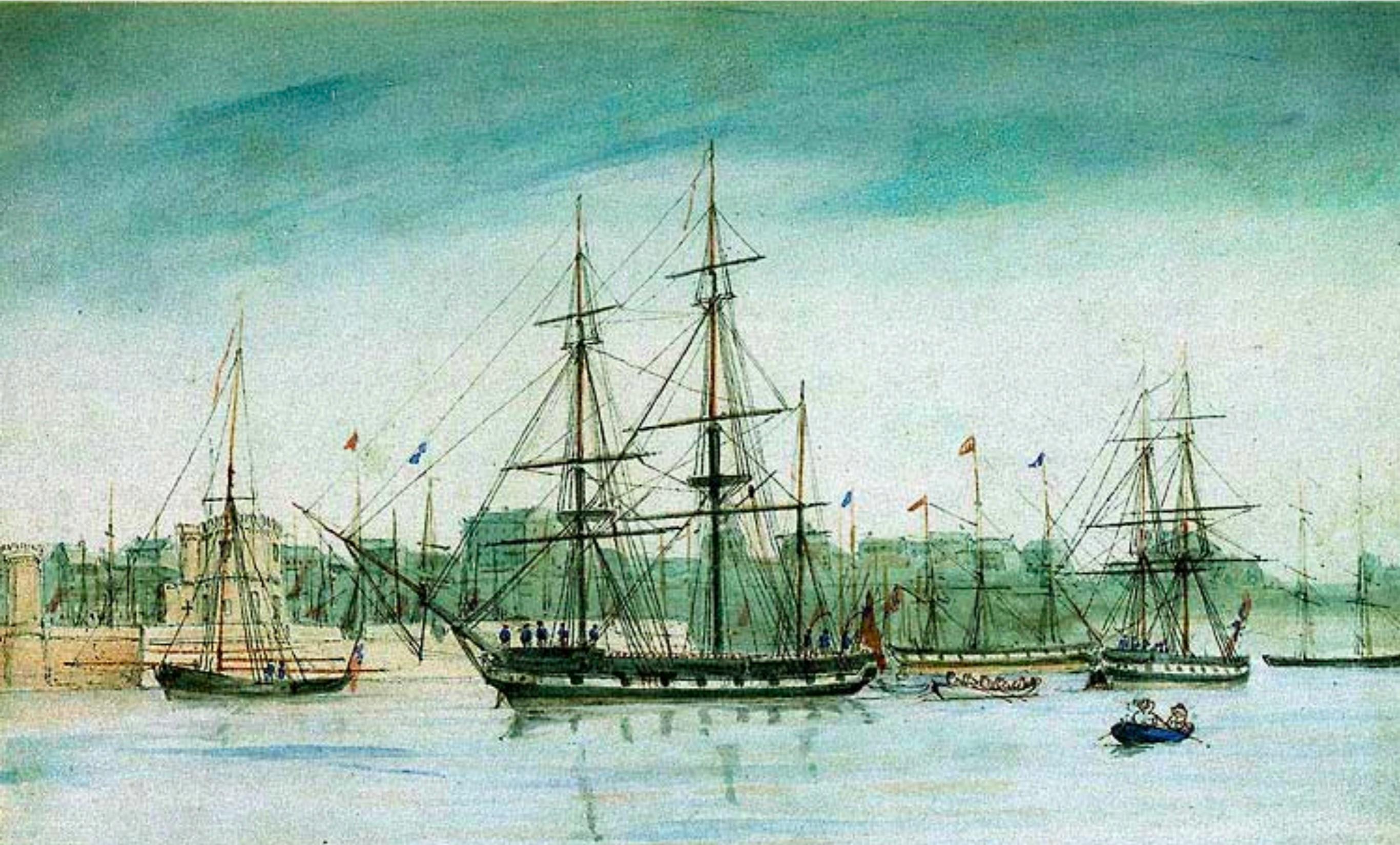


# The Research Journey

CSCI 8901:  
Research & Evaluation Methods

Prof. Tim Wood  
GWU

# HMS Beagle



# Charles Darwin

Born 1809

- Went to college in 1828, but “preferred riding and shooting to studying”. Collected beetles.

Applied to be the naturalist on HMS Beagle in 1831



# Charles Darwin

Born 1809

- Went to college in 1828, but “preferred riding and shooting to studying”. Collected beetles.

