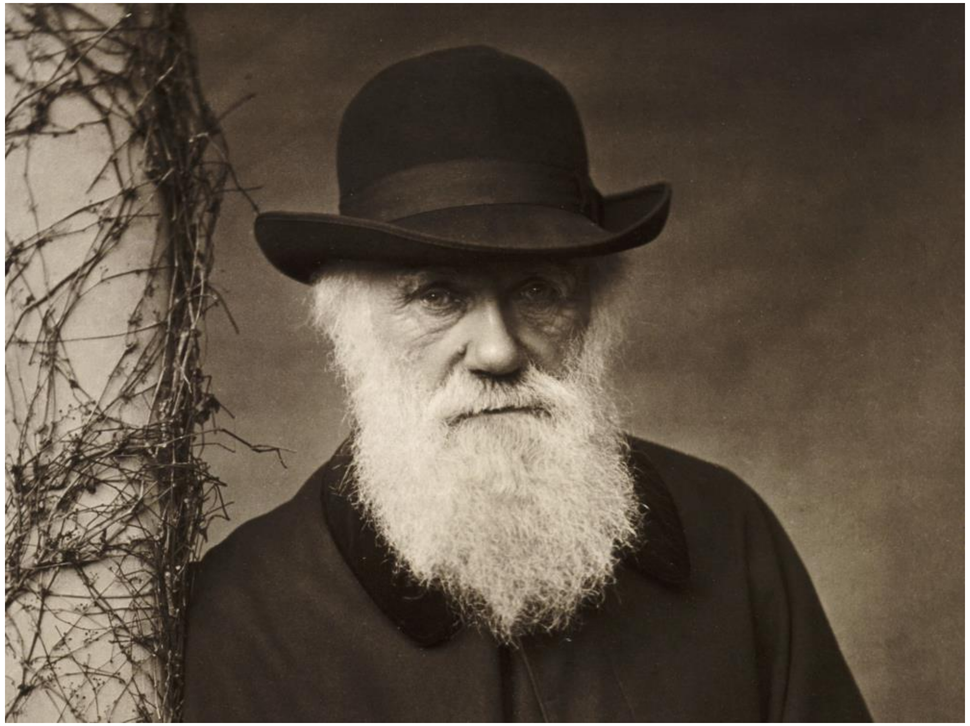
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| species 1 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |
| species 2 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |

But not all shared traits favor common ancestry equally.

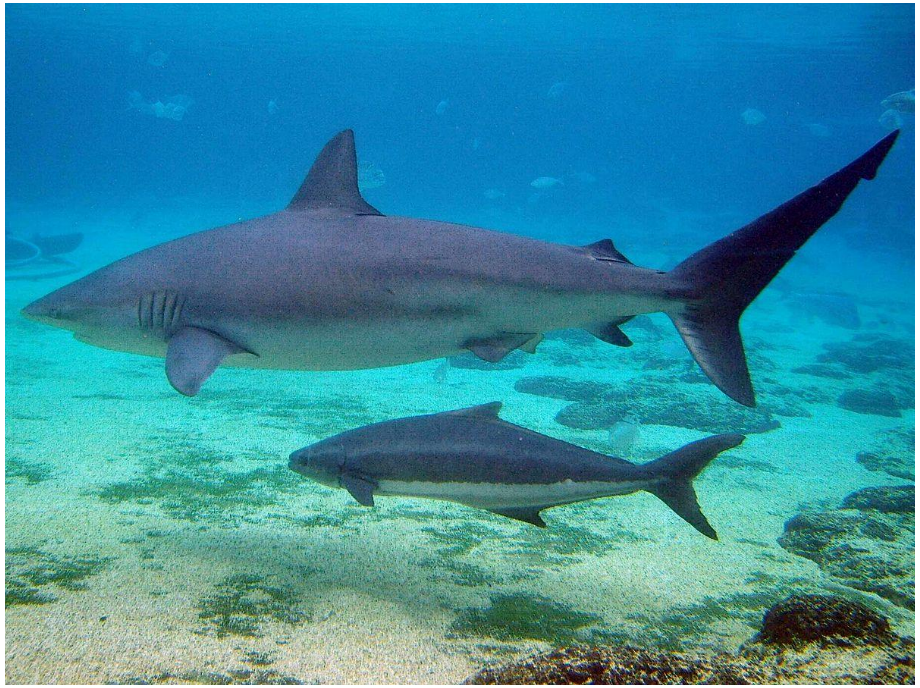
With this result we can now say, *qualitatively*, there are 8 observations in favor of common ancestry and zero against.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| species 1 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |
| species 2 | $+T_1$ | $+T_2$ | $+T_3$ | $+T_4$ | $+T_5$ | $+T_6$ | $+T_7$ | $+T_8$ |

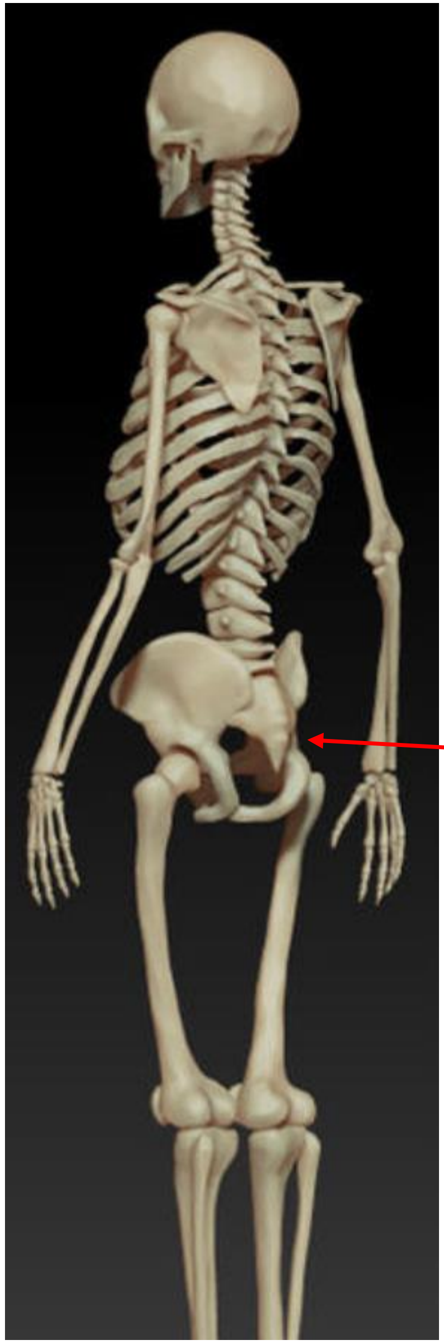
But not all shared traits favor common ancestry equally.



