

# CS 6000 Research Methods

---

Terrance E. Boult

El Pomar Prof. of Innovation and Security

The objectives of this course are for you to gain an understanding of :

- Some basic CS publishing/development tools  
(LaTex, Git etc..)
- How to effectively read CS research papers
- How to effectively write a research paper
- How to effectively present research papers
- How to develop a thesis proposal

- The objectives of this course are for you to gain an understanding of :
- How to design an experiment and the statistical analysis of the resulting experimental data.
- Basic usage of cloud computing for CS research.
- and appreciation for and understanding of different subfields of CS and their approaches to CS research,

# Course Ground Rules

- Try to understand
- Don't be afraid to ask (email okay)
- Don't be afraid to comment
- Be constructive
- Be polite
- Don't be afraid to criticize (constructively!)
- Be prepared & do the work on time

# Grading

- Writing assignments/Paper 30%
- Oral Presentation 15%
- Class participation and/or online discussion 15%
- Weekly journal 20%,
- Experimental Design Assignment 10%,
- Cloud assignment 10%

# In This Course

- We will give you many research papers to read and hear many research talks
- If the paper is published, free to discuss anywhere even online.
- If the paper is unpublished (it has been submitted and is currently under review)
  - ◆ Treat it as confidential
  - ◆ Do not show it to anyone outside of this class
- Be ready to put yourself in the mind of a reviewer, author, researcher.. as you'll need to be each

# First Assignment

- Week 1 journal is to be done in latex, and you are to upload the pdf which will include a link to your git repo.
- The document should have a title/author.
- The first section should be 200-300 words on your goals for this course, what you hope to learn, which degree you are studying and something personal plus your photo as a figure. This same content should be in your canvas profile.
- The second section should be about where to find your source code and your learning.
  - It should have hyperref/link to git repo (either public or shared with [tboult](#) on github or tboult on bitbucket) for the body of the latex.
  - Describe with whom you shared it and what tools you used. The body should discuss your learning from this week, problem/struggles you had and how you overcame them.
- The third section should contain a 1-3 sentence description, plus citation, to each of five papers that are related to your research. It should reproduce at least 1 equation from at least 1 paper and discuss it.

# Role of Research

- Should be about some problem that encourages enthusiasm (for you) and interest (for others)
- Is often generated from the thought “what we’ve got now/from the past isn’t quite right/good enough – we can do better...”
- Consists of work that leads to a meaningful contribution and evaluates that contribution
- Generates, in some way, a better solution to the problem – it advances the state of the art
- Needs to be shared and used to have an impact
- Probably needs to be reviewed before most

# What Is Research?

## ➤ **Merriam-Webster's definition:**

- 1** : careful or diligent search
- 2** : studious inquiry or examination;  
*especially* : investigation or experimentation  
aimed at the discovery and interpretation of  
facts, revision of accepted theories or laws in  
the light of new facts, or practical application  
of such new or revised theories or laws
- 3** : the collecting of information about a  
particular subject

# What Is Research?

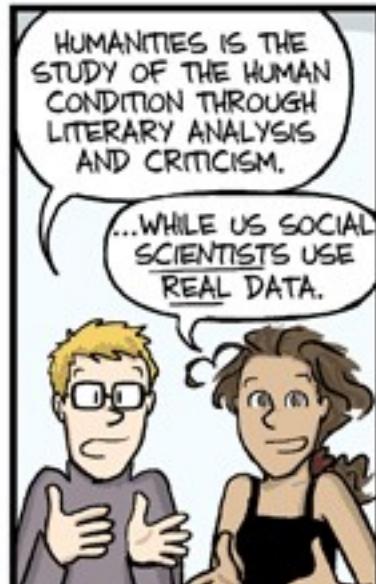
- “A combination of investigation of past work and effort in the present that will help others in the future”
- A set of opposites
  - ◆ Fun and frustration
  - ◆ Small steps and large insights
  - ◆ Building on others’ work and contributing your own work
- Finding or developing something new that impacts the world....

# What Isn't Research

- Playing with technology
- Book reports
- Programming project
- Commercial product development
- Doing what others have already done
  
- However, each of these can be done as part of research.

# What is fundamental research

- The endless debate about fundamental research, basic research, applied research, finalized research...
- A first and some say definitive answer : a study of something in pure form, with no application that leads to a result that has some scientific value as soon as it has been accepted by a scientific publication (a scientific committee)
- A second answer : a result has some scientific value if it interests people and if it is novel (original) (and if it has some genericity)
- A third and I say definitive answer : a result helps solve a problem and has some general scientific value cited by others, e.g. in an reviewed published paper.
- The most difficult (and so the most interesting) problems have often been raised by applications, e.g. medical applications. Investigating these problems have gone through theoretical scientific developments.
- My Bias: Fundamental research is that which help solve multiple-real world problems. Until then its just “speculative research.”



# Who Does Research?

- Graduate Students
  - ◆ Masters Degree (lower expected output)
  - ◆ Ph.D. Degree (higher expected output)
- Researchers at universities
  - ◆ Post-Doctoral students
  - ◆ Faculty members
- Researchers in industry
  - ◆ Research scientists
  - ◆ Many other technical workers
- Undergraduate students

# Who Does Research?

- Individuals
- Teams
- Teams almost always make the process easier
  - ◆ Feedback from team members
  - ◆ Multiple-skill sets and division of labor
  - ◆ Each member can work to own strengths

# Scope of Research

- Varies by level of work
  - ◆ Ph.D. students – contribution expected at global/world level; e.g.
    - background investigation on all past work
    - make meaningful addition to world knowledge
  - ◆ MS/Undergraduate students – contribution can be at local to national to world level; e.g.
    - background investigation at university up to world level
    - make meaningful addition to university up to world level of knowledge
    - much ugrad research is to learn the process and get started down the path

# What is Research?

- Quantitative vs. Qualitative Research
  - ◆ Quantitative – use of mathematical models, software statistical, formulaic or numerical analysis to analyze results
    - Main approach: discovery, analysis; causal determination, prediction, generalization of findings
    - Results: “This solution is N% better”
  - ◆ Qualitative – not quantitative; use of non-numeric techniques to analyse results
    - Main approach: discovery; analogy; understanding, extrapolation to similar circumstances
    - Results: “This is a new problem and here is how to solve it”

## Pragmatic method & Mixed methods:

- ◆ Generally problem driven
- ◆ combines both quantitative and qualitative approaches, using right tool for a given task.

## Advocacy/participatory research methods:

- ◆ Predefined agenda, often focused on improving outcomes for a group of (marginalized) people.

# Approaches to CS research

- Attack a known problem, find/compare tools to solve it.
- Formalize a new problem, then attack it and find/compare tools to solve it.
- Build a new tool and test it on your own private data.

# Rest of Today's Agenda

- Some common tools for the course:
- Latex
- Zotero
- Git

# **Experiment Design in Computer Science**

# Computer Science

- Origins
  - Mathematics,
  - Engineering, and
  - Commercial practice.
- Evolved into
  - Theoretical,
  - Experimental and
  - Design (or user) orientated aspects.
- Balance and synthesize these aspects
- Mathematics (what is?), CS (how to?)

# **How does this Drive progress?**

1. Theoretical advances due to new mathematical results
  - You prove a theorem
  - Fun, short, no argument!
  - Not our concern here. But rarely enough.
2. Experimental results
  - The real stuff?
  - The rest of this lecture ...

# **Experimentation is Always important in Computer Research**

- Why:
  - “It doesn’t matter how beautiful your theory is, it doesn’t matter how smart you are – if it doesn’t agree with the experiment, it’s wrong.”
    - -- R. Feynman
  - “It does matter if its ‘optimal’ unless is at least good in real experiments.
    - -- T. Boult

# **What is an Experiment in Computer Science (ECS)?**

- ECS is the creation of, or the experimentation with or on, nontrivial hardware and software systems
  - These systems, taken broadly, are called computational artefacts.
- ECS process:
  - Form a hypothesis
  - Construct a model and make a prediction
  - Design an experiment and collect data
  - Analyze results

# Outline

- What is Experimental Computer Science?
- **Debunking reasons not to experiment**
- How do you do Experimental Computer Science?
- What are Information Artefacts?
- In Conclusion

Source: W Tichy,  
“Should Computer  
Scientists Experiment  
More?”, *IEEE Computer*,  
31(5), May 1998, pp. 32-  
40

# Is Computer Science really an Experimental Science?

- Computer Science is “not a science, but a synthetic, an engineering discipline” [Brooks]:
  - Phenomena are manufactured
  - CS is a type of engineering
  - So experimentation is misplaced
- But other Sciences:
  - Study manufactured entities, e.g., super-heavy elements, lasers
  - Make inferences about models, e.g. simulations

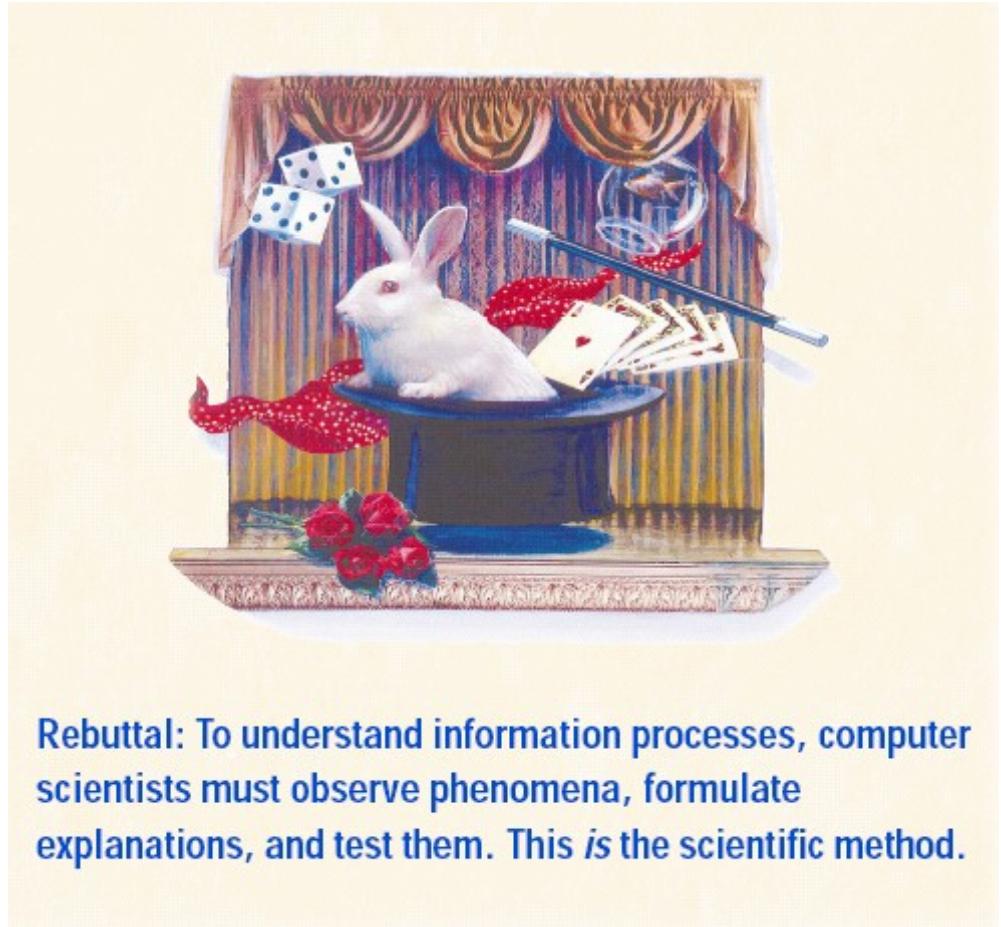


# **Why should we experiment?**

- Experiments cannot prove anything with absolute certainty
- But they are good for:
  - Reducing uncertainty about theories, models, and tools
  - Leading to new insights and whole new areas of investigation
  - Quickly eliminating fruitless approaches
  - Testing theoretical assumptions about the world.