Applied to be the naturalist on HMS Beagle in 1831

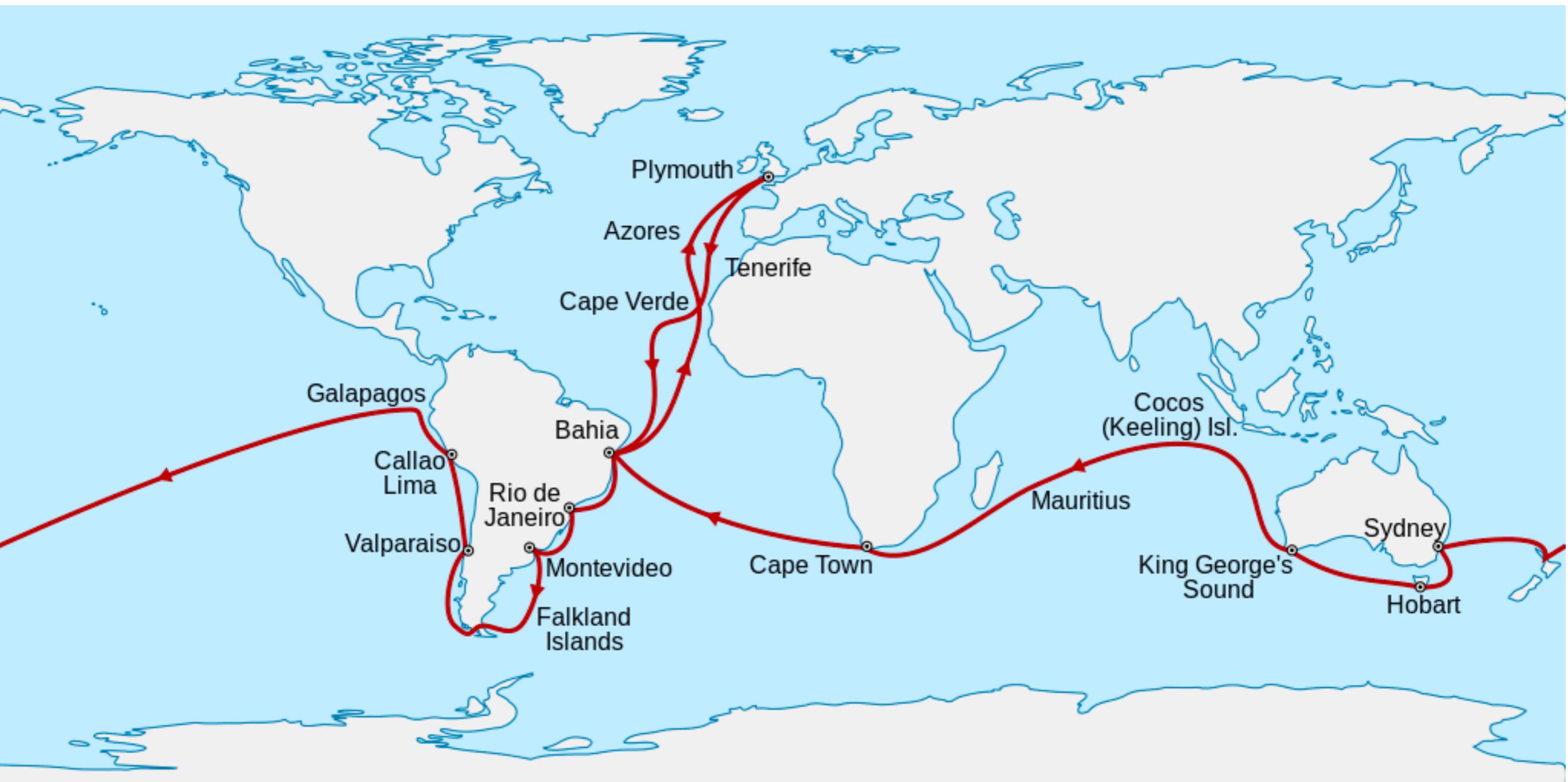
- Captain almost rejected his application:

“He was... convinced that he could judge a man’s character by the outline of his features; and he doubted whether [sic] anyone with my nose could possess sufficient energy and determination for the voyage.”



# 5 Year Journey

1831 - 1836 (a PhD?)



