Research Methods

CSCI 8901:
Research & Evaluation Methods

Prof. Tim Wood
GWU

These slides draw inspiration from similar courses by:
David Jensen, UMass
Omprakash Gnawali, University of Houston

Course Overview

Reading

Is "Computer Science" Science?

Experimental design

Data processing and statistics

Writing

Oral presentations

Creativity

(Not necessarily in this order, all subject to change)

Introductions…

Who are you?

- What should we call you? (3 times slowly)
- How long have you been at GW?
- Where did you come from?
- Who are you working with?
- What is your general research area?
- What is your favorite programming language/ movie/data structure/food/theory/algorithm/ vacation spot/etc? Why? (briefly)

Course Outcomes

What do you want to get out of this course?

-

Research "Craft"

The tools and techniques for daily research tasks

- Working on problems in your specific research domain
- Writing good code
- Using version control, latex, etc

Less focus on this…

Goal: Cover enough of the practicalities to make you efficient in your work

Research "Evaluation"

Statistics, modeling, and experiment analysis

- Metrics
- Significance testing
- Causal hypothesis testing
- Scripting, data parsing/scrubbing
- Visualization and graphing

Some focus on this…

Goal: Ensure you can run accurate, repeatable experiments

Research "Professionalism"

Turning your research skills into a successful career

- Job applications / interviewing
- Getting along with your advisor
- Writing grants or patents
- Ethics

Some focus on this…

Goal: Make you think about your future career path options, even if it is far away

Research "Methodology"

How to conduct a personal research program:

- Identifying important problems
- Selecting among possible research directions
- Identifying and effectively reading related work
- Designing useful experiments
- Drawing conclusions from data

This is the course's main focus

Goal: Jump start your Ph.D. research career by learning the high level path forward

What you will do

Read a lot

- Much of it will NOT be in your research area. Some of it will be way outside of CS

Write a lot

- Lots of small writing assignments building up to a full report
- I will try to give feedback on both grammar and content

Speak a lot

- Informal discussions during class
- Short, unplanned presentations
- Longer, prepared in advance presentations

In class: think about things outside your area
Out of class: apply these concepts to your area

Major Assignments

Research project

- Multi-stage assignment involving proposing project, surveying related work, defining and performing experiments, and writing a final report
- You should be doing all of this anyway as part of your normal research activities! Should add minimal "extra" work if you are research active

Tutorial (maybe)

- Presentation that teaches a concept or tool related to your research

Class Participation

You **must be active** in this course or I will be **SAD and disappointed in you**.

Ask and answer questions

Make suggestions to improve the course

Let everyone get a chance to participate

Challenge other people's ideas, but be respectful

You should speak every single class

How to Read

Research Papers

Why do people write research papers?

For the glory

- Lets other people learn your great ideas

For the future

- Makes a lasting record of an idea

For themselves

- Writing a paper crystalizes your ideas

Why Read?

Why should you read papers?

To keep up with your research community

- Shows you what **problems** others think are important

To inspire your own work

- Find better **solutions** by leveraging other smart ideas

What to Read?

Papers from the top conferences/journals in your field

- Browse the program once it is available
- Look at previous years to understand what types of papers are accepted there

Papers that are cited frequently

- Find the authors that repeatedly appear in your bibliographies

Old papers that are the basis for your field

- Becoming easier to find with things like google scholar
- Ask your advisor for suggestions

How do **you** read…

A novel?

A magazine?

A newspaper?

A research paper?

Phase 1: Skim

1) Read the abstract and introduction
   - Highlight each **contribution** they claim
2) Look at the title of each section/subsection
   - Guess what it will be about, but don't read it carefully
3) Examine the figures and tables
   - Understand what metrics they will evaluate
4) Read the conclusion and any parts that stand out

You now know:

- Paper type: theoretical, modeling, implementation, measurement
- The main goals of the paper
- What evaluation the authors think is important

Abstraction

What to do if you don't understand something?

1. Read more background or related work
   - Or ask for help!

2. Treat the idea/tool/algorithm/etc as a **black box**
   - What are the inputs?
   - What are the outputs?
   - What is the overall goal?
   - You may not need to understand the details of how it does these

Practice time!

Read/Skim for 10 minutes

Discuss with group for 5 minutes

Phase 1: Skim

What do you know?