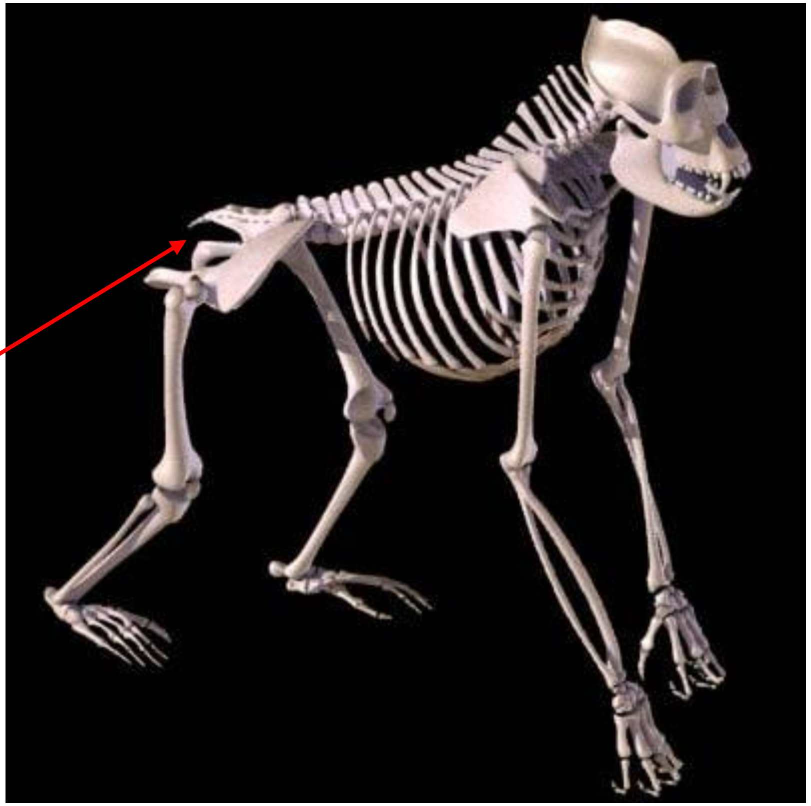
"…adaptive characters, although of the utmost importance to the welfare of the being, are almost valueless to the systematist. For animals belonging to two most distinct lines of descent, may readily become adapted to similar conditions, and thus assume a close external resemblance; but such resemblances will not reveal – will rather tend to conceal their blood-relationship to their proper lines of descent." (Darwin 1859, p. 427)

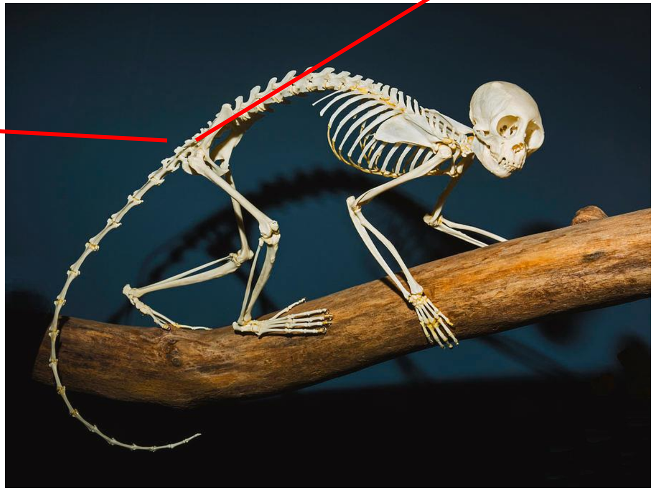


Shared adaptive traits are *not* evidence of common ancestry



Shared useless traits *are* evidence of common ancestry



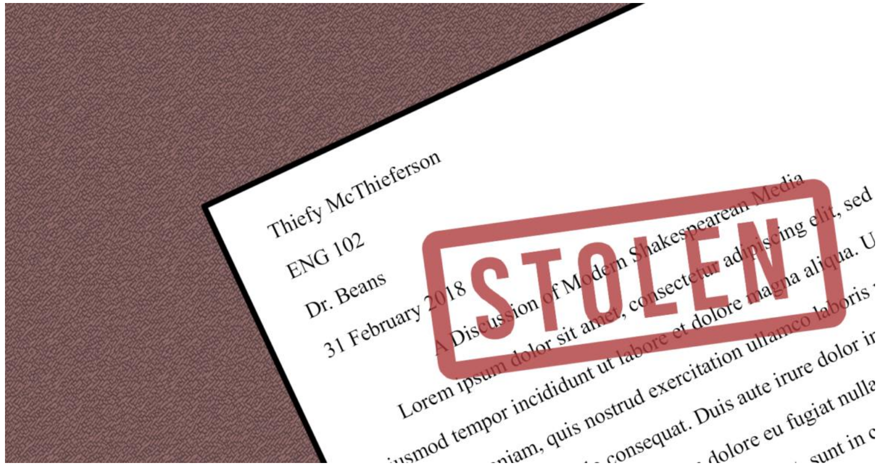


*Darwin's Principle:* adaptive similarities provide negligible evidence for common ancestry whereas neutral or deleterious similarities provide stronger evidence.

Expressed using likelihood ratios:

$$\frac{\Pr(\text{X and Y have trait T} \mid \text{CA})}{\Pr(\text{X and Y have trait T} \mid \text{SA})} \approx 1 \text{ when T is adaptive for both X and Y}$$

$$\frac{\Pr(\text{X and Y have trait T} \mid \text{CA})}{\Pr(\text{X and Y have trait T} \mid \text{SA})} \gg 1 \text{ when T is not adaptive for both X and Y}$$




Darwin’s Principle can be applied to the plagiarism example:

- That both essays are divided into paragraphs is scarcely evidence of plagiarism.
- That both essays misspell the same word in the same way is strong evidence of plagiarism.