# Evolution

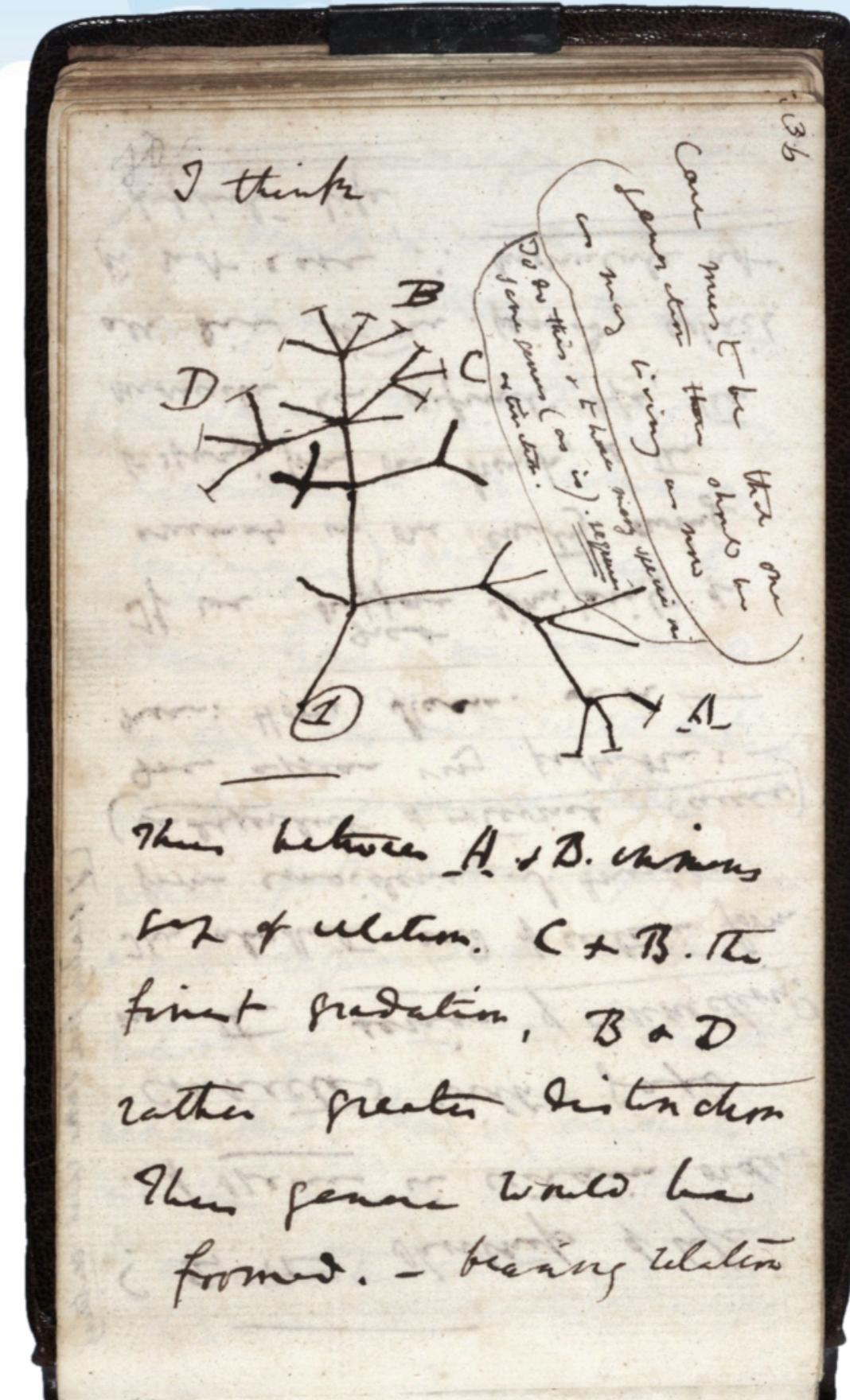
Ideas came together after his trip

In 1838 read work by Malthus on population growth

- Should be exponential!

But populations tend to be stable...

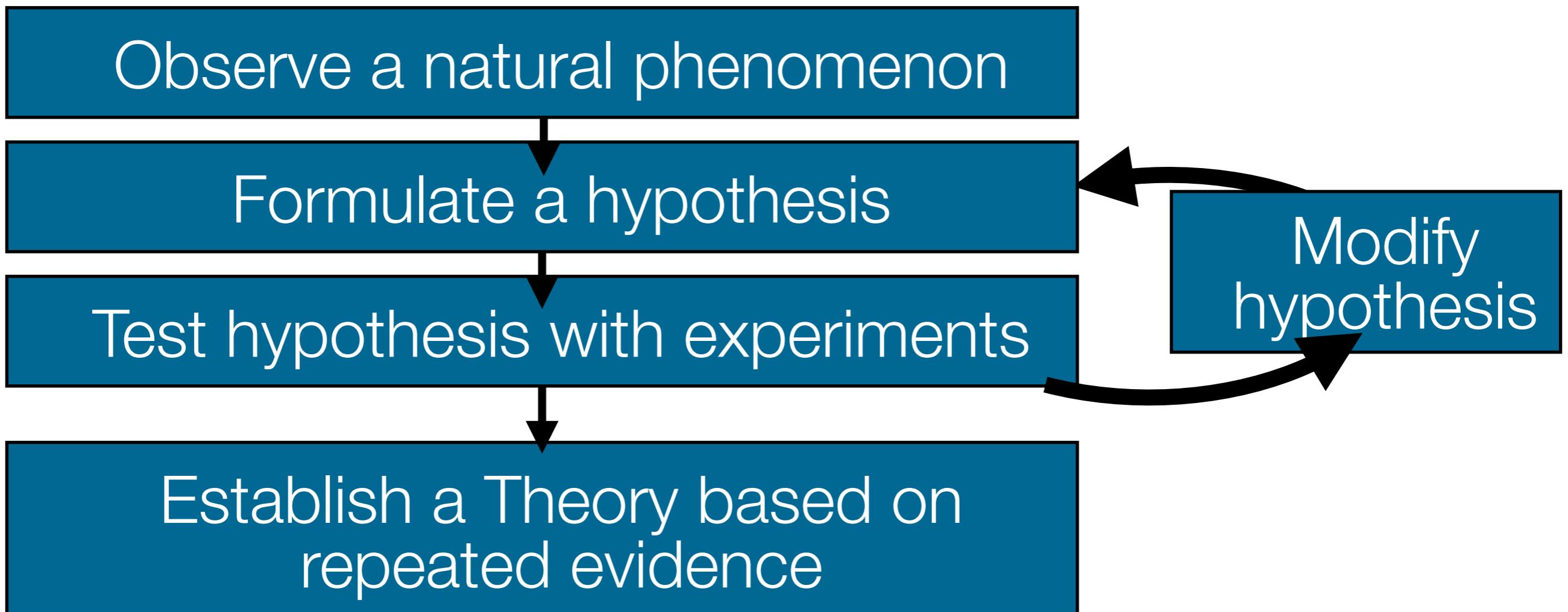
- Which ones will survive?