What do you not yet know about the paper?

Phase 2: Understand

From skimming you know ***what*** they have done

Hypothesize what you expect to be in the paper

- Based on Phase 1's skim, what do you think they are doing?
- How would you solve this problem?

Now read the paper to understand ***how*** and ***why***

- Skip non-essential implementation details, proofs, etc.

Make notes as you read

- Summarize main points
- List questions you have
- Circle references that you need to look at for background
- You should have a way to make notes ON the paper (digital or physical)

You now should clearly understand the problem and the proposed solutions

Phase 3: Critique

The goals of reading a paper:

- Learn about new tools/problems/algorithms
- Critique the paper's science
  - Not it's grammar or figure colors, unless you are a reviewer

Third phase: judge the paper's details

- Skip the background that you have already read
- Analyze the assumptions being made
- Consider how you would solve the problems and compare
- Think about what is missing (evaluation, assumptions, proofs…)

(This can be combined with Phase 2 once you have enough practice)

Read a lot

Practice will let you read more efficiently

Knowing the strengths and weaknesses of other papers will help you improve your own ideas

Knowing the hot problems in your community is crucial for guiding your own research

- Most papers don't fully solve a problem or don't solve it perfectly!

Set yourself a goal:

- Skim 3 papers a week and pick one to read thoroughly
- In a year you will know the basic idea behind 150 papers!
- I guarantee this will make you a better researcher

Be Organized

Have a scheme to manage the papers you read

- If you can remember all of them, you aren't reading enough
- Ideas cycle; 5 years from now you may want to recall an old one

I use Zotero

- Browser plugin + application
- Imports/exports bibtex
- Stores PDFs
- Searchable, tagable, note-able, shareable

You must install and learn how to use this!

How to cheat

Discuss papers with others

- Find someone else who has already read the paper
- Explaining a paper is the best way to really understand it
- You should have at least done a Phase 1 skim through the paper

Watch conference presentations

- Some conferences now post videos of all sessions
- Warning: a conference presentation does not give all the details!

Read conference reports

- Some conferences provide summaries of papers
- *Usenix Login* magazine has these every few months (free online)
- A good way to find which papers in a session are worth reading