**NB**: the assumptions backgrounding these examples are substantive.

With different assumptions, similarities could be evidence *against* a common cause.

“Common cause explanations are, in an intuitive sense, more parsimonious because they postulate fewer causes, but whether parsimony is epistemically relevant, and how it is relevant, depend on the background assumptions that are in place.”

**On Similarity**

What counts as a similarity?

When do two species have the "same" trait?

What counts as a trait?

These problems can be overcome by constructing models with realistic assumptions.

Sober argues *similarities and differences do not have intrinsic epistemic significance*;

All depends on the background assumptions in the model.

Return to Reichenbach's principle:

*The Principle of the Common Cause*: If an improbable coincidence has occurred, there probably exists a common cause.

Sober casts doubt on this as a general principle, appealing to spurious correlations.

**Bayesian Ockham's razor**

The CC/SC models compared so far are "not *very* Bayesian."

- No prior probabilities
- No probability distributions assigned

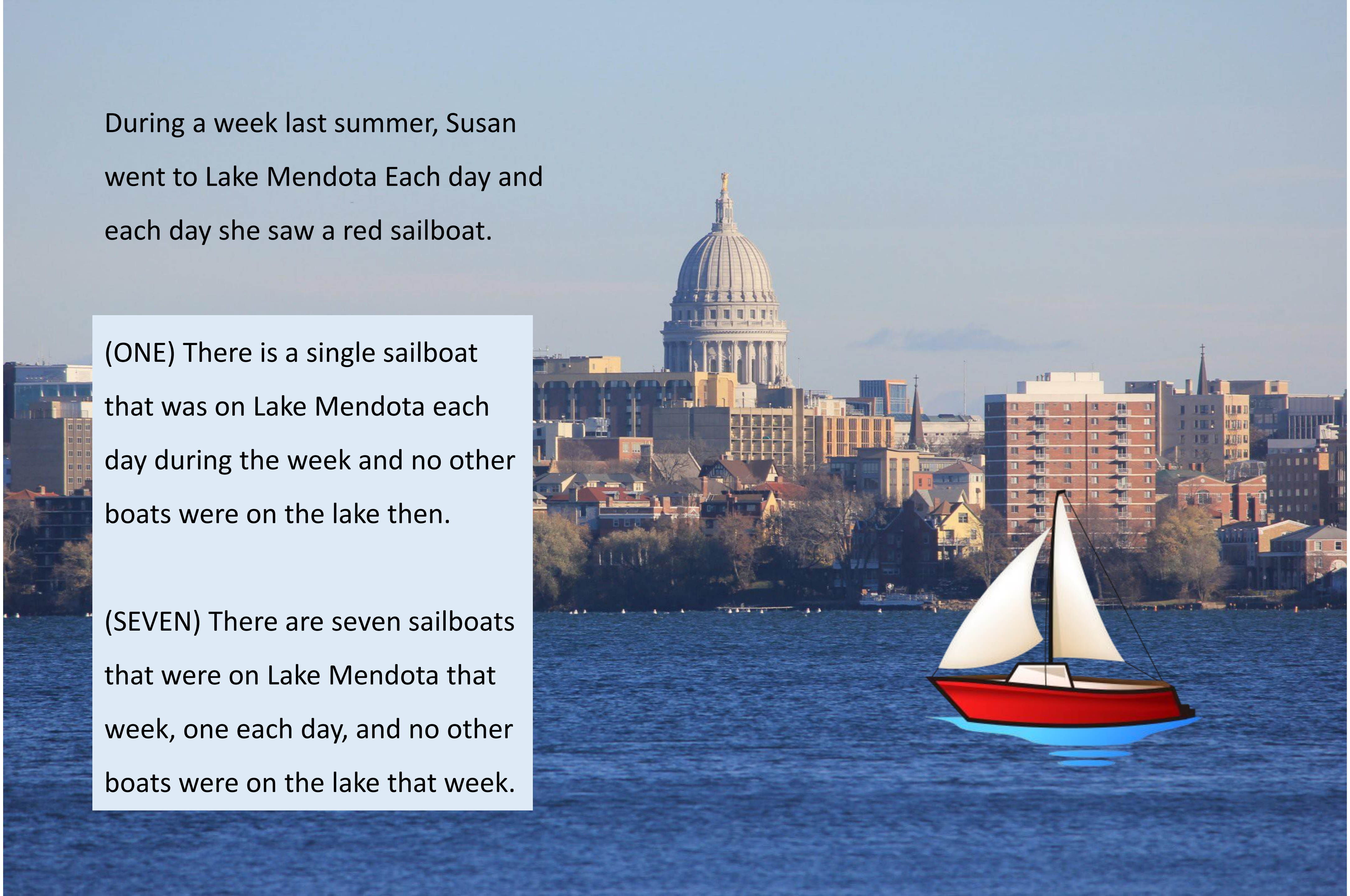
**Bayesian Ockham's razor**

The CC/SC models compared so far are "not *very* Bayesian."