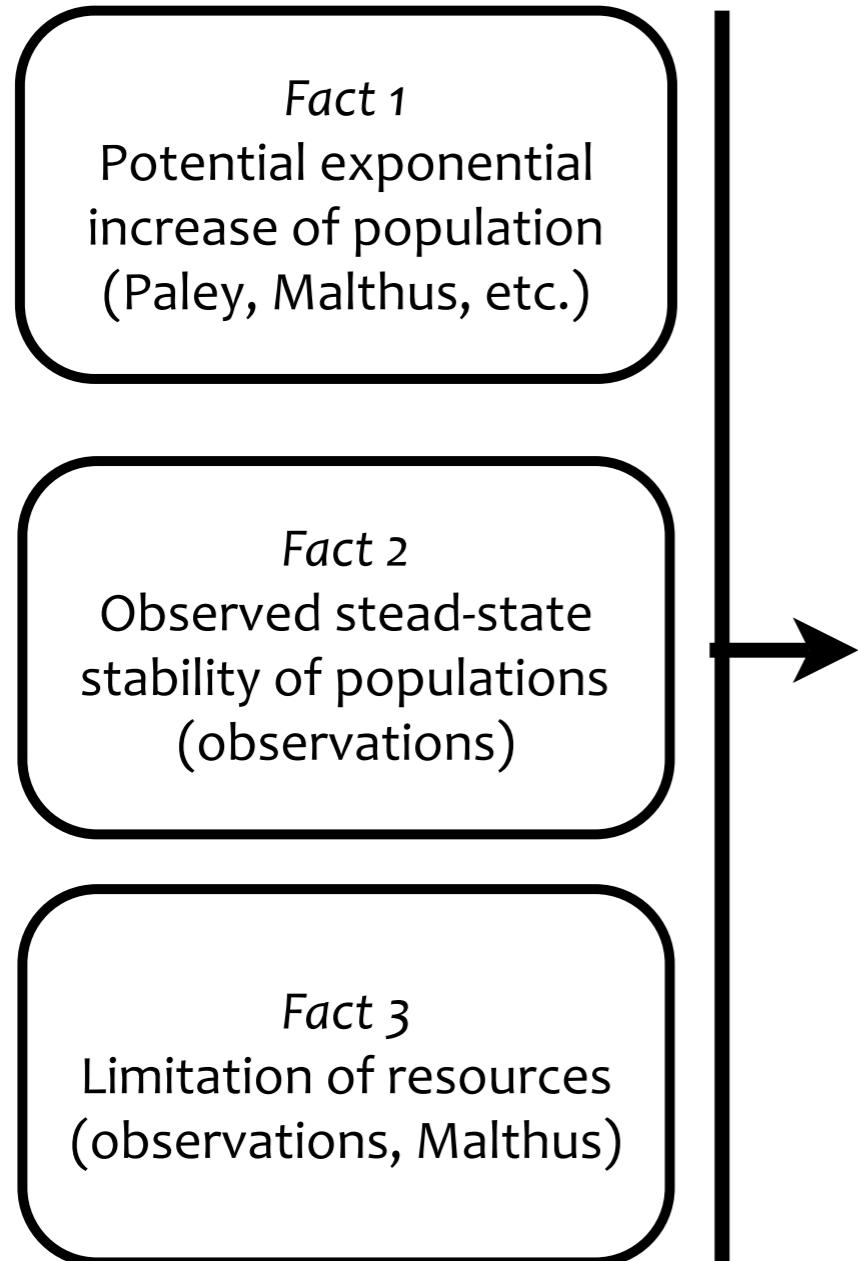
# How does Science work?