# Fallacy #1: Traditional scientific method isn't applicable

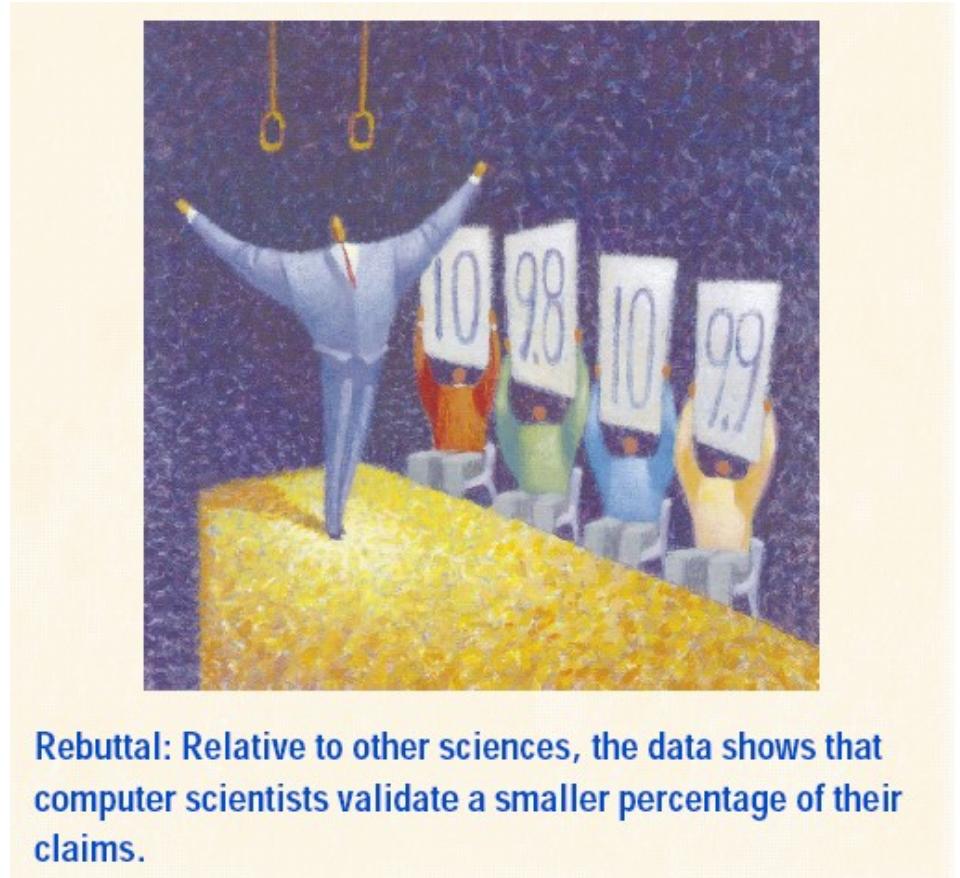
- Subject of inquiry is information unlike traditional sciences which study matter or energy
- Example:
  - Object-oriented programming, is it genuinely better?



**Rebuttal:** To understand information processes, computer scientists must observe phenomena, formulate explanations, and test them. This *is* the scientific method.

## Fallacy #2: Current levels of experimentation are enough

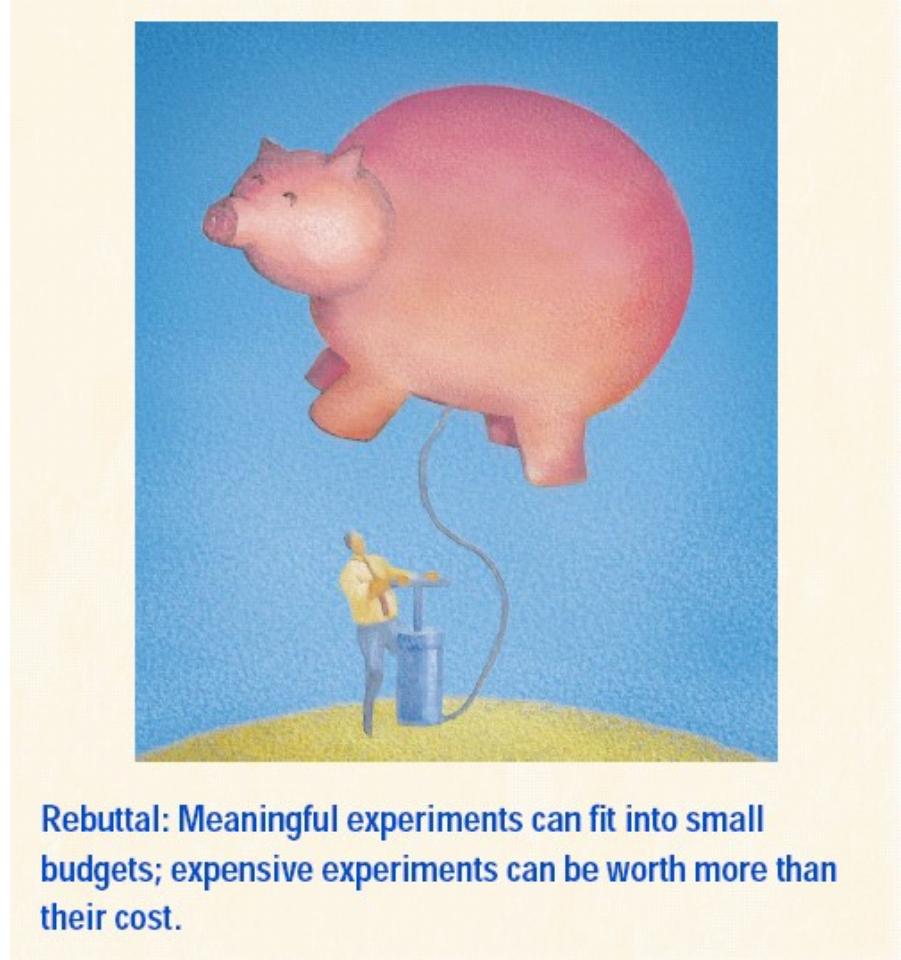
- In a late 1990s study of CS papers requiring empirical backup, 40-50% had none.
  - Compared to <15% in non-CS papers
  - The youth of CS as a discipline is not sufficient justification.
- 
- More recent studies show that is changing..



Rebuttal: Relative to other sciences, the data shows that computer scientists validate a smaller percentage of their claims.

# Fallacy #3: Experiments cost too much

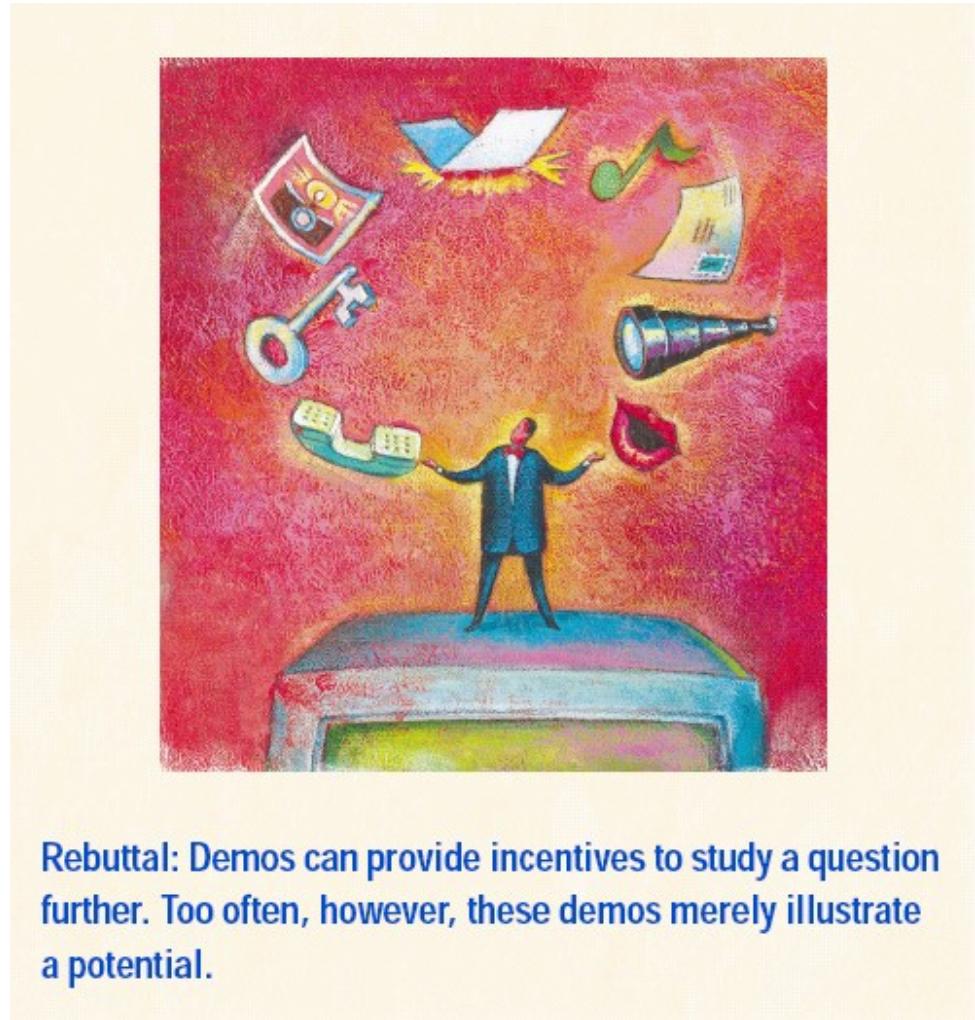
- Experiments can be expensive, but:
  - Often cheaper than the alternative
  - The cost may be worthwhile for important questions (general relativity)
  - Explore cheaper options (benchmarking)



Rebuttal: Meaningful experiments can fit into small budgets; expensive experiments can be worth more than their cost.

# Fallacy #4: Demonstration will suffice

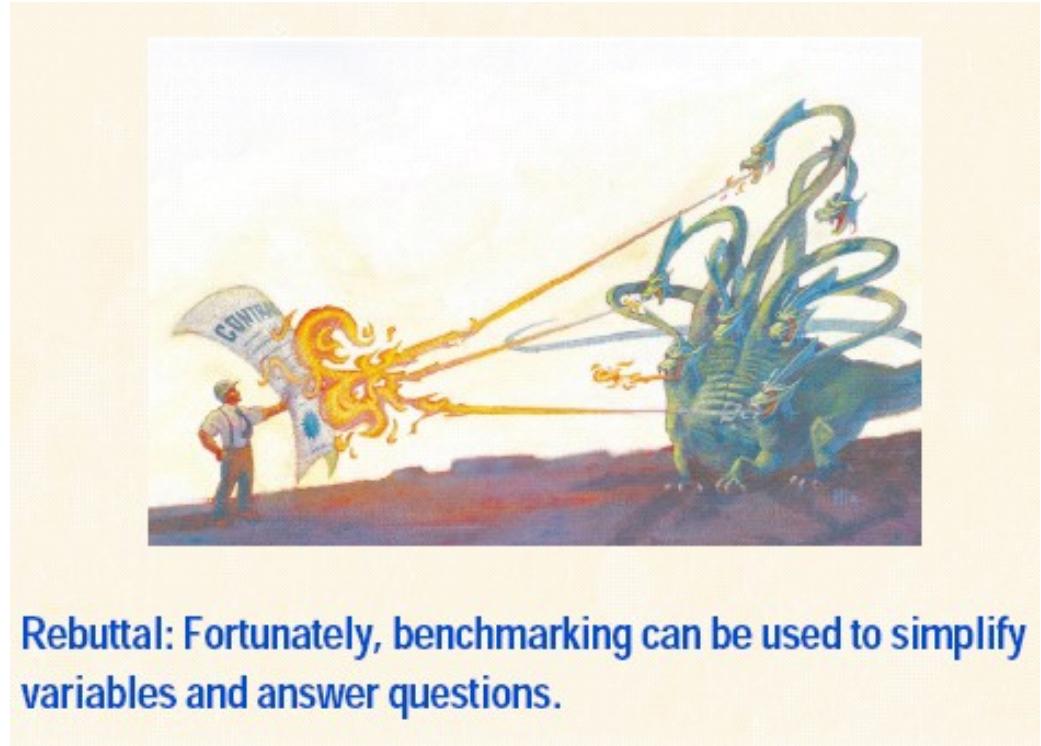
- Demos allow proof of concept and illustrate potential
- But they cannot provide solid evidence
- Easier for bias to hide in a “demo”



Rebuttal: Demos can provide incentives to study a question further. Too often, however, these demos merely illustrate a potential.