- No prior probabilities
- No probability distributions assigned

"Bayesian Ockham's razor"
assigns probability distributions
to parameters

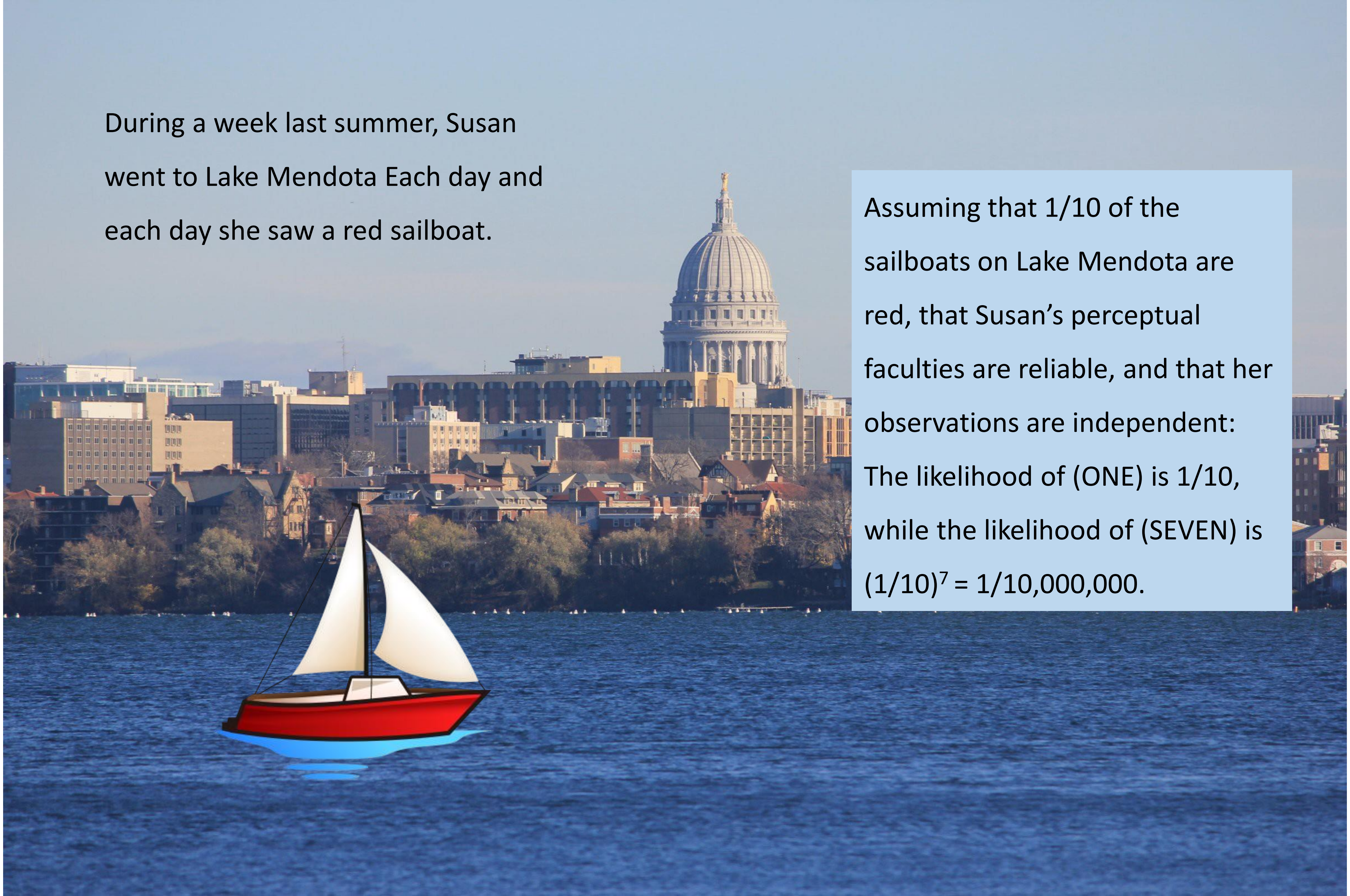
During a week last summer, Susan went to Lake Mendota Each day and each day she saw a red sailboat.



During a week last summer, Susan went to Lake Mendota Each day and each day she saw a red sailboat.

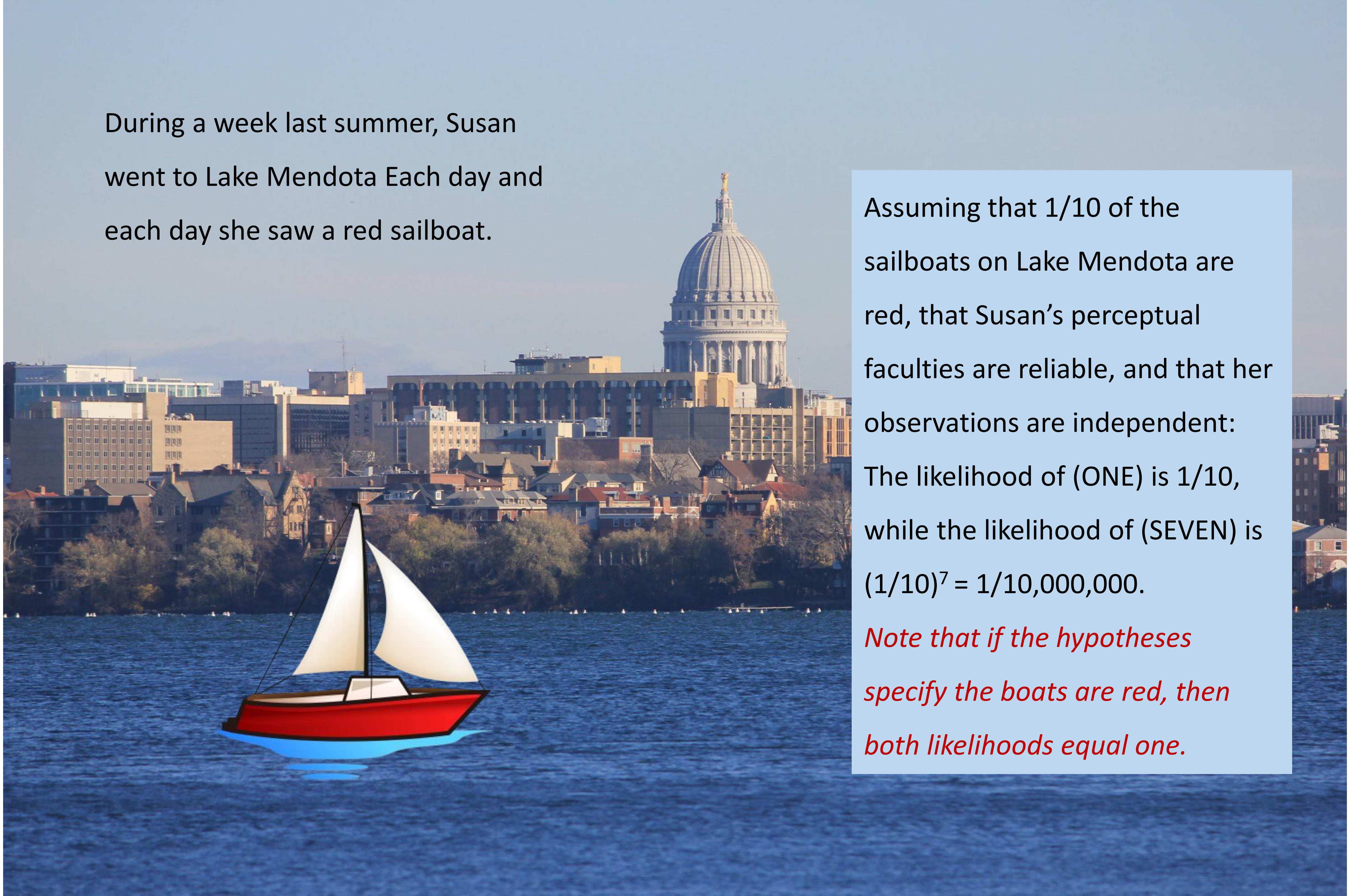
(ONE) There is a single sailboat that was on Lake Mendota each day during the week and no other boats were on the lake then.