**Facts:** data we can gather about the world

**Inferences:** Hypothesis that help explain those facts and tell us about what to expect in related situations.

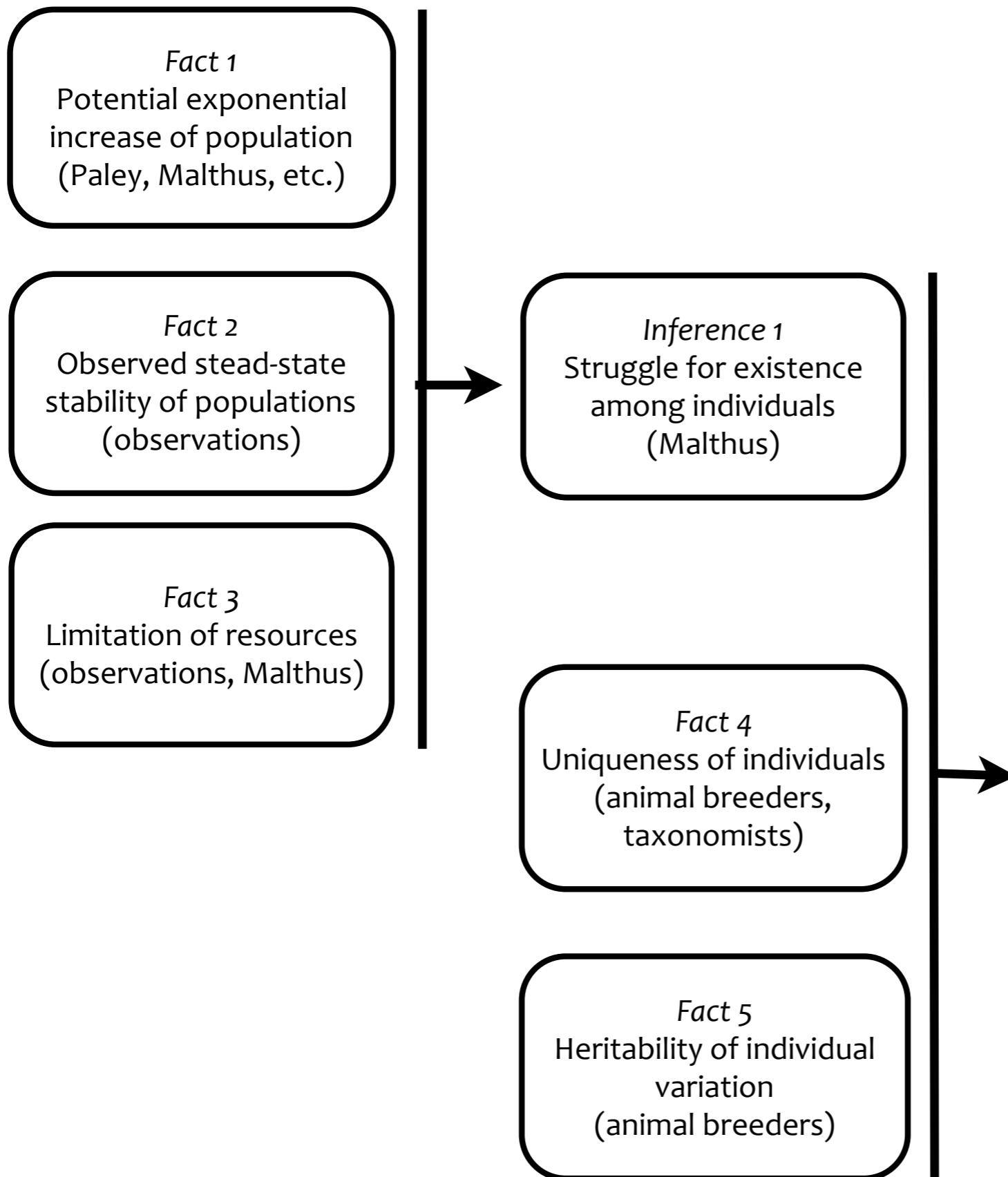


# Journey of Ideas



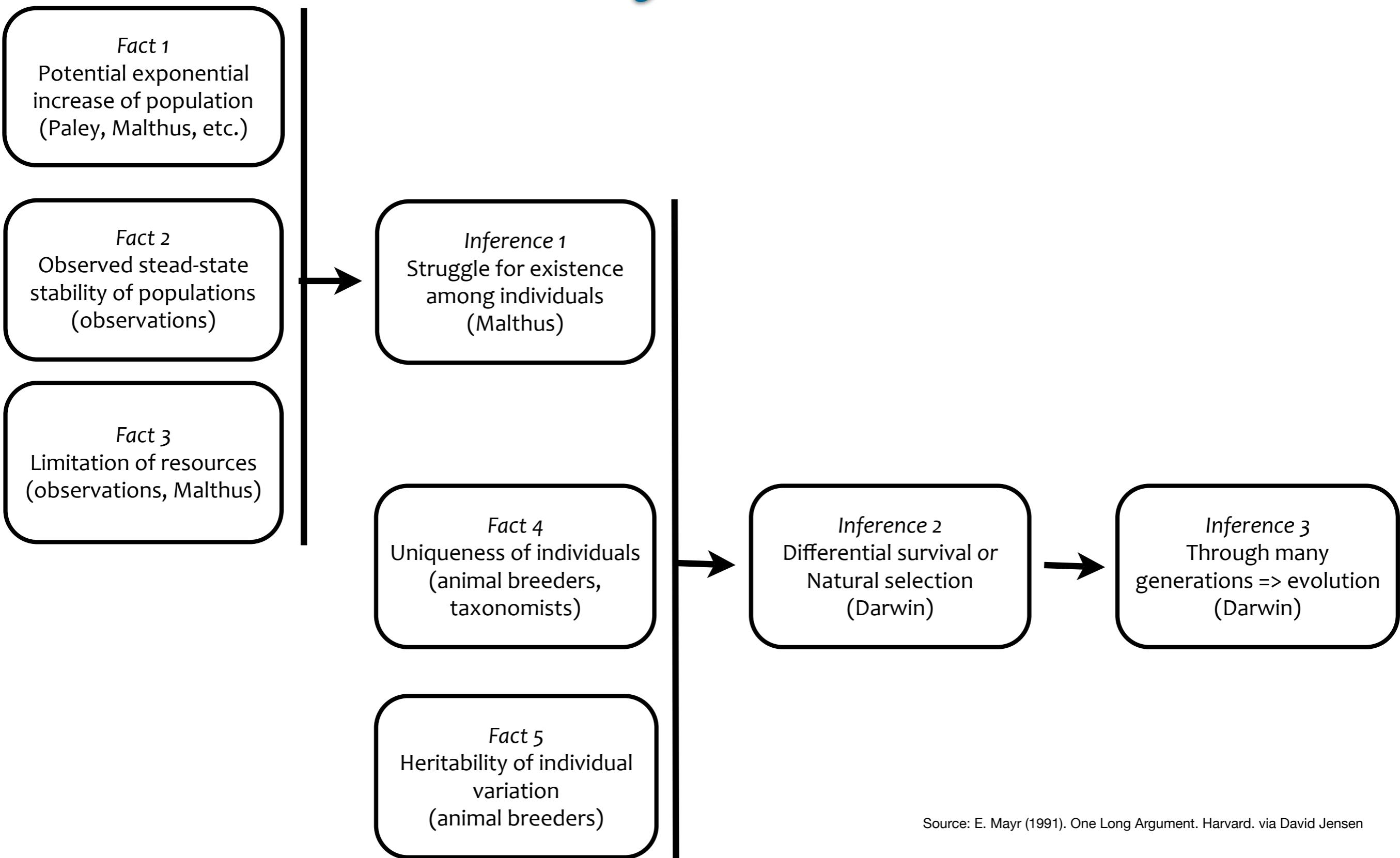
Source: E. Mayr (1991). One Long Argument. Harvard. via David Jensen

