Hypothesis-Driven Research

Research types

- Descriptive science: observe, describe and categorize the facts
- Discovery science: measure variables to decide general patterns based on inductive reasoning
- Hypothesis-driven science: make a hypothesis and then test the hypothesis using deductive reasoning
- Engineering science: use of theories and technologies to solve real world problems

Descriptive Science

- Fishing expeditions / pattern recognition / descriptive science is just information gathering, may not be a scientific method.
- Early biology: categories of plants
- It can be an integral part of hypothesis formation, but it is open to the criticism that the results are biased by "cherry-picking."

Discovery Science

- Scientists describe some aspect of the world and use inductive reasoning to draw general conclusions

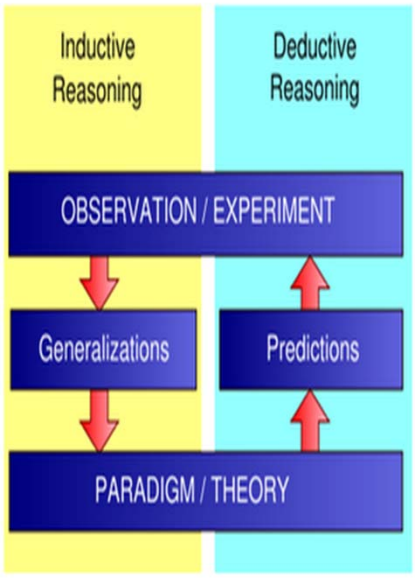
- Example:
  - Scientists observe that all the calico cats they have seen were female, through years of observations the general conclusion has been drawn that "All calico cats are female"

Hypothesis Driven Science

- Driven by deductive reasoning.
- Guided by a hypothesis (a tentative answer to a question) based on an observation.
- If a hypothesis is correct, and we test it, then we can expect a particular outcome.
- Experiments designed to test hypotheses must be controlled experiments.

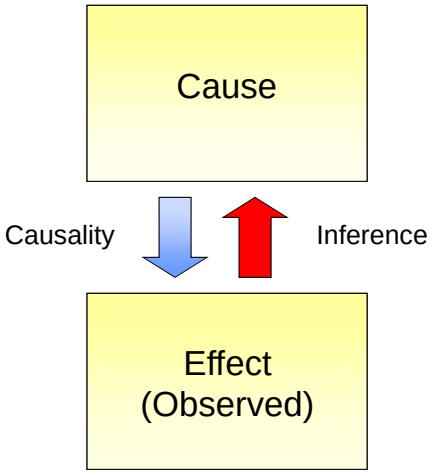
Inductive vs. Deductive Reasoning

- **Inductive Reasoning** – Goes from specific to general
- **Deductive Reasoning** goes from general to specific.



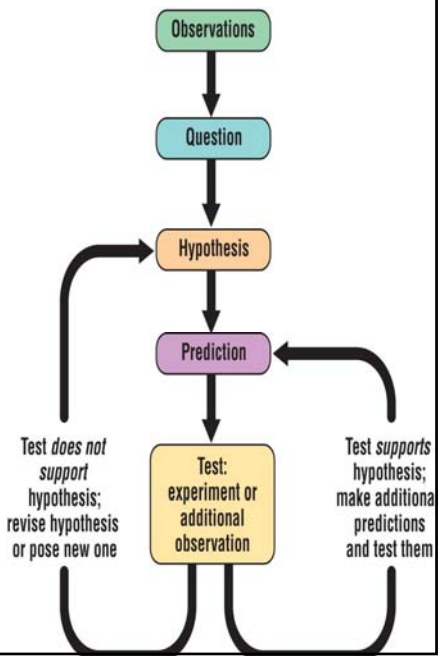
What is hypothesis-driven science?

- The accumulation of knowledge about the world through the *testing* of *causal theories (explanations).*
- Science attempts to *infer* causal relationships ("If A, then B") by application of the scientific method.
- Unique to modern science.



Hypothesis-Driven Science

– As a formal process of inquiry, the scientific method consists of a series of **steps.**
  - The key element of the scientific method is **hypothesis-driven** science.



A formula for (most) hypothesis-driven research

- Background
- Question
- Approach
- Experiment
- Results
- Literal Interpretation
- Author Interpretation

What is a Good Hypothesis?

A possible explanation for an observation or a scientific problem. . .

- Based on observations, inferences and previous knowledge.
- Must be written as a statement.
- Is a predictable or logical conclusion/result.
- Is testable.
- Found to be right or wrong at end of investigation.