(SEVEN) There are seven sailboats that were on Lake Mendota that week, one each day, and no other boats were on the lake that week.



During a week last summer, Susan went to Lake Mendota Each day and each day she saw a red sailboat.

Assuming that 1/10 of the sailboats on Lake Mendota are red, that Susan's perceptual faculties are reliable, and that her observations are independent: The likelihood of (ONE) is 1/10, while the likelihood of (SEVEN) is $(1/10)^7$ = 1/10,000,000.



During a week last summer, Susan went to Lake Mendota Each day and each day she saw a red sailboat.

Assuming that 1/10 of the sailboats on Lake Mendota are red, that Susan's perceptual faculties are reliable, and that her observations are independent: The likelihood of (ONE) is 1/10, while the likelihood of (SEVEN) is $(1/10)^7$ = 1/10,000,000.

*Note that if the hypotheses specify the boats are red, then both likelihoods equal one.*

| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

*Bayesian* Ockham's razor

(ONE-MT) There is a single sailboat that was on Lake Mendota on both Monday and Tuesday and, on each day, it was the only boat on the lake; there is a color *c* that that sailboat has.

(TWO-MT) There was a single sailboat on Lake Mendota on Monday and a different single sailboat out there on Tuesday; there is a color *c*M that Monday's sailboat has and a color *c*T that Tuesday's sailboat has.

| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

*Bayesian* Ockham's razor

One adjustable parameter

Two adjustable parameters

(ONE-MT) There is a single sailboat that was on Lake Mendota on both Monday and Tuesday and, on each day, it was the only boat on the lake; there is a color *c* that that sailboat has.

(TWO-MT) There was a single sailboat on Lake Mendota on Monday and a different single sailboat out there on Tuesday; there is a color *c*M that Monday's sailboat has and a color *c*T that Tuesday's sailboat has.

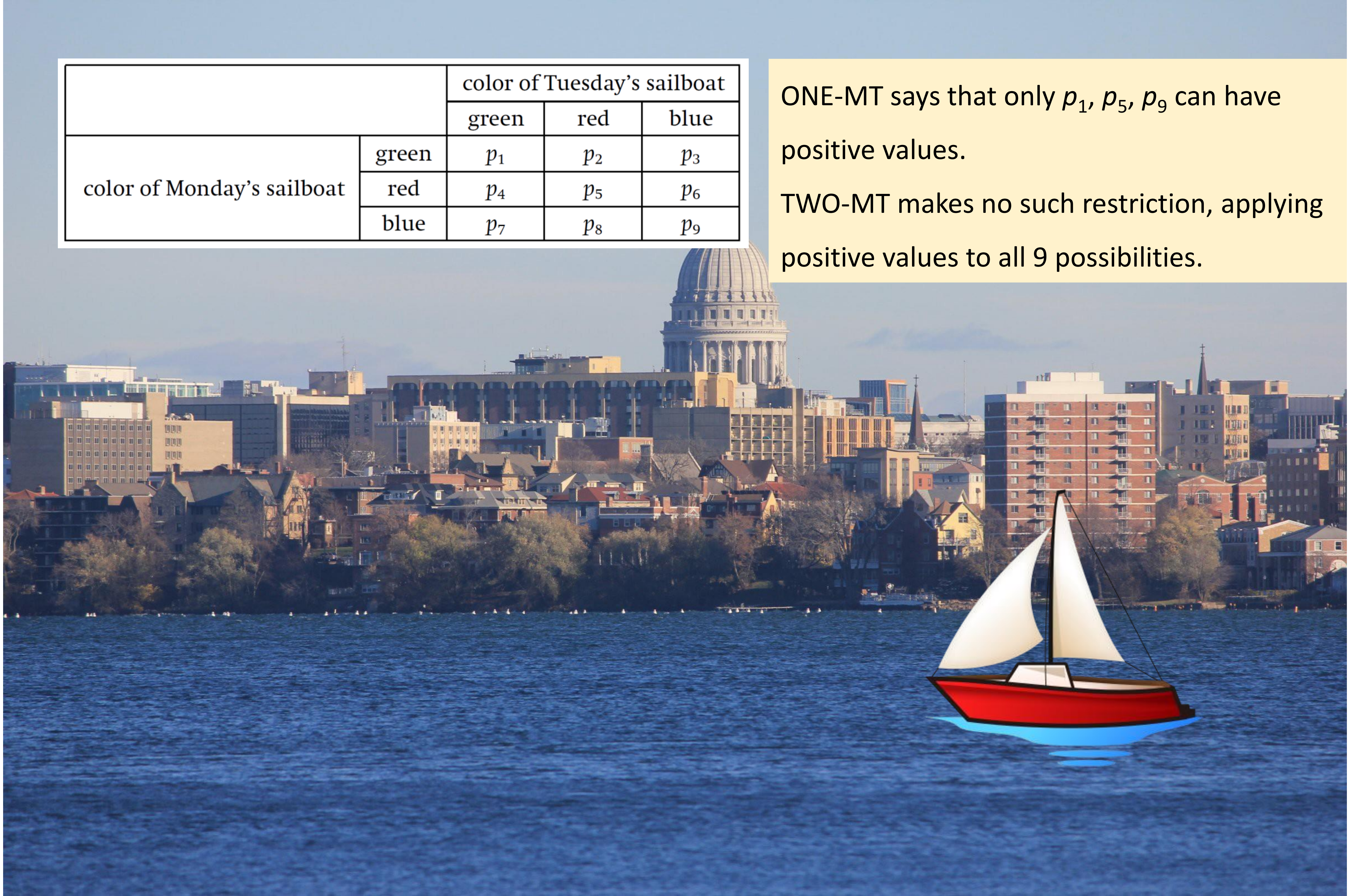
| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | 0 | 0 |
| | red | 0 | $p_5$ | 0 |
| | blue | 0 | 0 | $p_9$ |