# Journey of Ideas



Source: E. Mayr (1991). One Long Argument. Harvard. via David Jensen

# Journey of Ideas



Source: E. Mayr (1991). One Long Argument. Harvard. via David Jensen

# Why is science hard?

## Intrinsic:

- Science is about discovery and thus inherently about something that is unknown

## Personal:

- We as scientists make mistakes, have biases, get distracted, etc

## Communal:

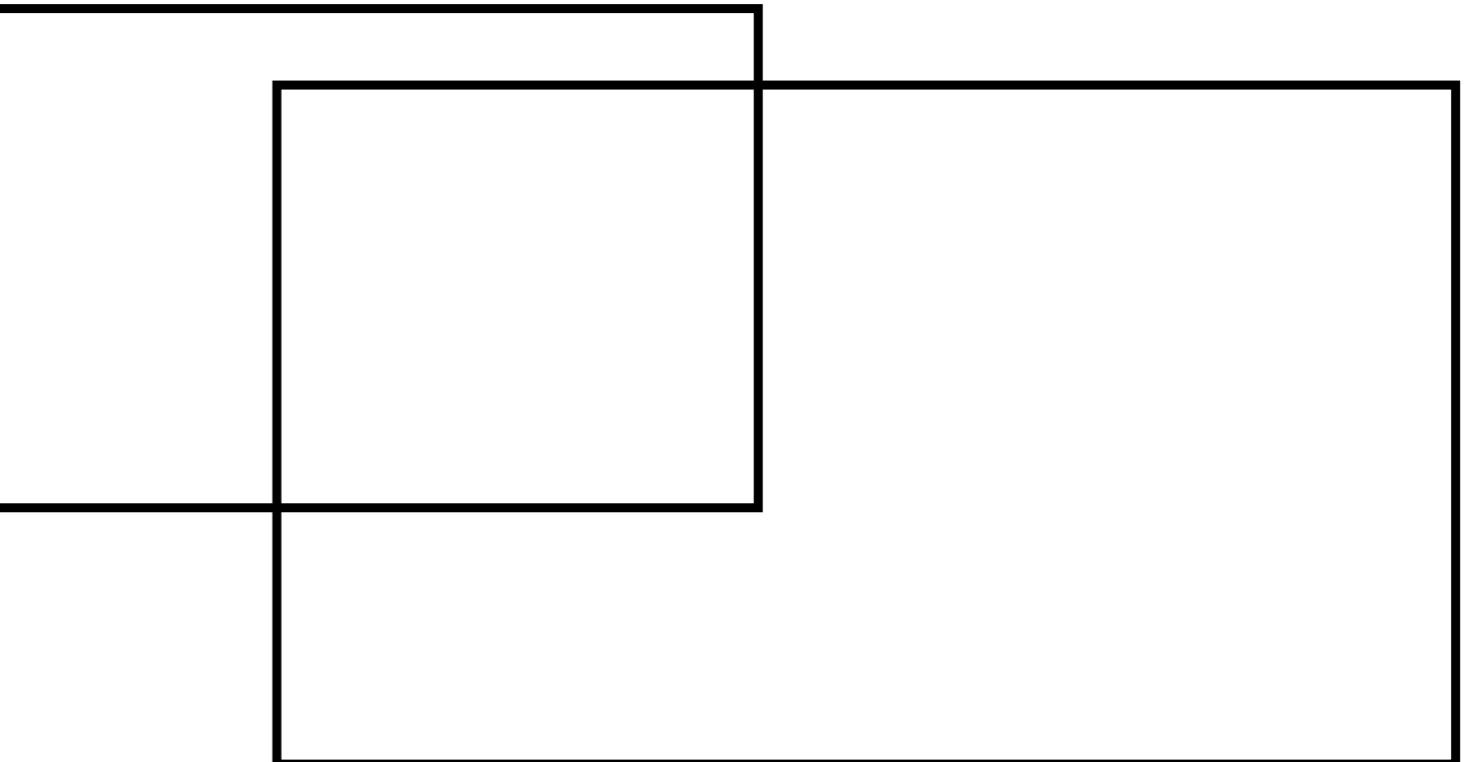
- Progress depends on many researchers coming together, yet our communities don't always recognize important work or share information

Following slides from D. Jensen, Research Methods, UMass

**All possible theories**



**Theories that are actually true**



***Theories we think are true***

***Theories that are actually true***

***Theories we think we have tested well***

***Theories we think are true***

***Theories that are actually true***

# Darwin's Natural Selection

Took a 5-year journey around the world, plus 23 years of further study, and data gathering