# Fallacy #5: There's too much noise in the way

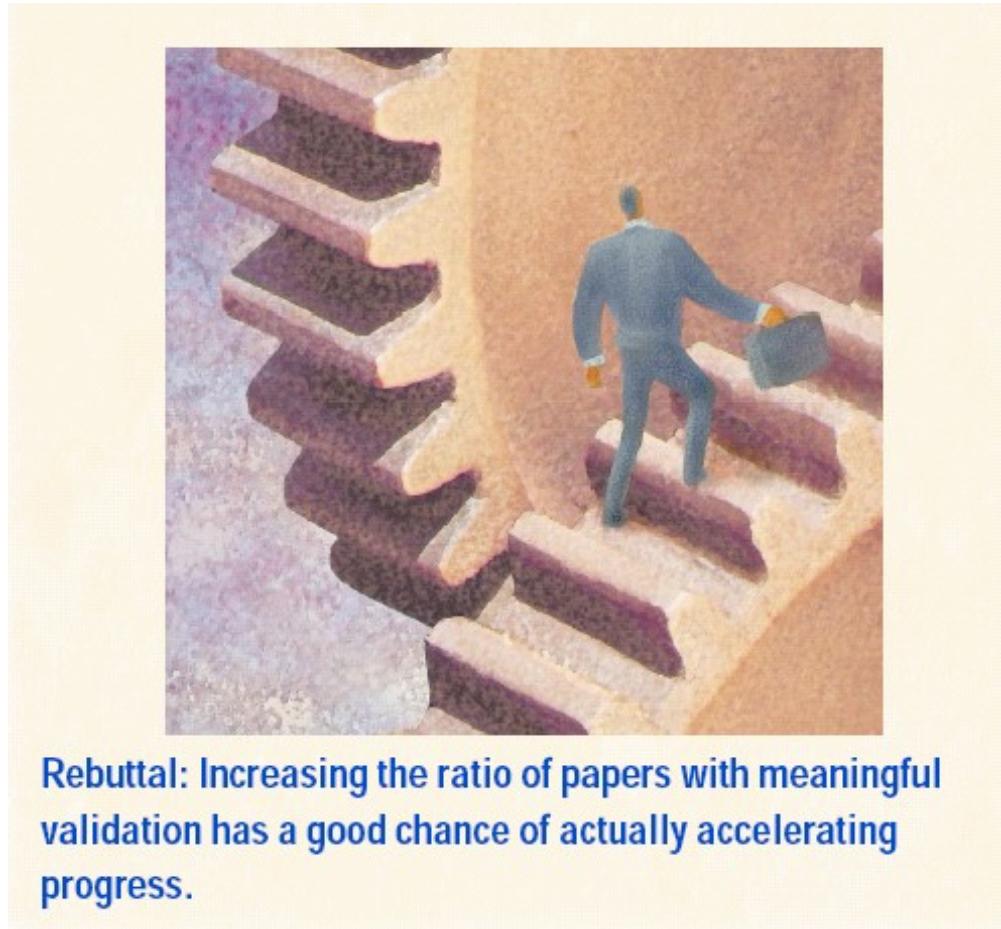
- Too many variables, effects swamped by noise
- Answers:
  - Use benchmarks
  - Apply statistical controls from medicine and psychology



Rebuttal: Fortunately, benchmarking can be used to simplify variables and answer questions.

# Fallacy #6: Experimentation will slow progress

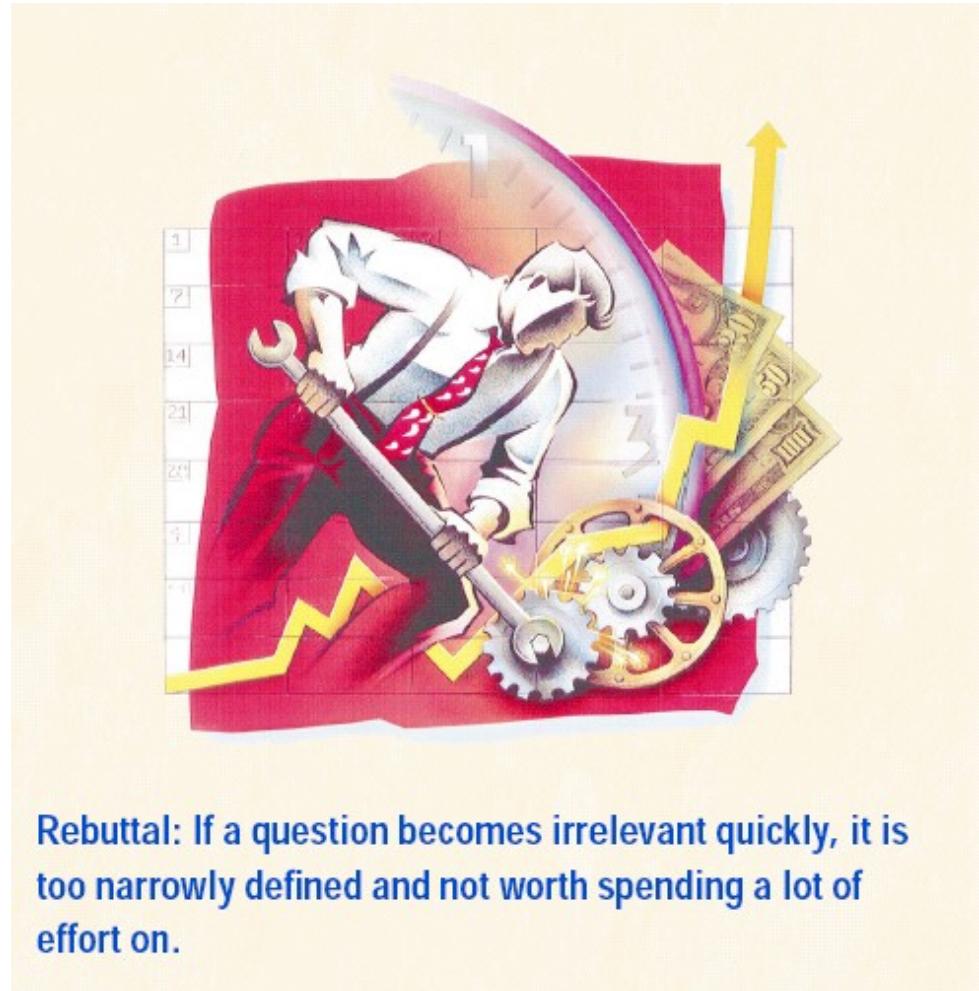
- Research takes longer → fewer ideas
- Actually weeds out questionable ideas and their offshoots
- Still a place for the hypothesis paper



Rebuttal: Increasing the ratio of papers with meaningful validation has a good chance of actually accelerating progress.

# Fallacy #7: Technology changes too fast

- “The rate of change in computing is so great that by the time results are confirmed they may no longer be of any relevance” [Mudge]
- Look to fundamental long term problems rather



**Rebuttal:** If a question becomes irrelevant quickly, it is too narrowly defined and not worth spending a lot of effort on.

# Fallacy #8: There are substitutes

## □ Theory

- Can be contradicted in practice by incorrect simplifying or modeling assumptions

## □ Intuition

- Fails in the face of counterintuitive results
- E.g., productivity is NOT necessarily improved by typechecking

## □ 37 Experts

- Science must always be backed up by evidence
- E.g., claims about cold fusion

## A Spherical Chicken

The report "Energy expenditure in animal locomotion" by Albert Gold (20 July, p. 275) may indeed pose an essentially correct solution to an important bioenergetic problem. Nevertheless, Gold's supposition, "For a sphere of unit density,  $k^{1/2} = 1.6 \text{ g}^{1/2} \text{ cm}^{-1}$ ," irresistibly calls to mind the following famous story.

A certain commercial farm was having great difficulty raising its egg productivity. Every suggestion they attempted failed to increase the output of its hens. For 1 year they tried special feed formulas, hormones, minerals in the hens' drinking water, piped in music (rock and classical), soft lights, ambient temperature variation, and even specially imported roosters, all with the same notable lack of success. In desperation, they finally took the suggestion of an extension officer to hire a theoretical physicist. After 3 more months of agonized waiting, the theoretician announced to the anxious farmers that he had the solution to their egg problem. He strode up to the blackboard and confidently began, "Postulate a spherical chicken . . ."

STEVEN D. STELLMAN

Department of Biochemical Sciences,  
Princeton University,  
Princeton, New Jersey 08540



Stellman, Steven.

"*A Spherical Chicken*". *Science* 28 Dec 1973: Vol. 182, Issue 4119, pp. 1296-1297.

<https://science.Scienmag.org/content/182/4119/1296.3>

# Outline

- What is Experimental Computer Science?
- Debunking reasons not to experiment
- How do you do Experimental Computer Science?**

What your research supposedly looks like:

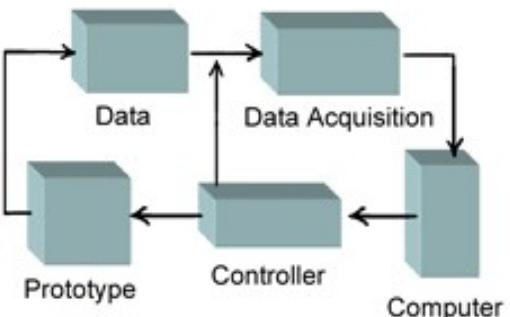
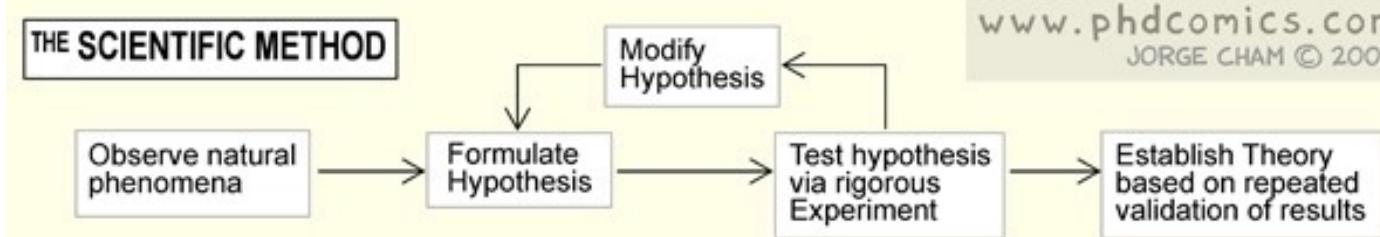


Figure 1. Experimental Diagram

What your research *actually* looks like:

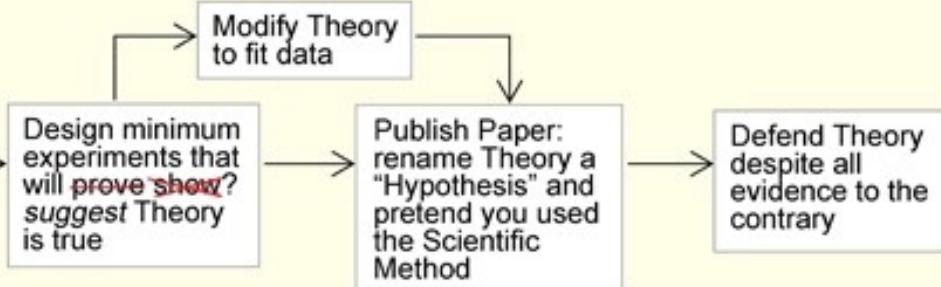


Figure 2. Experimental Mess



THE ACTUAL METHOD

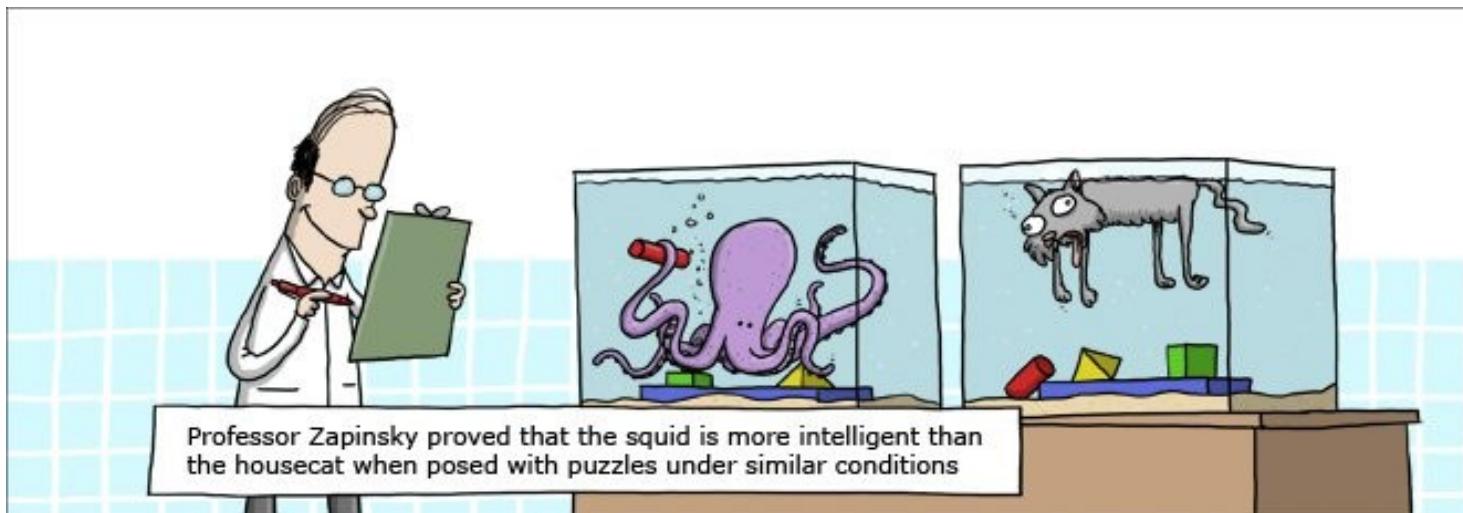
Make up Theory based on what Funding Agency Manager wants to be true