ONE-MT says that only $p_1$, $p_5$, $p_9$ can have positive values.

| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

ONE-MT says that only $p_1$, $p_5$, $p_9$ can have positive values.

TWO-MT makes no such restriction, applying positive values to all 9 possibilities.

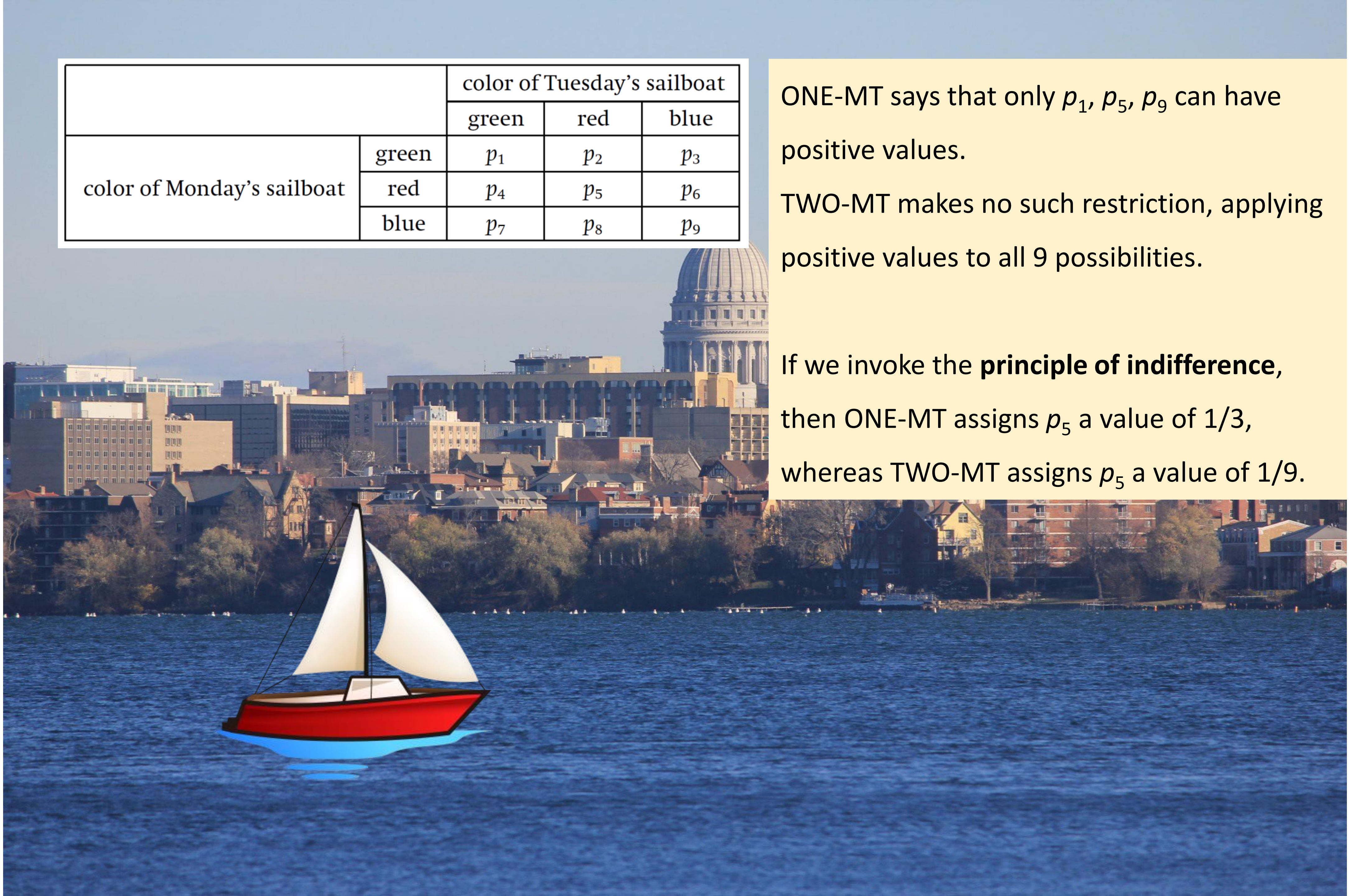


| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

ONE-MT says that only $p_1$, $p_5$, $p_9$ can have positive values.

TWO-MT makes no such restriction, applying positive values to all 9 possibilities.

If we invoke the **principle of indifference**, then ONE-MT assigns $p_5$ a value of 1/3, whereas TWO-MT assigns $p_5$ a value of 1/9.