Broke from prior theories

Proposed a new theory based on extensive evidence

**Theories we have even considered**

**Theories we think we have tested well**

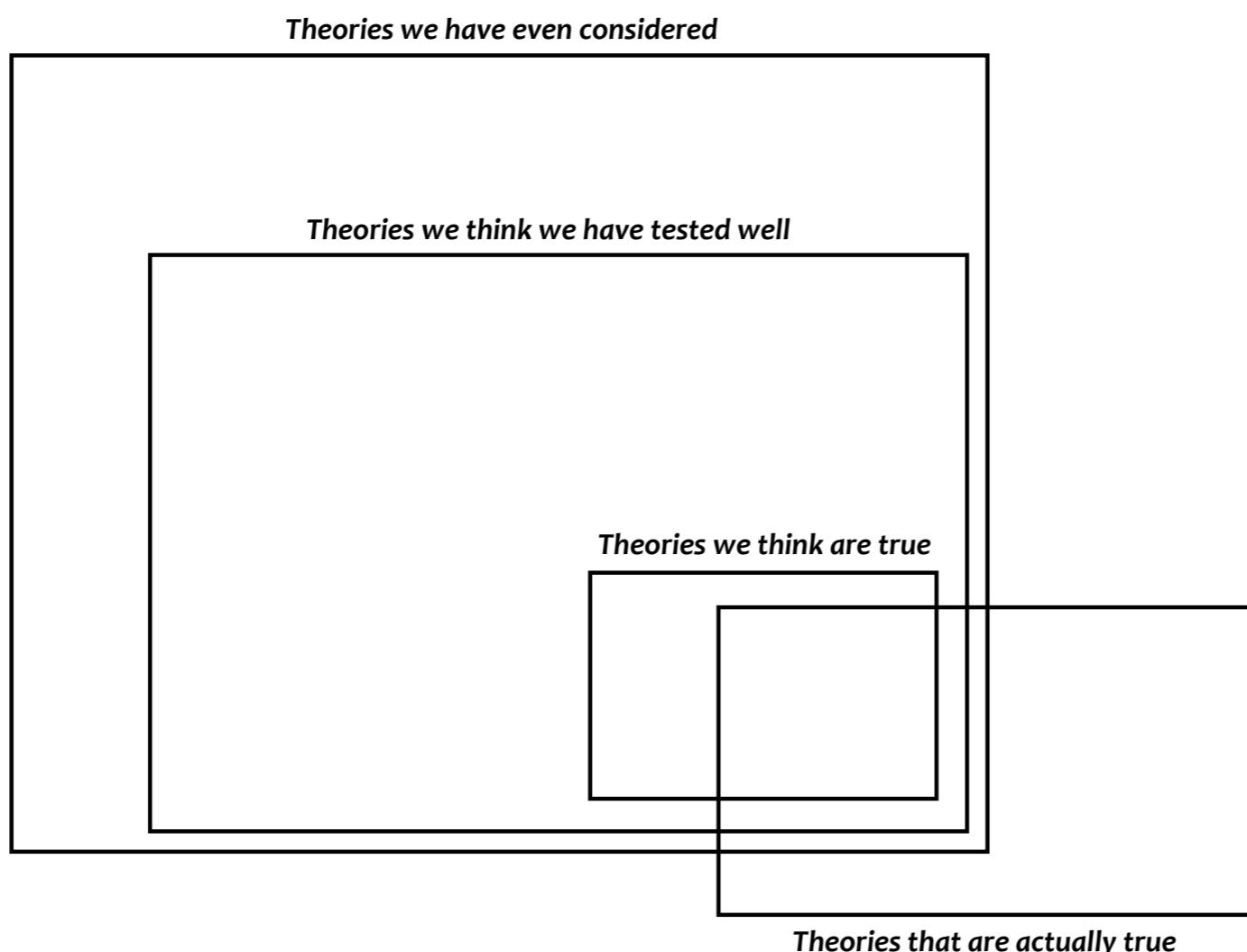
**Theories we think are true**

**Theories that are actually true**

# Darwin's Intrinsic Challenges

Took a 5-year journey around the world, plus 23 years of further study, and data gathering

Proposed a new theory based on extensive evidence



**Many fields in CS  
don't think about  
“theories” that are  
true or false, but  
the same picture  
holds for  
“solutions” that  
are good or bad!**

# Why is science hard?

## Intrinsic:

- Science is about discovery and thus inherently about something that is unknown

## Personal:

- We as scientists make mistakes, have biases, get distracted, etc

## Communal:

- Progress depends on many researchers coming together, yet our communities don't always recognize important work or share information

# Researcher Sins

## Slop:

- Doing research in such a way that it is impossible to know for certain what was done or observed;
- Confused or unclear procedures and data-recording techniques;
- Imprecise theorizing, unexpressed assumptions, and informal derivation of predictions.

## Sloth:

- Doing too little;
- Laziness such that important potential data are not obtained or recorded;
- Partial or incomplete analysis of data.

*From:* Donald D. Jensen (circa 1995), Unpublished lecture notes. University of Nebraska - Lincoln via David Jensen, UMass

# Researcher Sins

## Precipitance:

- Jumping to a conclusion;
- Premature decision on an issue;
- Accepting as established something that deserves further investigation.