# **Experimental Hypothesis**

- Any scientific research should state a hypothesis, then
  - Provide evidence for/against
  - Conclude whether it is supported or refuted
- Should be:
  - Precise, explicit statement
  - Falsifiable
- Run experiment to confirm or refute
- Ensure that the experiment really tests the hypothesis in a meaningful way

# Experimental Design



# Elaborating the Hypothesis

- Hypothesis might be that system/theory/technique/ parameter P is:
  - Good for task X
  - Better than rivals Q and R for task X
- According to:
  - Behavior - correctness or quality of solution
  - Coverage - range of problems to which it applies
  - Efficiency - resources consumed
- Evidence can be theoretical, experimental or both
  - Theoretical evidence - theorem based
  - Experimental evidence - testing on a range of examples

# Designing an Experiment

- Specification needs to be complete and explicit
- Make sure the experiment really tests the hypothesis
- Plan for managing variance
- Requirements:
  - Controlled - other factors must be kept constant
  - Quantitative - provide numbers
  - Coverage - are tests representative of the full range of the hypothesis

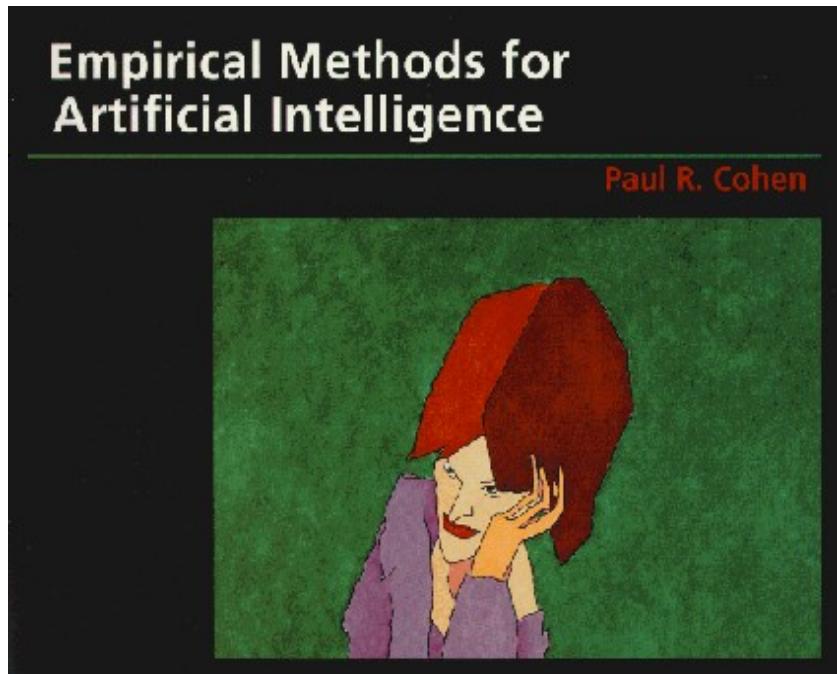
# Analyzing and Reporting Results

- Analyze the measured data:
  - Does statistical evidence really support (or refute) the hypothesis?
  - Make sure differences are not due to chance or natural variability
- Be Careful:
  - Better to admit to flaws in your methodology
  - Don't generalize without adequate support
- Report everything:
  - Procedures, results and conclusions
  - So that others can replicate the experiment
  - And build on your conclusions

# Message - Prove your Claims

- **which** means that you have to have claims to prove
- **and** evidence to back you up
- **and** evidence is almost always convincing numbers, e.g. statistically significant
- **from** well constructed, all influences considering, set of experiments
- **that** are discussed
- **and** from which a series of conclusions are drawn

# A Good read.. Cohen's Book..



Textbook, MIT Press, 1995

Other material:

Empirical Methods [Tutorial](#) at the Pacific Rim AI Conference, 2008

[Assessing the Intelligence of Cognitive Decathletes](#). Paul Cohen.  
Presented at the NIST Workshop on Cognitive Decathlon.  
Washington DC. January 2006.

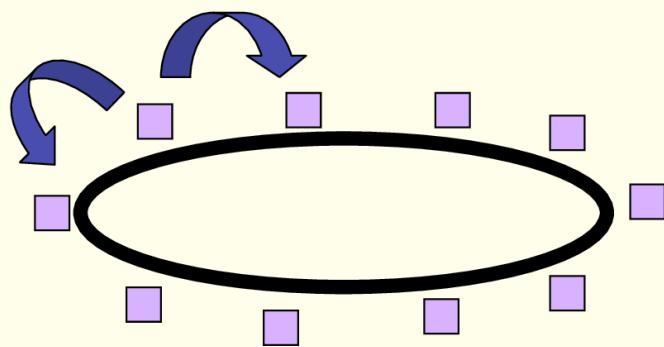
[If Not the Turing Test, Then What?](#) Paul Cohen. Invited Talk at the  
National Conference on Artificial Intelligence. July, 2004.

Various [papers](#) on empirical methods.

# Lesson 1: Evaluation begins with claims

- The most important, most immediate and most neglected part of evaluation plans.
- What you measure depends on what you want to know, on what you claim.
- Claims:
  - X is bigger/faster/stronger than Y
  - X varies linearly with Y in the range we care about
  - X and Y agree on most test items
  - It doesn't matter who uses the system (no effects of subjects)
  - My algorithm scales better than yours (e.g., a relationship between size and runtime depends on the algorithm)
- Non-claim: I built it and it runs fine on some test data

## Case Study: Comparing two algorithms



- Scheduling processors on ring network; jobs spawned as binary trees
- KOSO: keep one, send one to my left or right arbitrarily
- KOSO\*: keep one, send one to my least heavily loaded neighbor

Theoretical analysis went only so far, for unbalanced trees and other conditions it was necessary to test KOSO and KOSO\* empirically

"An Empirical Study of Dynamic Scheduling on Rings of Processors" Gregory, Gao, Rosenberg & Cohen Proc. of 8th IEEE Symp. on Parallel & Distributed Processing, 1996

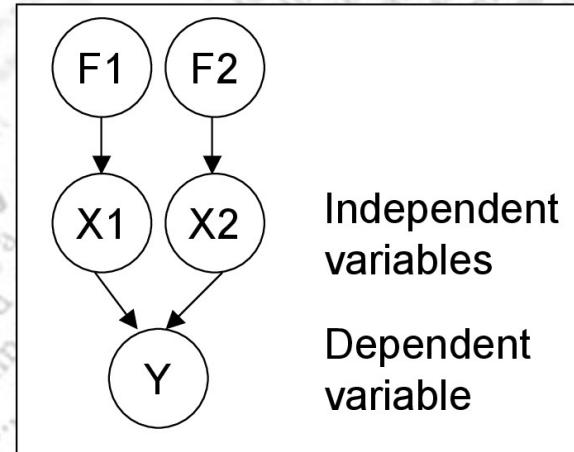
## Evaluation begins with claims

- Hypothesis (or claim): KOSO takes longer than KOSO\* *because* KOSO\* balances loads better
  - The “because phrase” indicates a hypothesis about why it works. This is a better hypothesis than the “beauty contest” demonstration that KOSO\* beats KOSO
- Experiment design
  - *Independent variables*: KOSO v KOSO\*, no. of processors, no. of jobs, probability job will spawn,
  - *Dependent variable*: time to complete jobs

## Useful Terms

**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables

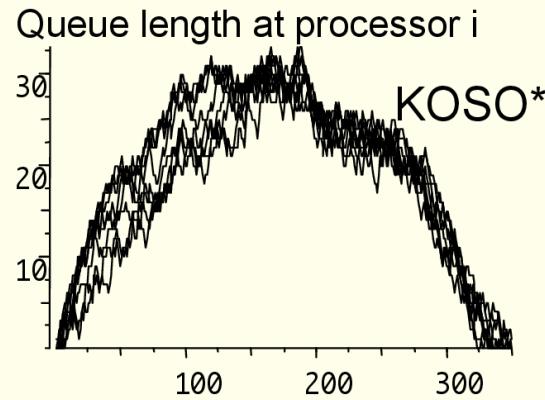
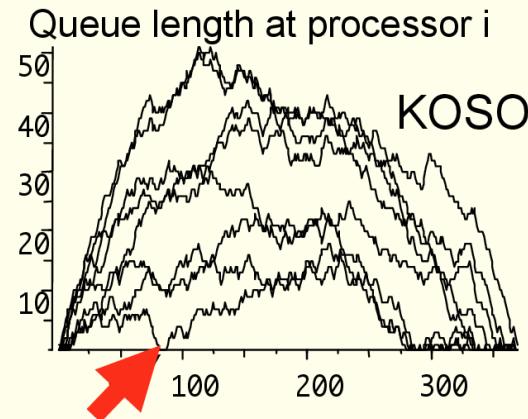


## Initial Results

- Mean time to complete jobs:
  - KOSO: 2825      (the "dumb" algorithm)
  - KOSO\*: 2935      (the "load balancing" algorithm)
- KOSO is actually 4% *faster* than KOSO\* !
- This difference is not statistically significant (more about this, later)
- What happened?

## Lesson 2: *Exploratory data analysis means looking beneath results for reasons*

- Time series of queue length at different processors:

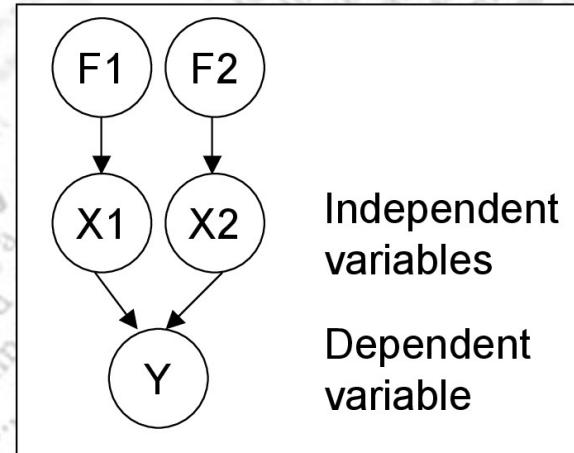


- Unless processors starve (red arrow) there is no advantage to good load balancing (i.e., KOSO\* is no better than KOSO)

## Useful Terms

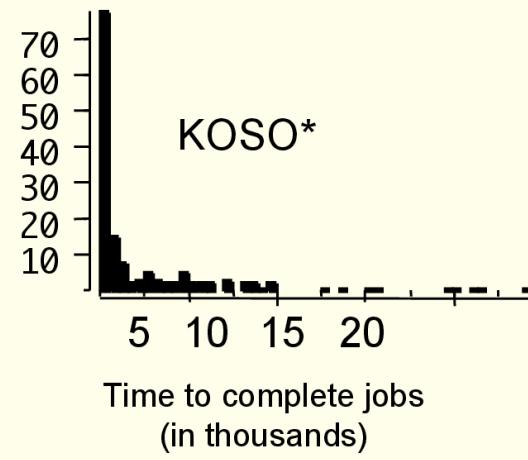
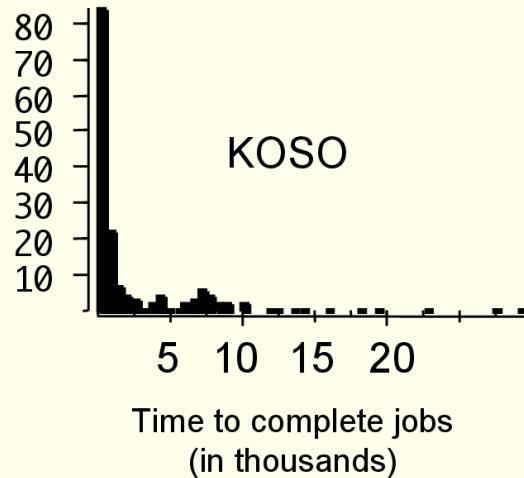
**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables



## Lesson 2: Exploratory data analysis means looking beneath results for reasons

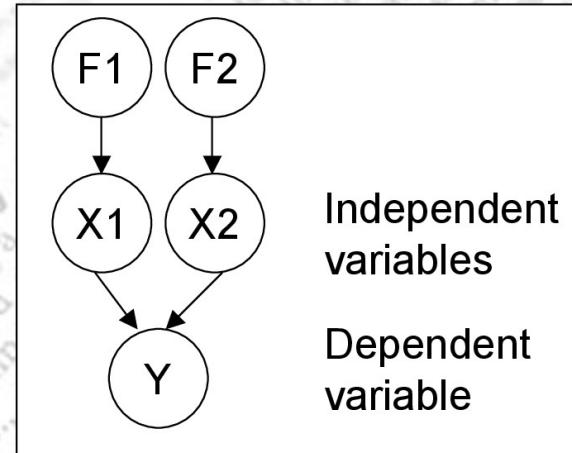
- KOSO\* is statistically no faster than KOSO. Why????