CS as Science

Edwin Smith Papyrus

Egyptian medical textbook from ~1,600 BCE

Proposes a system for learning how to treat disease

1. Examination
2. Diagnosis
3. Treatment
4. Prognosis

This is the basic form of **Empirical Research**

Empirical Method

This course will focus on *empirical* computer science

- Science where you experimentally evaluate a phenomenon

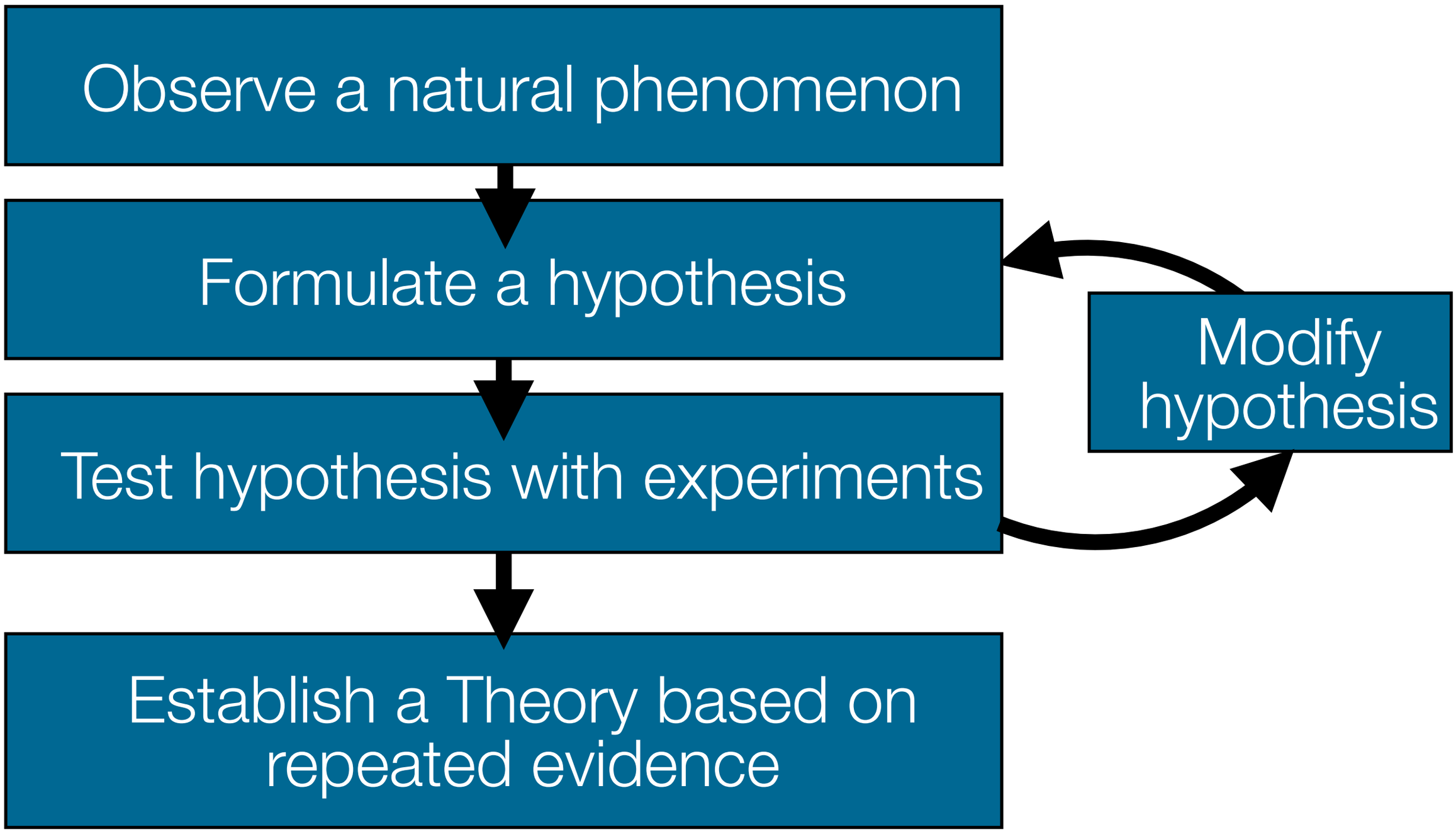
This applies to most fields of CS…

- AI: develop new algorithms and then evaluate their effectiveness
- Systems: build new architectures and evaluate performance
- HCI: design new interfaces and measure user satisfaction

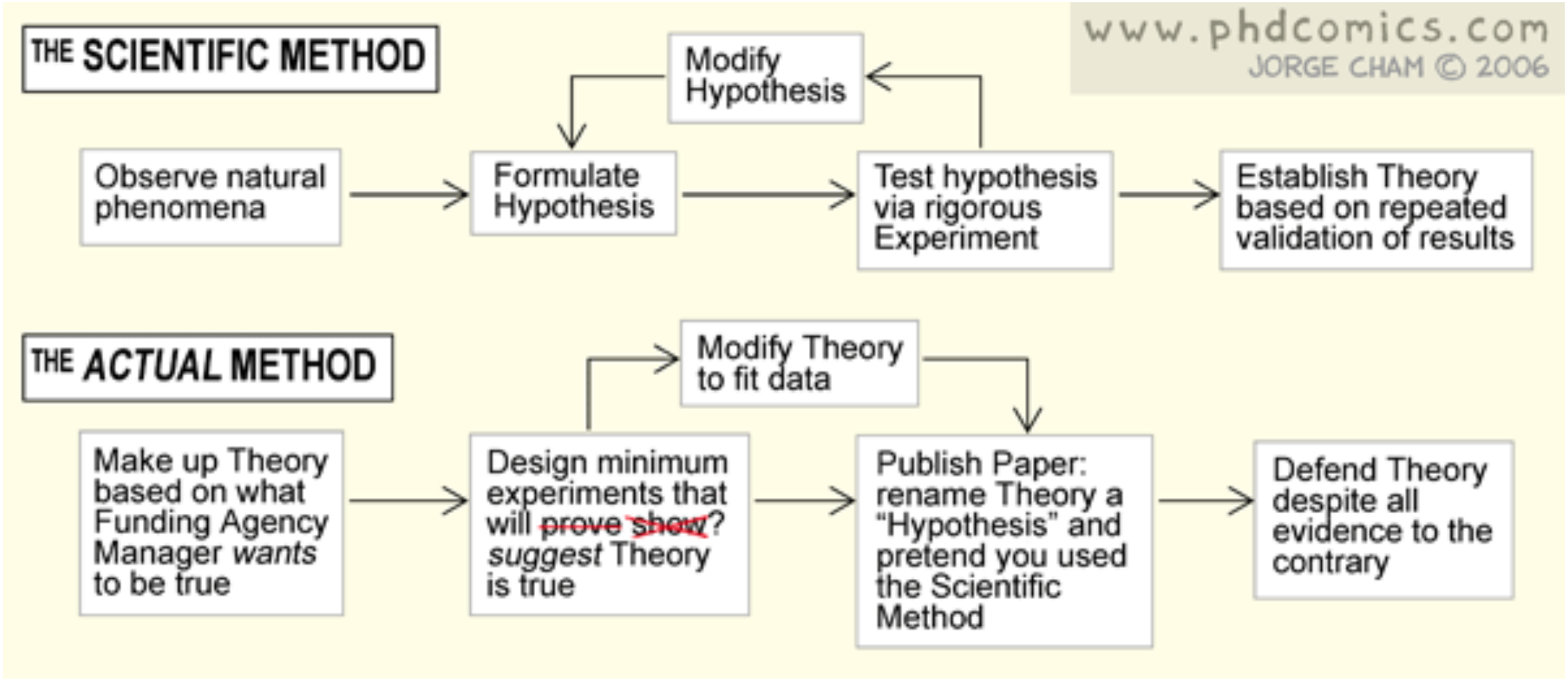
…but not so much to purely math-based fields:

- Theory of computation: prove things to be true or false
- If this is your area portions of the class may be less relevant

Scientific Method

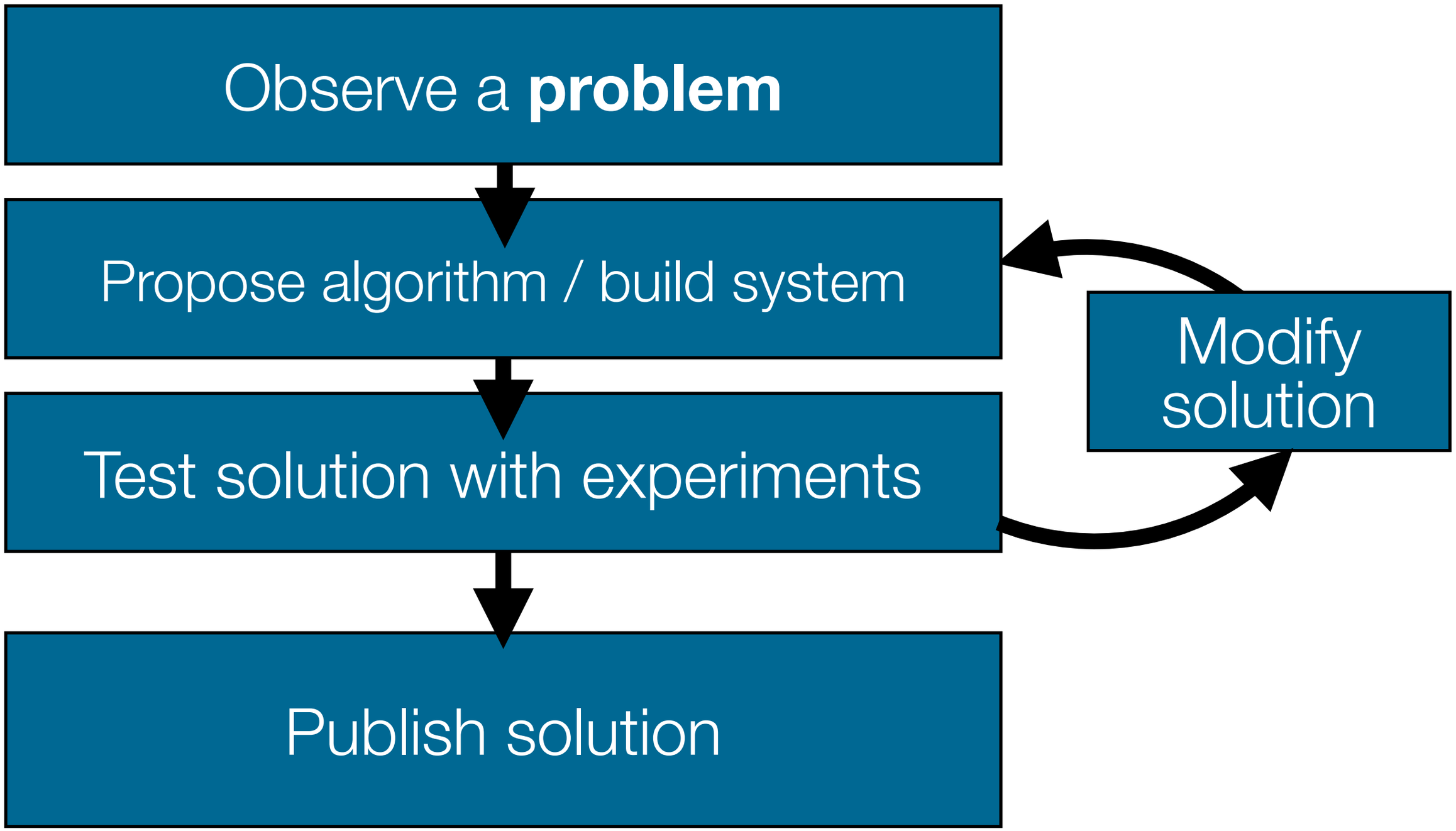


Scientific Method



http://phdcomics.com/comics/archive.php?comicid=761

CS Scientific Method



Computer Science

"Science is not science fiction.
It accepts the tests of observation and experiment, acknowledges the supremacy of fact over wish or hope.
The smallest experiment can crash to earth the most attractive theory."

— Herbert A. Simon

Research Project

Semester-long research project

- Can be a small-ish part of a larger project
- Should NOT involve significant development
- Ideally you will have a readymade system to explore
- Need a concrete artifact you can evaluate
- Project should focus on empirical evaluation of an algorithm, system, model, etc

Assignment 1

Describe your research project

- **What** is the problem you will look at?
- **How** do you plan to solve it?
- **Why** is the problem important and why will your solution be better than other options?

Total length: at most 1 page

- Submit in PDF format
- Suggestion: use latex (possibly www.overleaf.com)

Bring a printed copy to next class (if you are in person)

Reading 1

A Guide to Increased Creativity in Research — Inspiration or Perspiration?

by C. Loehle, in Bioscience February 1990

A guide to increased creativity in research— inspiration or perspiration?

There are four requirements for a successful career in science: knowledge, technical skill, communication, and originality or creativity. Many succeed with largely the first three. Those who are meticulous and skilled can make a considerable name by doing the critical experiments that test someone else's ideas or by measuring something more accurately than anyone else. But in such areas of science as biology, anthropology, medicine, and theoretical physics, more creativity is needed because phenomena are complex and multivariate.

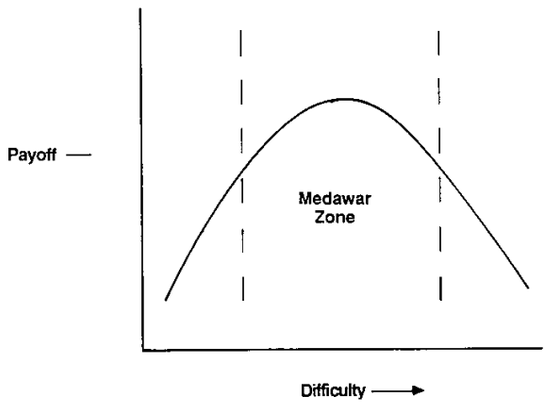


Figure 1. Relationship between degree of difficulty and payoff from solving a problem. Solving problems that are too easy

and work intensely. After you have finished writing your paper, you can go back and remove the comments about what an imbecile the other person is. The effort to refute someone can even lead to evidence supporting them or to a different topic altogether. Intensive rivalries, as in the race to discover DNA (Watson 1968), can also provide this essential intensity. Thus whereas the finished product may appear dispassionate, truly creative work is often driven by strong passions.

File is in Zotero group

Acknowledgements

Much of the slide content, and almost all of the amazing quotations, are 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!