## Propaganda:

- Biased presentation of a theory or data;
- Also called "special pleading";
- Acting as a proponent rather than an disinterested presenter of facts and interpretation;
- Salesmanship rather than science.

*From:* Donald D. Jensen (circa 1995), Unpublished lecture notes. University of Nebraska - Lincoln via David Jensen, UMass

# Researcher Sins

## Prejudice:

- Biased evaluation of theory and data;
- expecting more of other theories than of one's own;
- "Tilting the playing field" in favor of one's own theory.

## Perseveration:

- Holding to a theory despite clear evidence that it is false.

*From:* Donald D. Jensen (circa 1995), Unpublished lecture notes. University of Nebraska - Lincoln via David Jensen, UMass

# Researcher Sins

## Finagle:

- “Adjusting” data so that it fits a favored theory. Minor fraud.

## Filch:

- Stealing ideas or data without giving appropriate credit;
- Plagiarism or other unauthorized use of the work of others.

## Fraud:

- Falsifying data and investigation

*From:* Donald D. Jensen (circa 1995), Unpublished lecture notes. University of Nebraska - Lincoln via David Jensen, UMass

# Reflection

What are your sins?

**Slop, Sloth, Precipitance, Propaganda,  
Prejudice, Preservation, Finagle, Filch, Fraud**

Reflect on this... and think about how you can improve!

- We don't expect you to be perfect, but we do expect you to try to continually improve

# Integrity and Ethics

These are **incredibly** important!

- If people question your integrity, they will doubt all of your science! And rightfully so!

We can only make progress if we can trust each other and trust each other's results!

- Never violate this trust!

It is always better to be late/wrong/not the best than to lie/cheat for temporary success

It doesn't matter if you get caught or not

- Always try to do the right thing

# Ethics and Integrity

We can only make progress if we can trust each other and trust each other's results!

- Never violate this trust!

These are **incredibly** important!

- If people question your integrity, they will doubt all of your science! And rightfully so!

It is always better to be late/wrong/not the best than to lie/cheat for temporary success

It doesn't matter if you get caught or not

- Always try to do the right thing

# Darwin's Personal Challenges

Had many personal difficulties in his life

- Recurrent illnesses, several of his children died

Lack of focus?

- Hard to say if this was a strength or a weakness
- Spent many years on less impactful work like barnacles

# Darwin's Personal Challenges

Had many personal difficulties in his life

- Recurrent illnesses, several of his children died

Lack of focus?

- Hard to say if this was a strength or a weakness
- Spent many years on less impactful work like barnacles

But his barnacle classification schemes inspired all his later work!

Didn't feel a rush to complete his work

- Helps to be an independently wealthy English Gentleman
- Delayed several years and only published when he realized Wallace was reaching similar conclusions!

# Why is science hard?

## Intrinsic:

- Science is about discovery and thus inherently about something that is unknown

## Personal:

- We as scientists make mistakes, have biases, get distracted, etc

## Communal:

- Progress depends on many researchers coming together, yet our communities don't always recognize important work or share information

# Research Communities

Historically were not broadly inclusive

Getting better, but still tends to be some bias towards past stars

Fast advances in CS make it difficult to keep up

- Hard to judge what will have lasting effect
- Good work can get ignored or overlooked

Fragmentation between communities limits progress and sharing of knowledge

# Systemic Challenges

Research is primarily funded by grants

- Most are short term, 3-5 years
- Is that enough time to make a difference?

Race to get more publications

- Hard to judge impact in short term; is being at a good conference all that matters?

Encourages Minimal Publishable Unit (MPU)

# Darwin's Community Challenges

Prejudice and skepticism were major deterrents

“Vestiges of the Natural History of Creation”

- Published 1844 **anonymously**
- Proposed transmutation: a single linear chain of evolution, culminating in the white, English man...
- Suggested this was not necessarily guided by an active god!
- Quickly became popular for its radical ideas, but...

*"The Vestiges of the Natural History of Creation," has started into public favour with a fair chance of poisoning the fountains of science, and sapping the foundations of religion. — Sir David Brewster*

Darwin delayed publication for years because of an unwelcome community!

# How to overcome?

## Intrinsic:

- Science is about discovery and thus inherently about something that is unknown

## Personal:

- We as scientists make mistakes, have biases, get distracted, etc

## Communal:

- Progress depends on many researchers coming together, yet our communities don't always recognize important work or share information

## Solution:

- Systematic, research methodology that avoids bias and lets us empirically validate our ideas and their impact!

# Papers



# Zotero Demo

Useful tool for tracking citations

# Paper Review / Lit Survey

Carefully read N papers related to your project

- $1 \leq N \leq 6$  - your choice based on where you are in your project

Describe the paper - 1 paragraph

List strengths and weaknesses - bullet points with 1-2 sentences each

Describe how this work relates to your project/problem

PhD students: Write 1 paragraph on how you will organize papers and track your bibliography

Due in class week 6

# Acknowledgements

Much of the slide content is derived from the *Research Methods for Empirical Computer Science* course taught by **David Jensen**

- <http://dx.doi.org/11084/10002>
- <https://people.cs.umass.edu/~jensen/courses/index.html>
- <https://people.cs.umass.edu/~jensen>
- Many thanks for allowing me to make use of his materials!