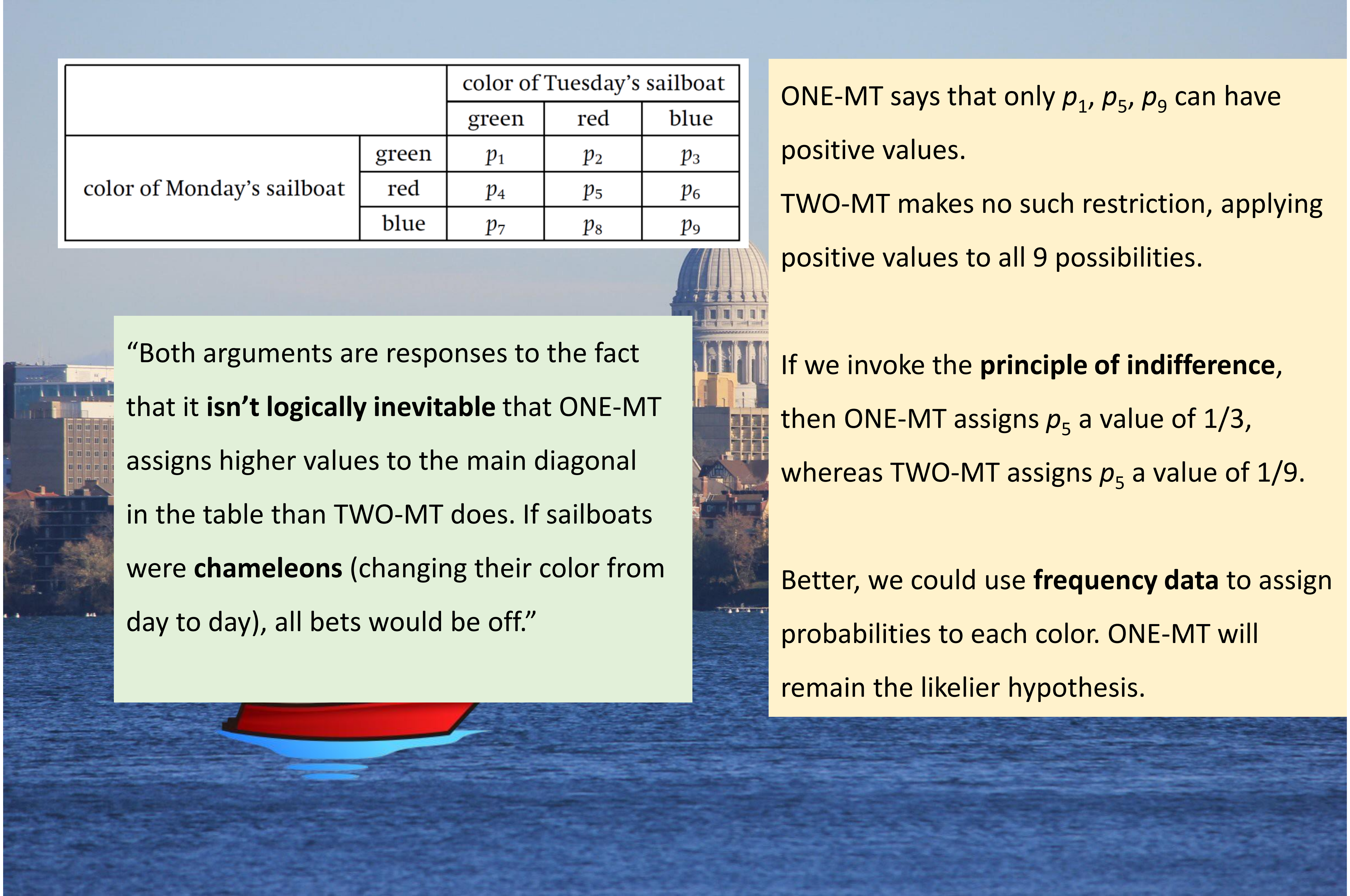
| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

ONE-MT says that only $p_1$, $p_5$, $p_9$ can have positive values.

TWO-MT makes no such restriction, applying positive values to all 9 possibilities.

If we invoke the **principle of indifference**, then ONE-MT assigns $p_5$ a value of 1/3, whereas TWO-MT assigns $p_5$ a value of 1/9.

Better, we could use **frequency data** to assign probabilities to each color. ONE-MT will remain the likelier hypothesis.



| | | color of Tuesday's sailboat | | |
|---|---|---|---|---|
| | | green | red | blue |
| color of Monday's sailboat | green | $p_1$ | $p_2$ | $p_3$ |
| | red | $p_4$ | $p_5$ | $p_6$ |
| | blue | $p_7$ | $p_8$ | $p_9$ |

"Both arguments are responses to the fact that it **isn't logically inevitable** that ONE-MT assigns higher values to the main diagonal in the table than TWO-MT does. If sailboats were **chameleons** (changing their color from day to day), all bets would be off."

ONE-MT says that only $p_1$, $p_5$, $p_9$ can have positive values.

TWO-MT makes no such restriction, applying positive values to all 9 possibilities.

If we invoke the **principle of indifference**, then ONE-MT assigns $p_5$ a value of 1/3, whereas TWO-MT assigns $p_5$ a value of 1/9.

Better, we could use **frequency data** to assign probabilities to each color. ONE-MT will remain the likelier hypothesis.

Bayesian Ockham’s razor handled the sailboats pretty well.

What about the **precession of the perihelion of Mercury**?

N: Newtonian mechanics

N*: Newtonian mechanics with an adjustible parameter ε

GTR: Einstein’s general relativity

N: Newtonian mechanics

N*: Newtonian mechanics with an adjustible parameter ε

GTR: Einstein's general relativity

*Assuming ε has a normal probability distribution centered on 0,*

Pr(Mercury's precession | GTR) > Pr(Mercury's precession | N*)