## Useful Terms

**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables



## More exploratory data analysis

- Mean time to complete jobs:  
KOSO: 2825  
KOSO\*: 2935
- Median time to complete jobs  
KOSO: 498.5  
KOSO\*: 447.0
- Looking at means (with outliers) KOSO\* is 4% slower  
but looking at medians (robust against outliers) it is 11% faster.

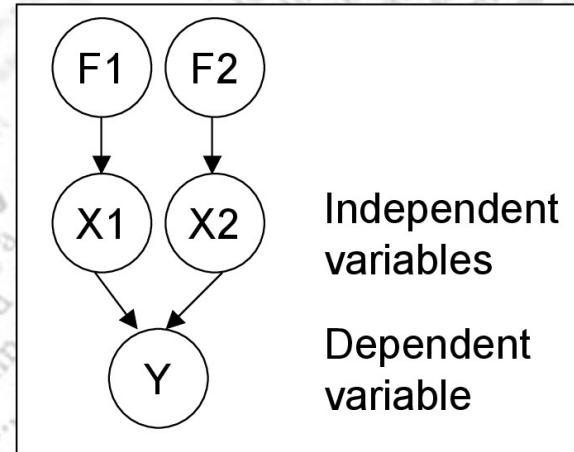
## More exploratory data analysis

- Mean time to complete jobs:  
KOSO: 2825  
KOSO\*: 2935
- Median time to complete jobs  
KOSO: 498.5  
KOSO\*: 447.0
- Looking at means (with outliers) KOSO\* is 4% slower  
but looking at medians (robust against outliers) it is 11% faster.

## Useful Terms

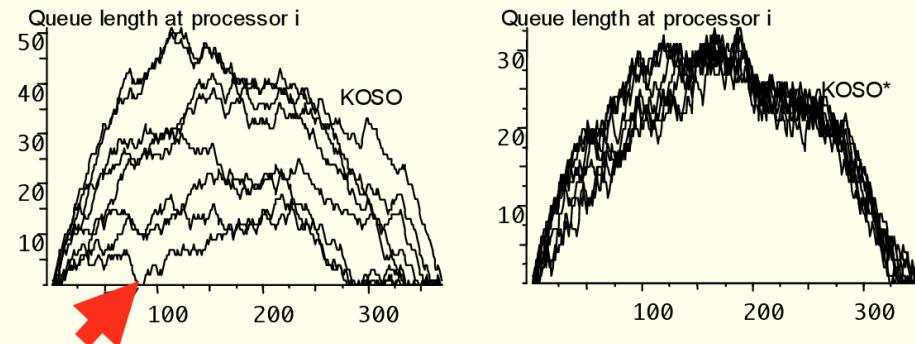
**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables



## How are we doing?

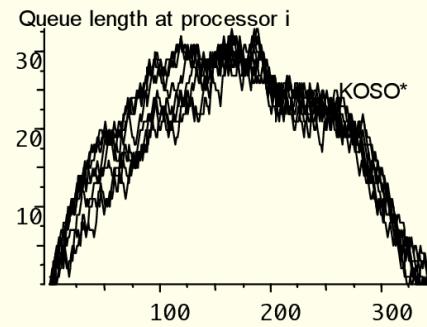
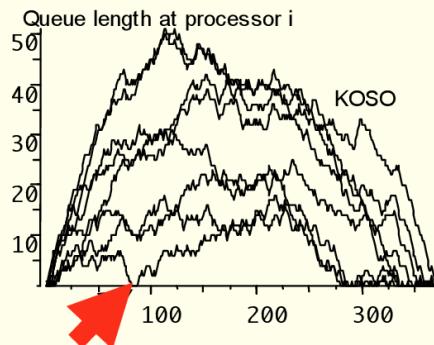
- Hypothesis (or claim): KOSO takes longer than KOSO\* *because* KOSO\* balances loads better
- Mean KOSO is shorter than mean KOSO\*, median KOSO is longer than KOSO\*, no evidence that load balancing helps because there is almost no processor starvation in this experiment.
- Now what?



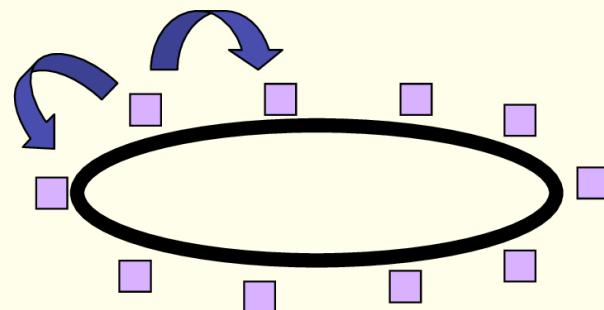
## Lesson 3: Always run pilot experiments

- A pilot experiment is designed less to test the hypothesis than to test the experimental apparatus to see whether it *can* test the hypothesis.
- Our independent variables were not set in a way that produced processor starvation so we couldn't test the hypothesis that KOSO\* is better than KOSO because it balances loads better.
- Use pilot experiments to adjust independent and dependent measures, see whether the protocol works, provide preliminary data to try out your statistical analysis, in short, test the *experiment design*.

## Next steps in the KOSO / KOSO\* saga...



**It looks like KOSO\* does balance loads better (less variance in the queue length) but without processor starvation, there is no effect on run-time**

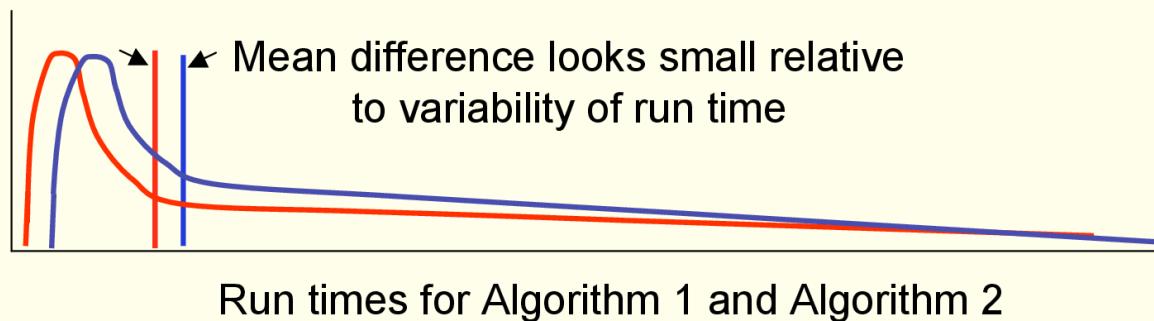


**We ran another experiment, varying the number of processors in the ring: 3, 9, 10 and 20**

**Once again, there was no significant difference in run-time**

## Variance-reducing transforms

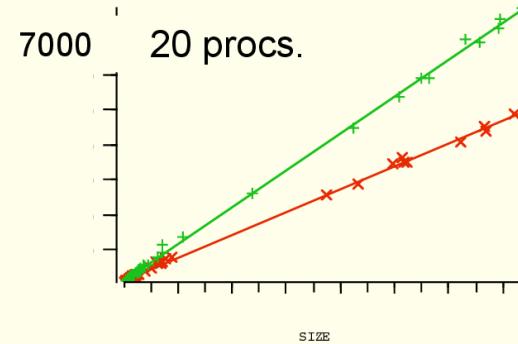
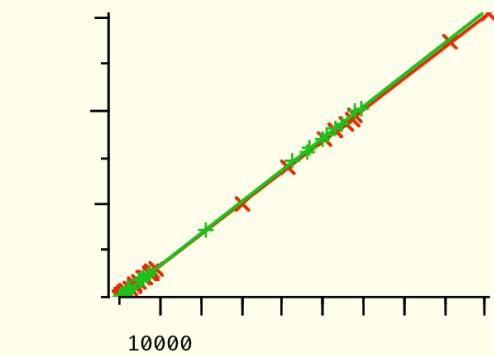
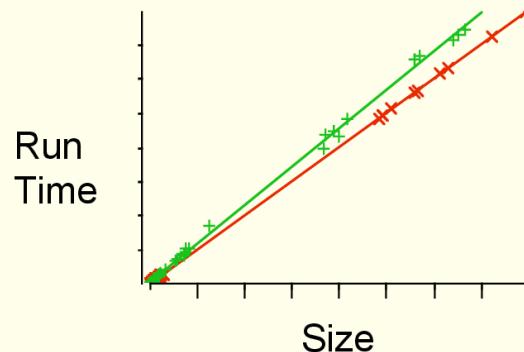
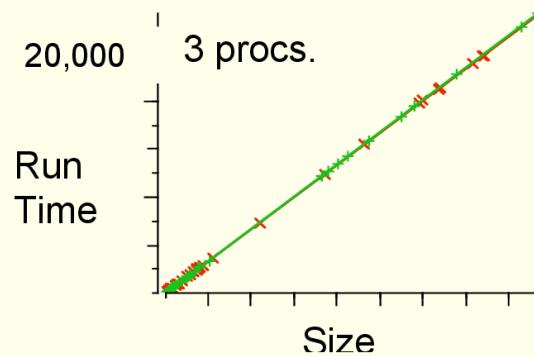
- Suppose you are interested in which algorithm runs faster on a batch of problems but the run time depends more on the problems than the algorithms



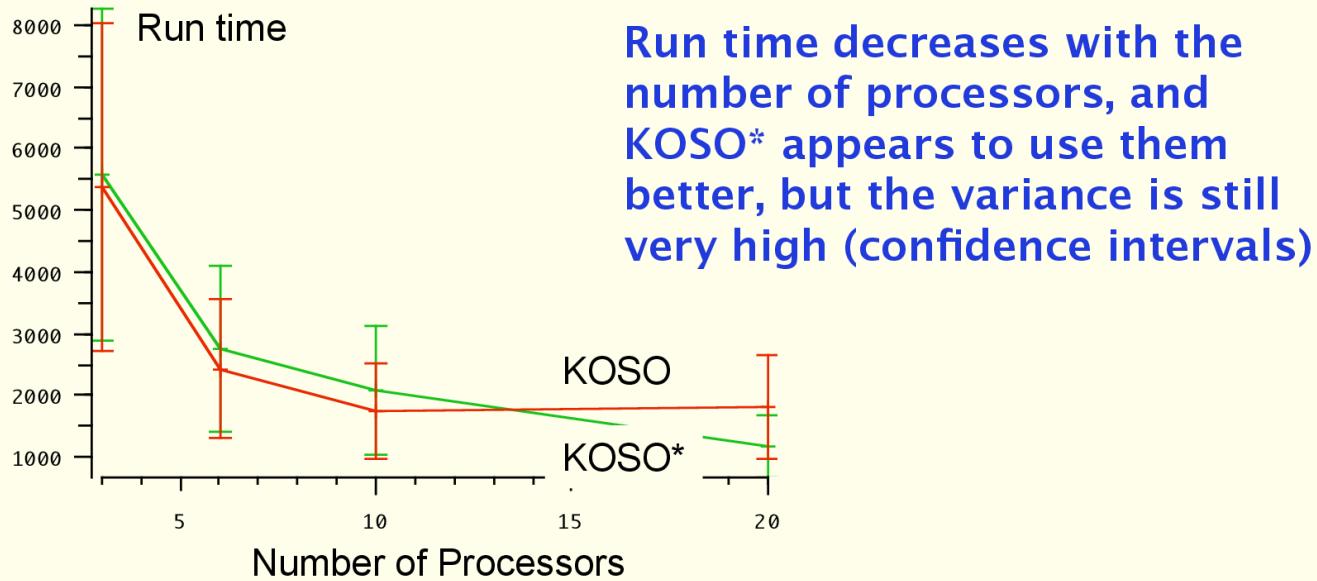
- You don't care very much about the problems, so you'd like to transform run time to "correct" the influence of the problem. This is one kind of *variance-reducing transform*.