Hypothesis-Driven Research

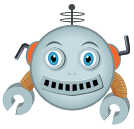
A good hypothesis proposes the obvious missing links between the known phenomena

A good hypothesis makes specific statements on the causal relationships among a few key phenomena

Newton's law of universal gravitation (apple vs. star)

Testing Hypotheses

Account for the "Variables"



- Variables are factors or conditions that can vary and thus influence the results.
- Only one variable should be tested at a time.
- **3 Kinds of variables:**
  - Independent variable = what you change or manipulate
  - Dependent variable = changes because of independent variable changing
  - Control variables = no changes at all, these must all be kept the same in experiment

Hypothesis in Engineering

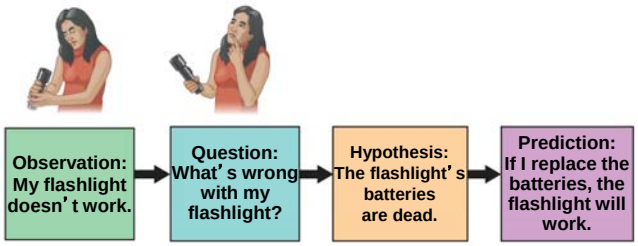


Figure 1.15-1

Hypothesis in Engineering

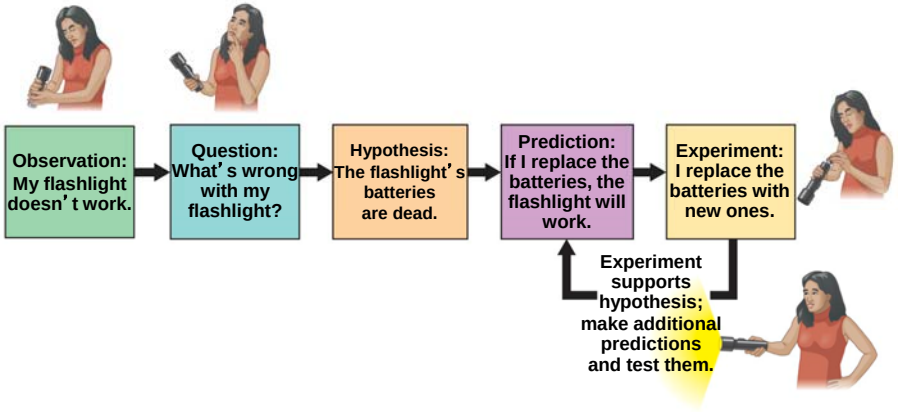
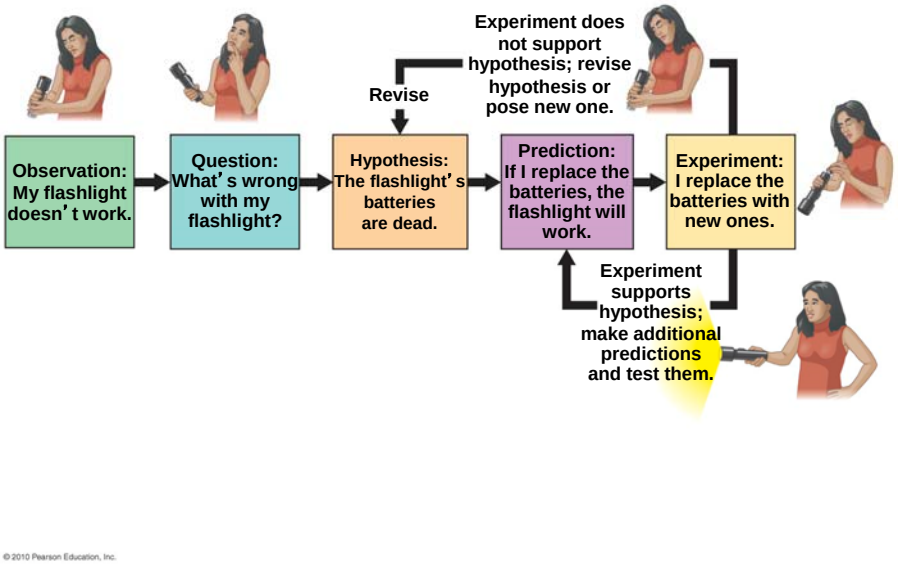
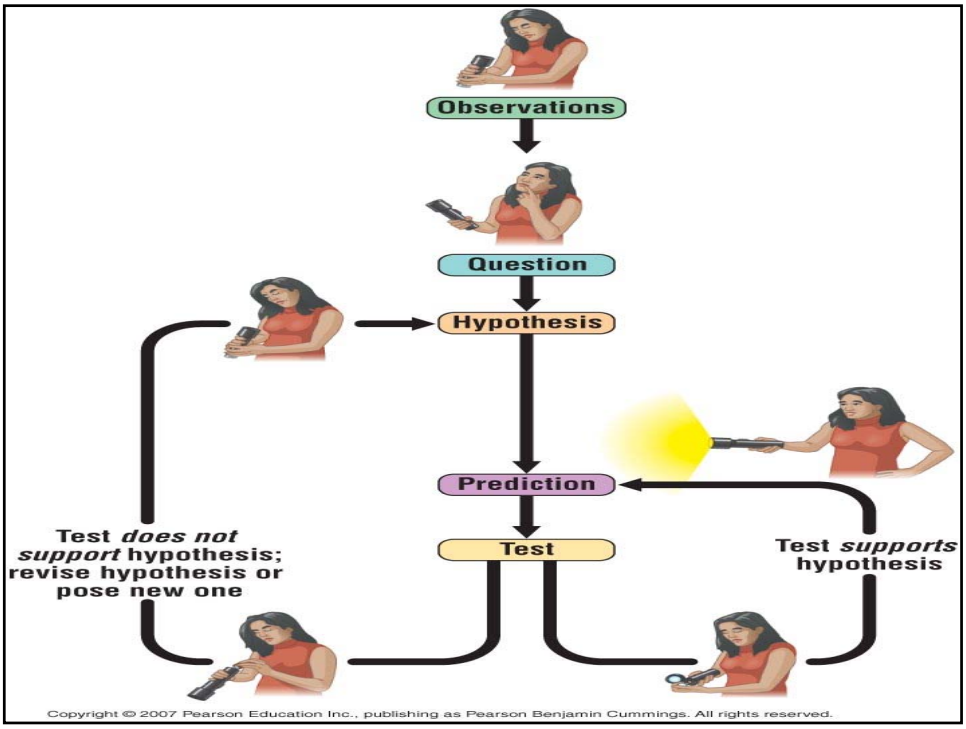