Jeffreys and Berger 1992

N: Newtonian mechanics

N*: Newtonian mechanics with an adjustible parameter ε

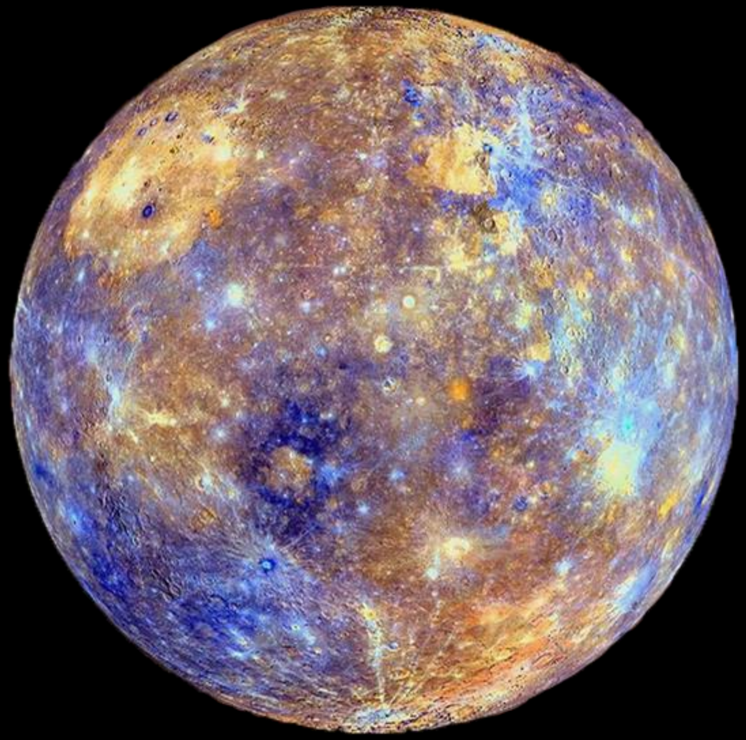
GTR: Einstein's general relativity

*Assuming ε has a normal probability distribution centered on 0,*

Pr(Mercury's precession | GTR) > Pr(Mercury's precession | N*)

And GTR is at least **27 times** likelier than N*!!!

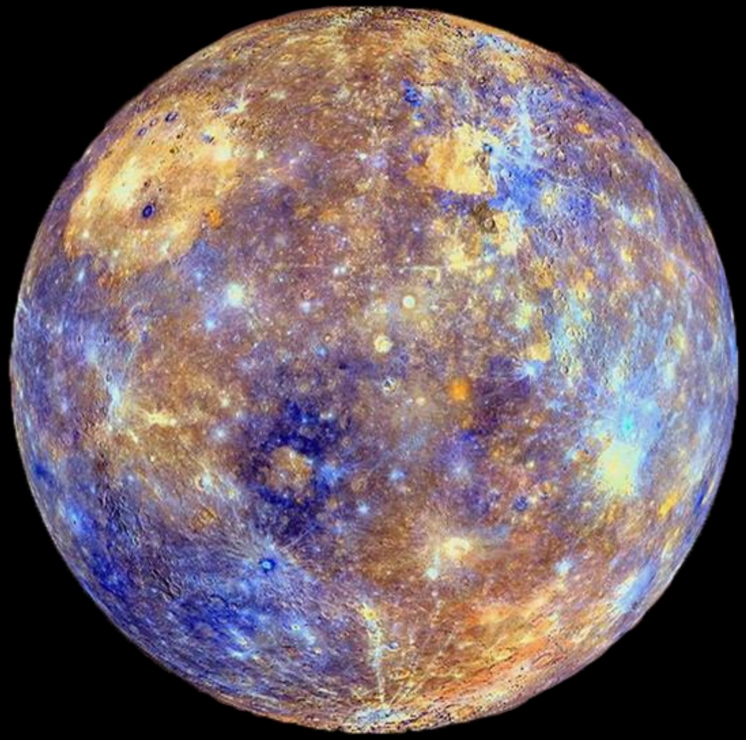
Jeffreys and Berger 1992



But why center the distribution of ε on 0?

“We assert that, prior to seeing the Mercury data, one would have no reason to differentiate between positive and negative values of ε – hence symmetry.” (Jeffreys and Berger)

Jeffreys and Berger 1992

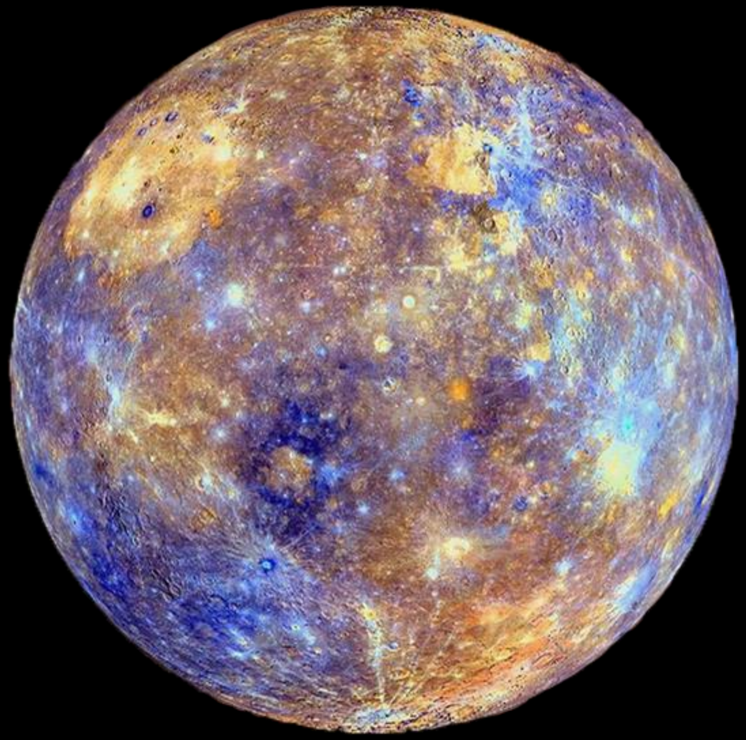


But why center the distribution of ε on 0?

“We assert that, prior to seeing the Mercury data, one would have no reason to differentiate between positive and negative values of ε – hence symmetry.” (Jeffreys and Berger)

“I sense an appeal to the **principle of indifference** in the first sentence; my reply is that having no reason to assume asymmetry is not a reason to assume symmetry.” (Sober)

Jeffreys and Berger 1992



“Bayesian Ockham’s razor works better on the sailboat problem than it does on Newcomb’s N*. You can gather frequency data on Lake Mendota sailboat colors and use that evidence to ground assumptions about the marginal probabilities in the 3×3 table. In contrast, **it is unclear how observation or theory would allow you to justify a value for the average likelihood of N*.”**

Epistemic relevance of simplicity depends on background assumptions (facts?)

Use of toy examples for clarity, contrasted with real examples from science

Is Bayesian likelihood the one and only measure of epistemic relevance?