## What causes run times to vary so much?

Run time depends on the number of processors and on the number of jobs (size). The relationships between these and run time are different for KOSO and KOSO\*. Green: KOSO Red: KOSO\*



## What causes run times to vary so much?



Run time decreases with the number of processors, and KOSO\* appears to use them better, but the variance is still very high (confidence intervals)

- Can we transform run time with some function of the number of processors and the problem size?

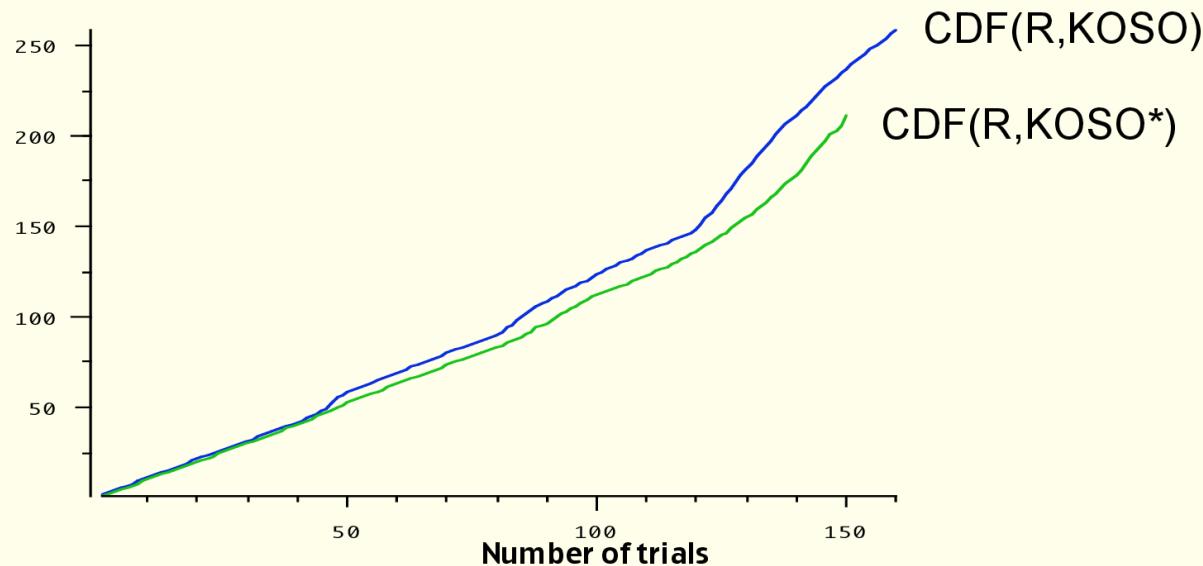
## Transforming run time

- Let  $S$  be the number of tasks to be done
- Let  $N$  be the number of processors to do them
- Let  $T$  be the time required to do them all (run time)
- So  $k_i = S_i/N_i$  is the theoretical best possible run time on task  $i$  (i.e., perfect use of parallelism)
- So  $T_i / k_i$  is how much worse than perfect a particular run time is
- The transform we want is  $R_i = (T_i N_i) / S_i$ . This restates the run time in a way that's independent of the size of the problem and the number of processors, both of which caused variance.

## A small difference

|       | Mean | Median |
|-------|------|--------|
| KOSO  | 1.61 | 1.18   |
| KOSO* | 1.40 | 1.03   |

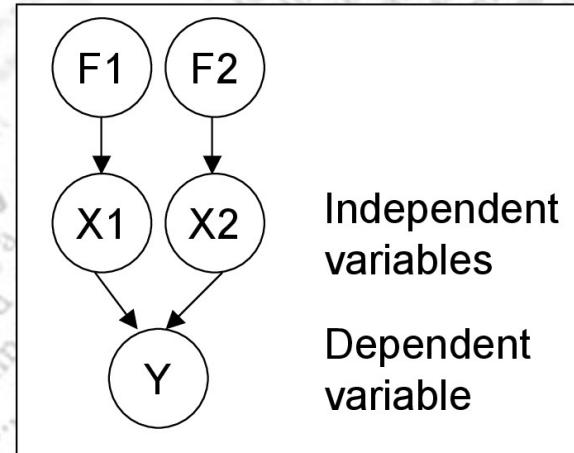
Median KOSO\* is almost perfectly efficient



## Useful Terms

**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables



## The logic of statistical hypothesis testing

1. Assume KOSO = KOSO\*

2. Run an experiment to find the sample statistics

$$R_{\text{koso}} = 1.61, R_{\text{koso}^*} = 1.4, \text{ and } \Delta = 0.21$$

3. Find the distribution of  $\Delta$  under the assumption KOSO = KOSO\*

4. Use this distribution to find the probability  $p$  of  $\Delta = 0.21$  if KOSO = KOSO\*

5. If the probability is very low (it is,  $p < .02$ ) reject KOSO = KOSO\*

6.  $p < .02$  is your residual uncertainty that KOSO *might* equal KOSO\*

difference between the means

probability of this result if the difference  
between the means were truly zero

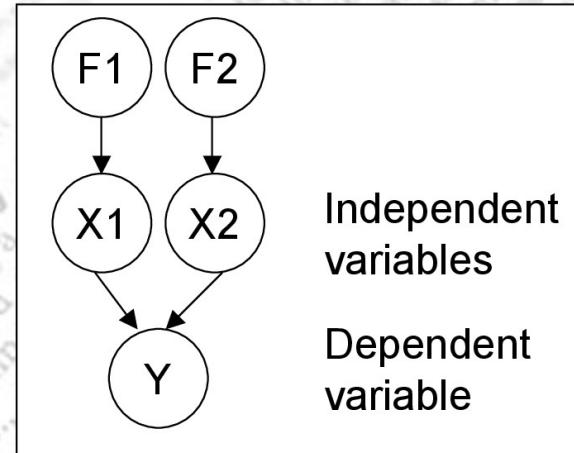
$$t = \frac{1.61 - 1.4}{.084} = 2.49, p < .02$$

estimate of the variance of the difference between the means

## Useful Terms

**Independent variable:** A variable that indicates something you manipulate in an experiment, or some supposedly causal factor that you can't manipulate such as gender (also called a **factor**)

**Dependent variable:** A variable that indicates to greater or lesser degree the causal effects of the factors represented by the independent variables



## A statistically significant difference!

|       | Mean | Standard deviation |
|-------|------|--------------------|
| KOSO  | 1.61 | 0.78               |
| KOSO* | 1.40 | 0.7                |

Two-sample t test:

$$t = \frac{\bar{x}_{koso} - \bar{x}_{koso*}}{\hat{\sigma}(\bar{x}_{koso} - \bar{x}_{koso*})}$$

difference between the means

probability of this result if the difference  
between the means were truly zero

$$t = \frac{1.61 - 1.4}{.084} = 2.49, p < .02$$

estimate of the variance of the difference between the means

# More on Statistical Tests later in term

JELLY BEANS  
CAUSE ACNE!

SCIENTISTS!  
INVESTIGATE!

BUT WE'RE  
PLAYING  
MINECRAFT!  
...FINE.

WE FOUND NO  
LINK BETWEEN  
JELLY BEANS AND  
ACNE ( $P > 0.05$ ).

THAT SETTLES THAT.  
I HEAR IT'S ONLY  
A CERTAIN COLOR  
THAT CAUSES IT.

SCIENTISTS!  
BUT  
MINECRAFT!

WE FOUND NO  
LINK BETWEEN  
PURPLE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BROWN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
PINK JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BLUE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TEAL JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
GREY JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TAN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
CYAN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND A  
LINK BETWEEN  
GREEN JELLY  
BEANS AND ACNE  
( $P < 0.05$ ).  
WHOA!

WE FOUND NO  
LINK BETWEEN  
MAUVE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
SALMON JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
RED JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TURQUOISE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
MAGENTA JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

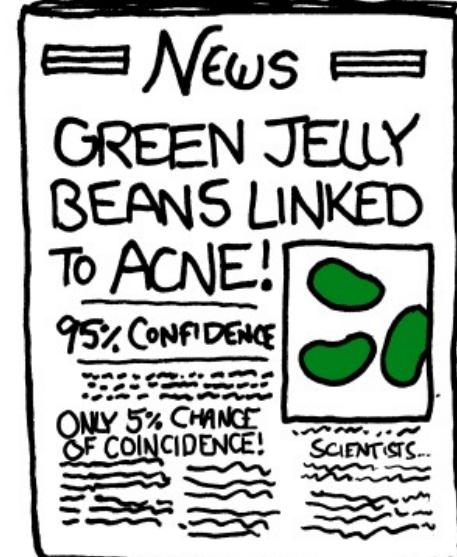
WE FOUND NO  
LINK BETWEEN  
YELLOW JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BEIGE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
LILAC JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BLACK JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

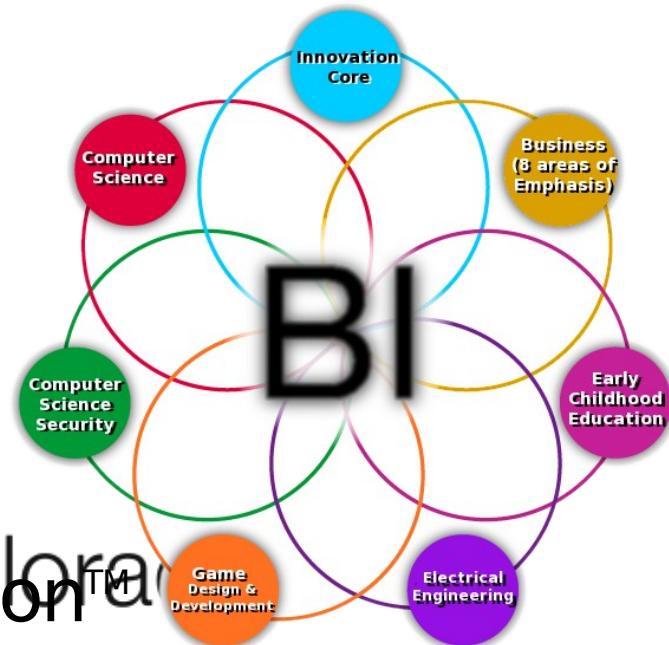
WE FOUND NO  
LINK BETWEEN  
ORANGE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).



# Some Lessons for Evaluators of “DARPA Programs”

Based on slides and the book of Paul Cohen

# Some Lessons for Evaluation of Programs



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs



# Outline

- Some general lessons about how to conduct evaluations of DARPA programs
- Some specific methodological lessons that every DARPA program manager should know
  - illustrated with a case study of a large IPTO program evaluation
- A checklist for evaluation designs

# General lessons from DARPA program evaluations

- All DARPA program evaluations serve three masters: The director, the program manager, and the research(ers).
- A well-designed evaluation gives these stakeholders what they need, but compromise is necessary and the evaluator should broker it
- The evaluator is not there to trip up the performer, but to design a test that *can* be passed. Whether it *is* passed is up to the performer.
- Start early. Ideally, the program claims, protocols and metrics are ready before the BAA/solicitation is even released.
- Keep the claims simple, but make sure there *are* claims
- Write (no Powerpoint!) the *protocol*, including claims, materials and subjects, method, planned analyses and expected results
- Run pilot experiments. Really. It's too expensive not to. Really. I mean it.
- Provide adequate infrastructure for the experiments. Don't be cheap.

# General lessons from DARPA program evaluations

- You are spending tens of millions on the program, so require the evaluation to provide more than one bit (pass/fail) of information (Lesson 5, below: demos are good, explanations better; or as Tony Tether said, “passing the test is necessary but not sufficient for continued funding.”)
- Stay flexible: Multi-year programs that test the same thing each year quickly become ossified. Review and refine claims (metrics, protocol...) annually.
- Stay flexible II: Let some parameters of the evaluation (e.g., number of subjects or test items) be set pragmatically and don’t freak if they change.
- Stay flexible III: Avoid methodological purists. Any fool can tell you why something is “not allowed” or your “sample size is wrong,” etc. A good evaluator finds workarounds and quantifies confidence.