Figure 1.15-2

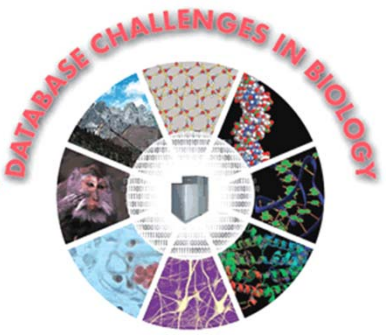
Hypothesis in Engineering





Science, then, and now…

- Today's experiment yields massive amounts of data
- From hypothesis-driven to exploratory data analysis:
  - data are used to formulate new hypotheses
  - computers help formulate hypotheses
- No single person, no group has an overview of what is known



Data Mining for Big Data

- **Hypothesis-driven data mining**
  Begins with a proposition by the user, who then seeks to validate the truthfulness of the proposition
- **Discovery-driven data mining**
  Finds patterns, associations, and relationships among the data in order to uncover facts that were previously unknown or not even contemplated by an organization (don't want to have descriptive science)

From Hypothesis-driven to Data-driven Science

- *Genomics*: measure all genes at once.
- Don't have to assume a hypothesis as basis for designing the experiment.
- Objective: let the data speak for themselves.
- Reality: vast amounts of data, very complex, hard to interpret.

Poor Science: Data-driven Science Done Wrong

- No hypothesis.
- Assumptions: alternative models not explicitly enumerated, weighed.
- Statistical basis of model either neglected or only implicit (and therefore poor).
- No cross-validation: just one form of evidence.
- Greedy algorithms, sensitive to noise.
- Measures of significance weak or absent, both computationally and experimentally.

Data-driven Science Done Right

- Multiple competing hypotheses.
- Alternative models explicitly included, computed, to eliminate assumptions.
- Statistical models clear, well-justified.
- Multiple, independent types of evidence.
- Robust algorithms w/ well demonstrated convergence to global optimum.
- Rigorous posterior probability calculated for all possible models of the data. Priors derived from data. False +/- measured.

Implications of Data-driven Science

- Don't confuse observations & interpretations.
- To get strong posteriors that can distinguish multiple models, you need LOTS of data.
- The end of (purely) human analysis.
- Big data
  - Variety:
  - Velocity:
  - Volume
- Product affinities

Summary

- Hypothesis-driven study is a fundamental, useful skill; applicable to many other things
- Be aware of applying hypothesis-driven research
- Limitation of hypothesis-driven research
- Design your research before doing it: Require a lot of thinking, get input from others
- Integrate hypothesis-driven and discovery-driven research

Acknowledgments

The content is extracted from a presentation by Dr. Dong Xu.