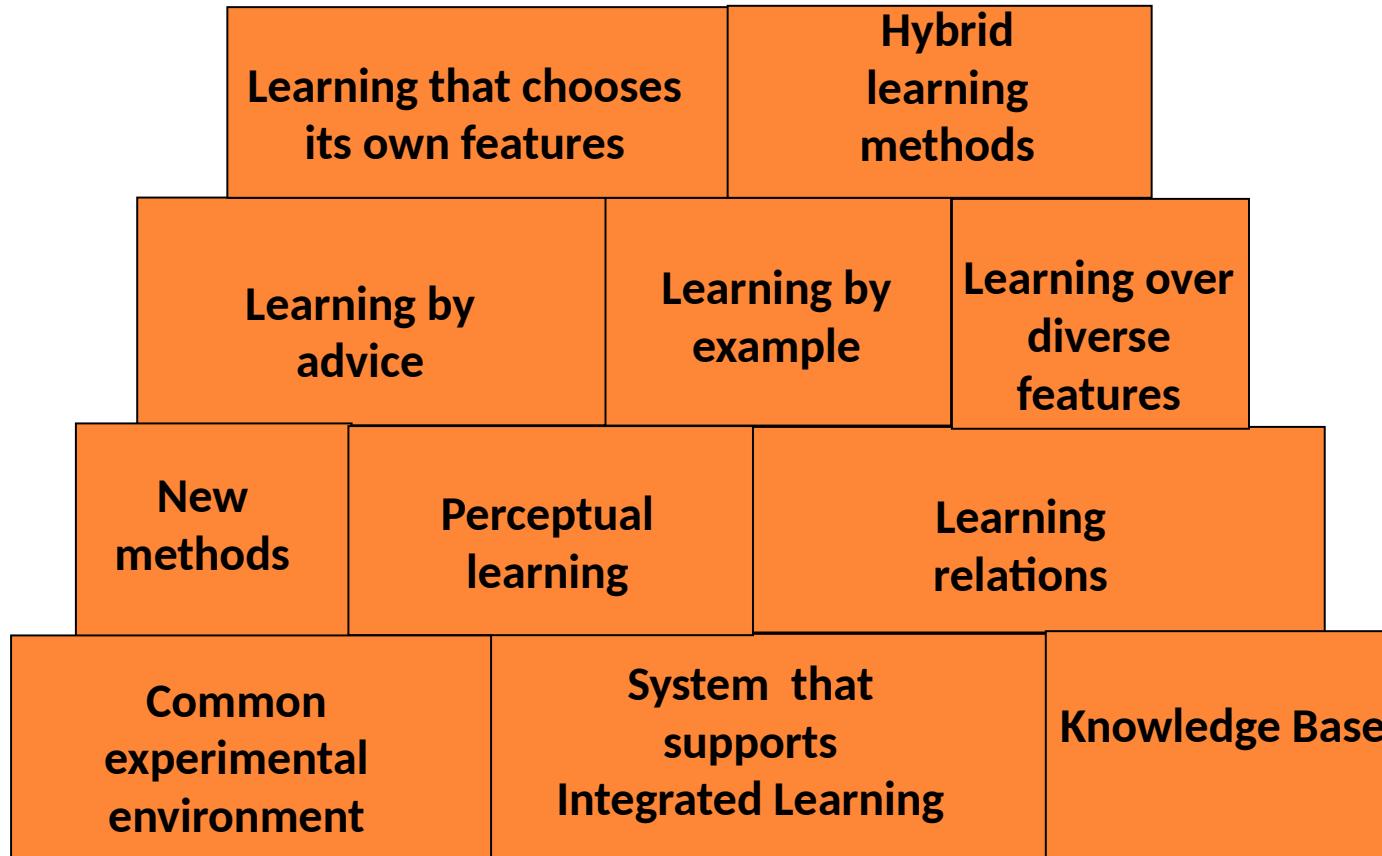
# **Some methodological lessons that every DARPA program manager should know**

1. Evaluation begins with claims; metrics without claims are meaningless
2. The task of empirical science is to explain variability
3. Humans are great sources of variability
4. Of sample variance, effect size, and sample size, control the first before touching the last
5. Demonstrations are good, explanations are better
6. Most explanations involve additional factors; most interesting science is about interaction effects, not main effects
7. Exploratory Data Analysis: use your eyes to look for explanations in data
8. Not all studies are experiments, not all analysis hypothesis testing;
9. Significant and meaningful are not synonyms

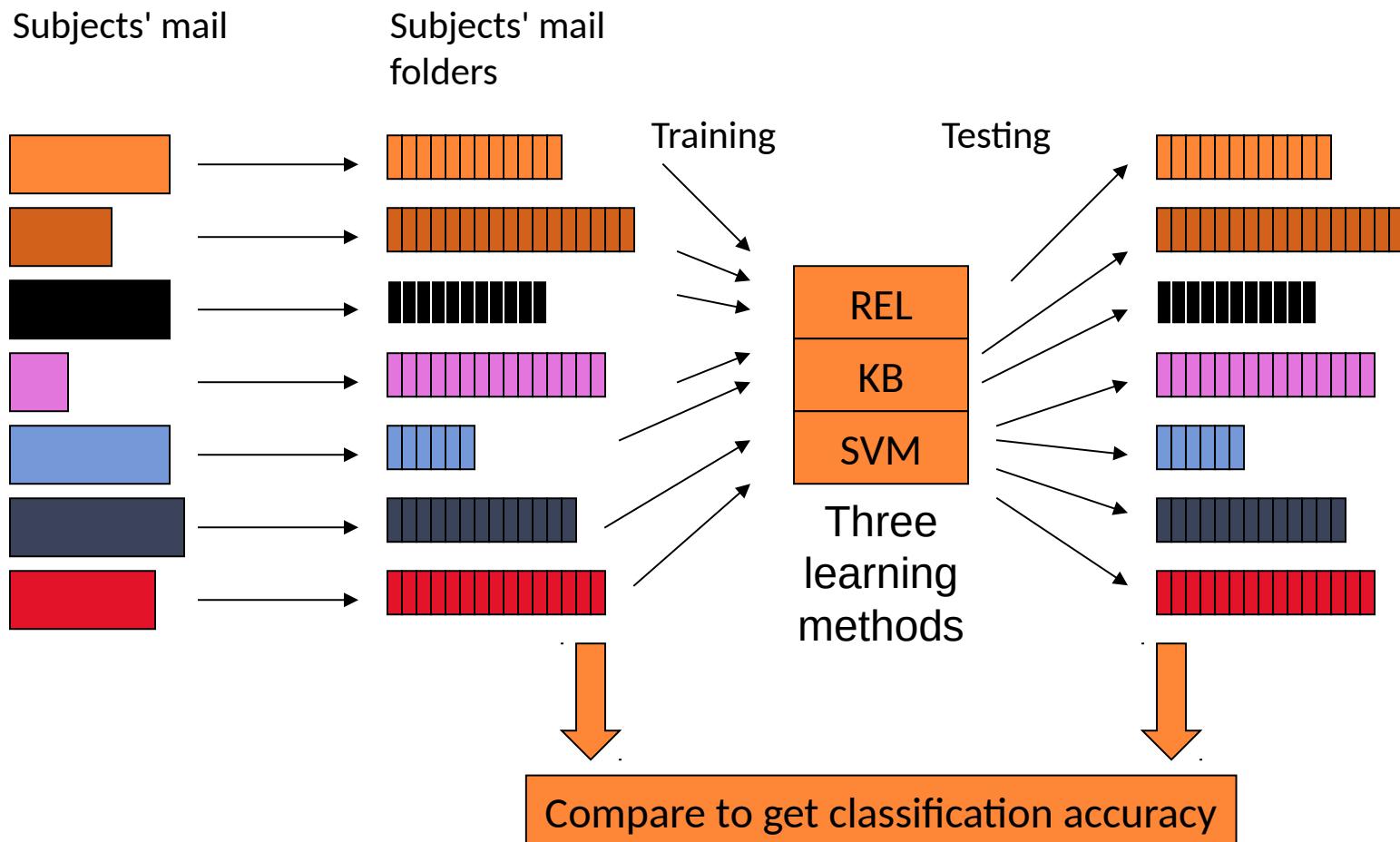
# Lesson 1: Evaluation begins with claims

- The most important, most immediate and most neglected part of evaluation plans.
- What you measure depends on what you want to know, on what you claim.
- Claims:
  - X is bigger/faster/stronger than Y
  - X varies linearly with Y in the range we care about
  - X and Y agree on most test items
  - It doesn't matter who uses the system (no effects of subjects)
  - My algorithm scales better than yours (e.g., a relationship between size and runtime depends on the algorithm)
- Non-claim: I built it and it runs fine on some test data

# THE TEAM CLAIMS THAT ITS SYSTEM PERFORMANCE IS DUE TO LEARNED KNOWLEDGE

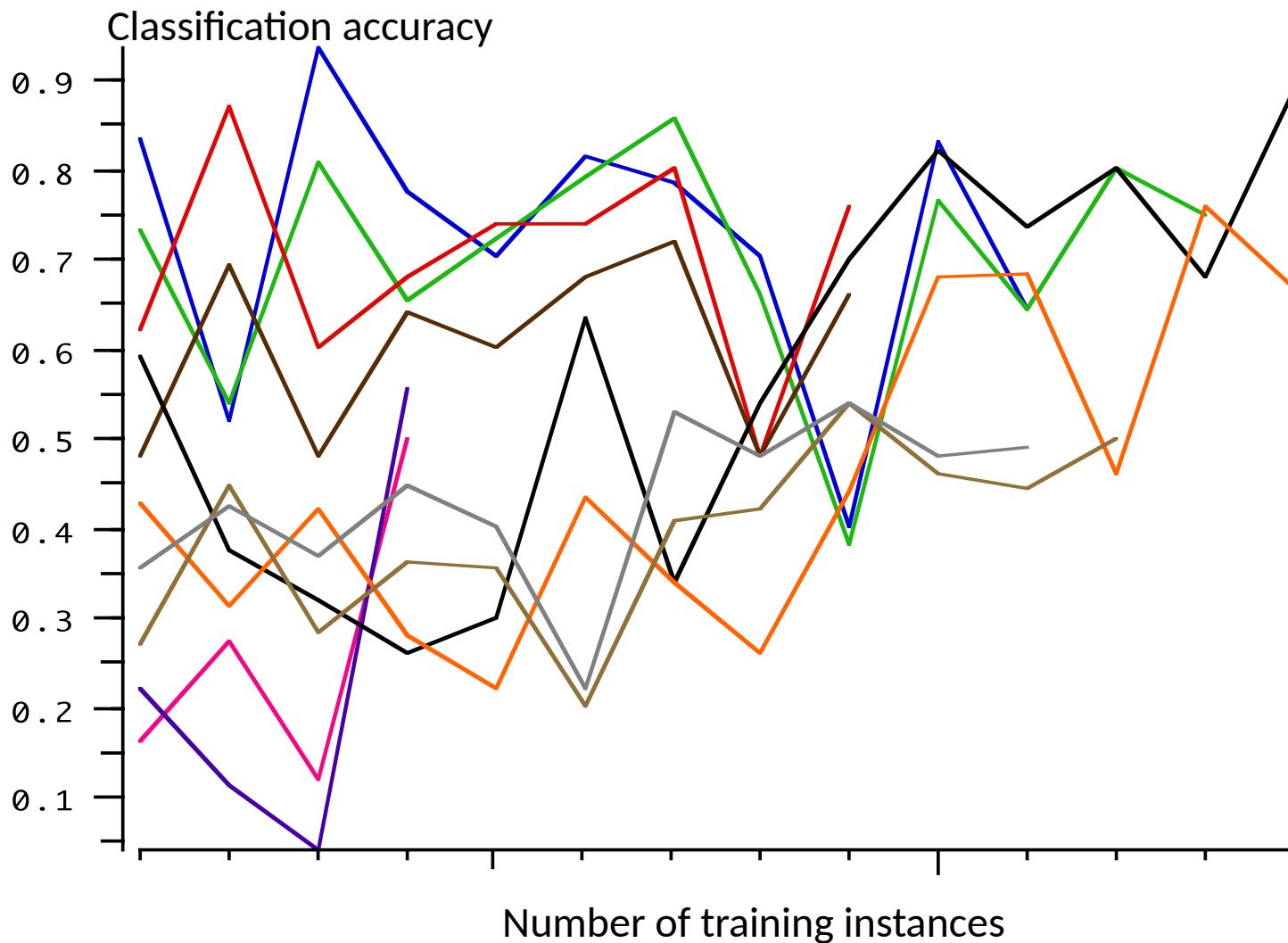


# Learning to put email in the right folders



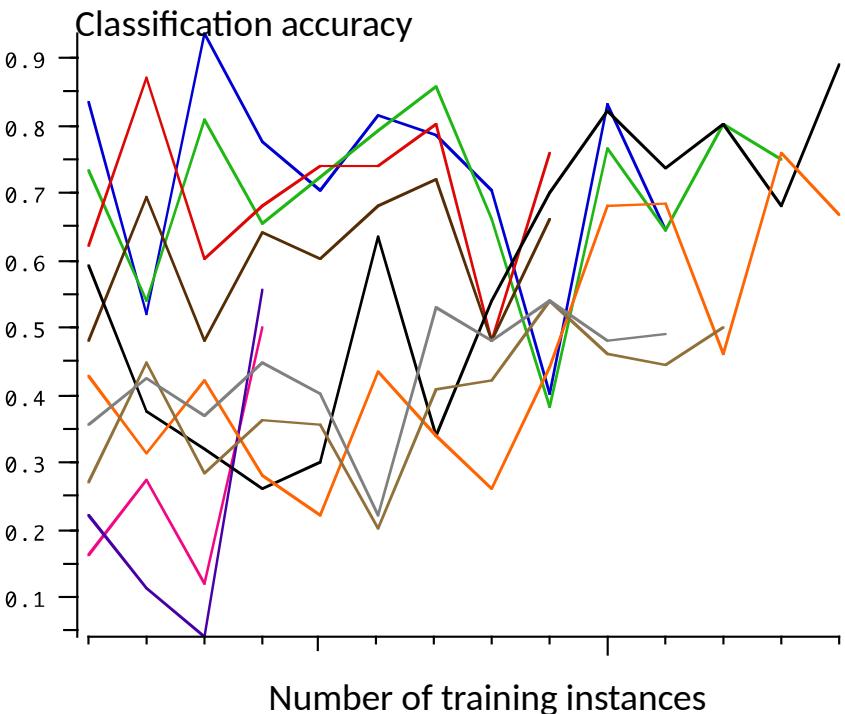
## **Lesson 2: The task of empirical science is to explain variability**

## **Lesson 3: Humans are a large source of variability**



## Lesson 2: The task of empirical science is to explain variability

## Lesson 3: Humans are a large source of variability



Why do you need statistics?

When something obviously works, you think don't need statistics

When something obviously fails, you think don't need statistics

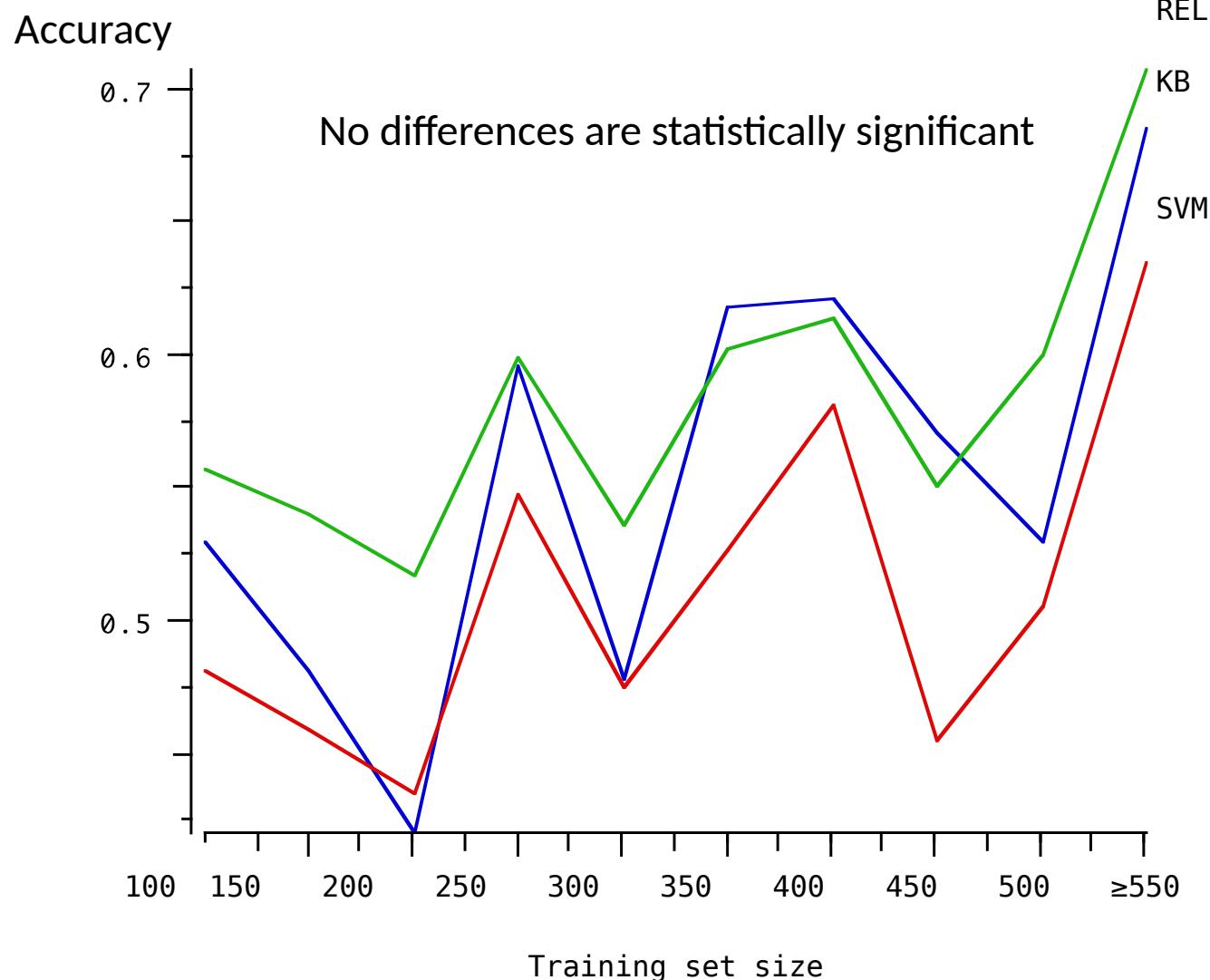
Statistics is about the ambiguous cases, where things don't obviously work or fail.

Ambiguity is generally caused by variance, some variance is caused by lack of control

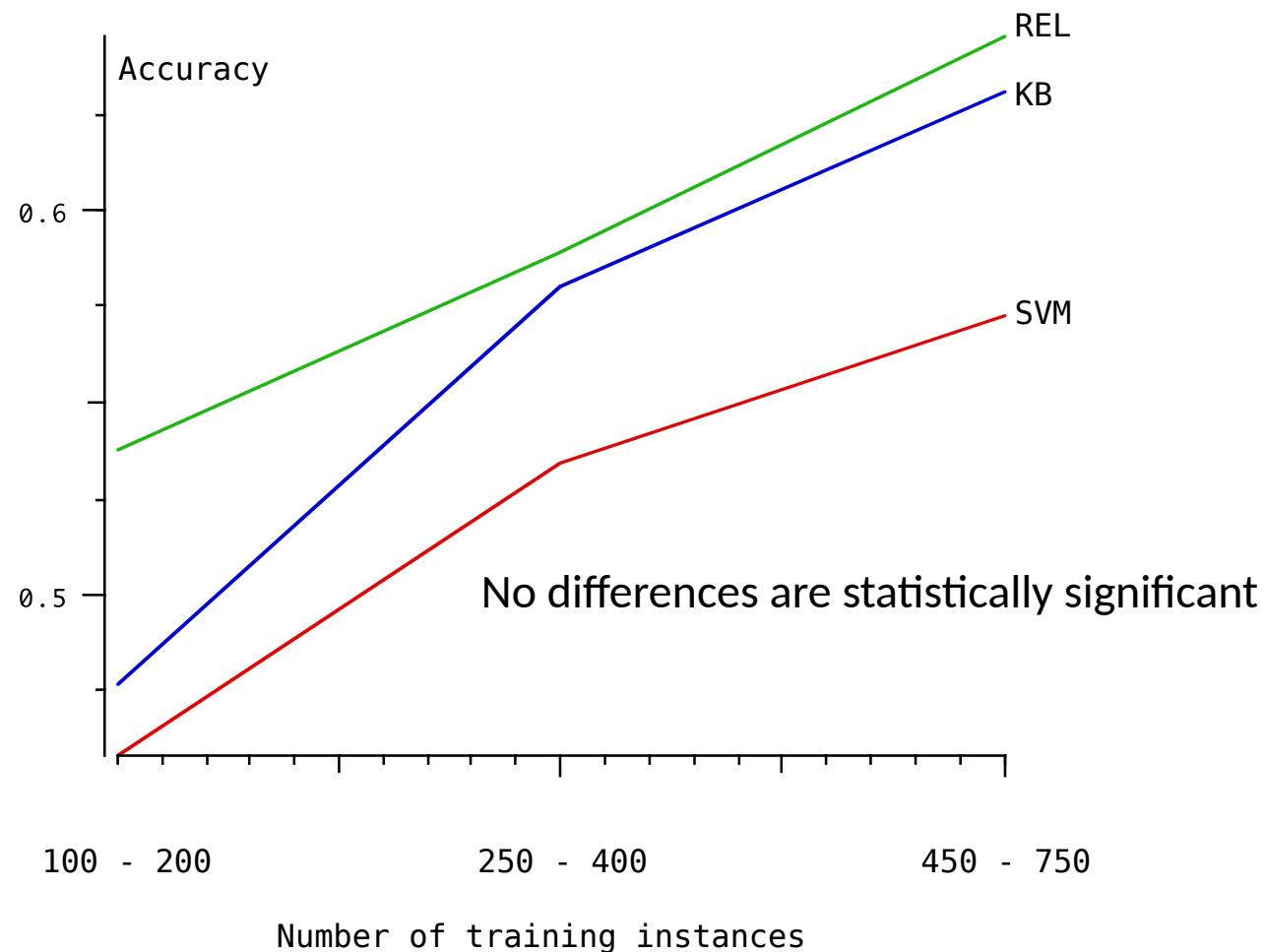
If you don't get control in your experiment design, you try to supply it post hoc with statistics

# Accuracy vs. Training Set Size

## Averaged over subject

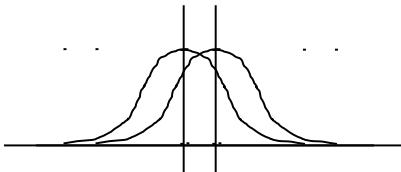


# Accuracy vs. Training Set Size (100% Coverage, Grouped)

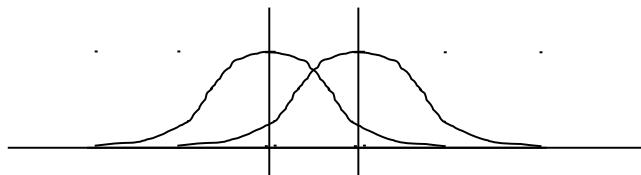


# Why are things not significantly different?

## Lesson 6: Most explanations involve additional factors



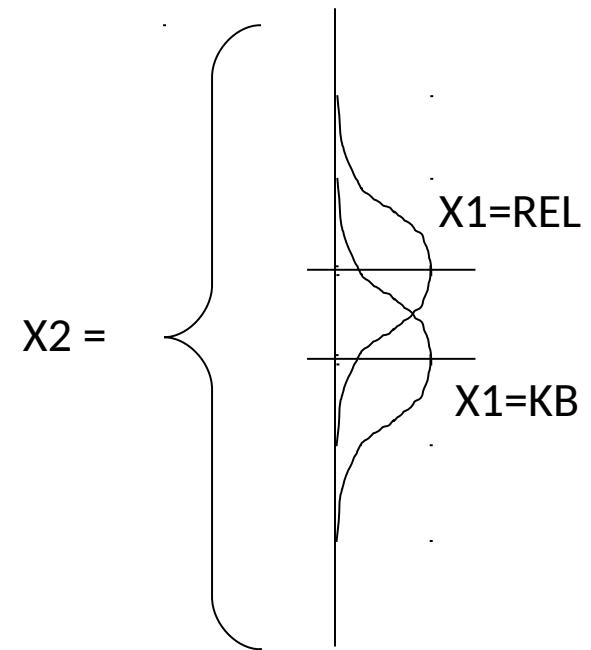
Means are close together and variance is high



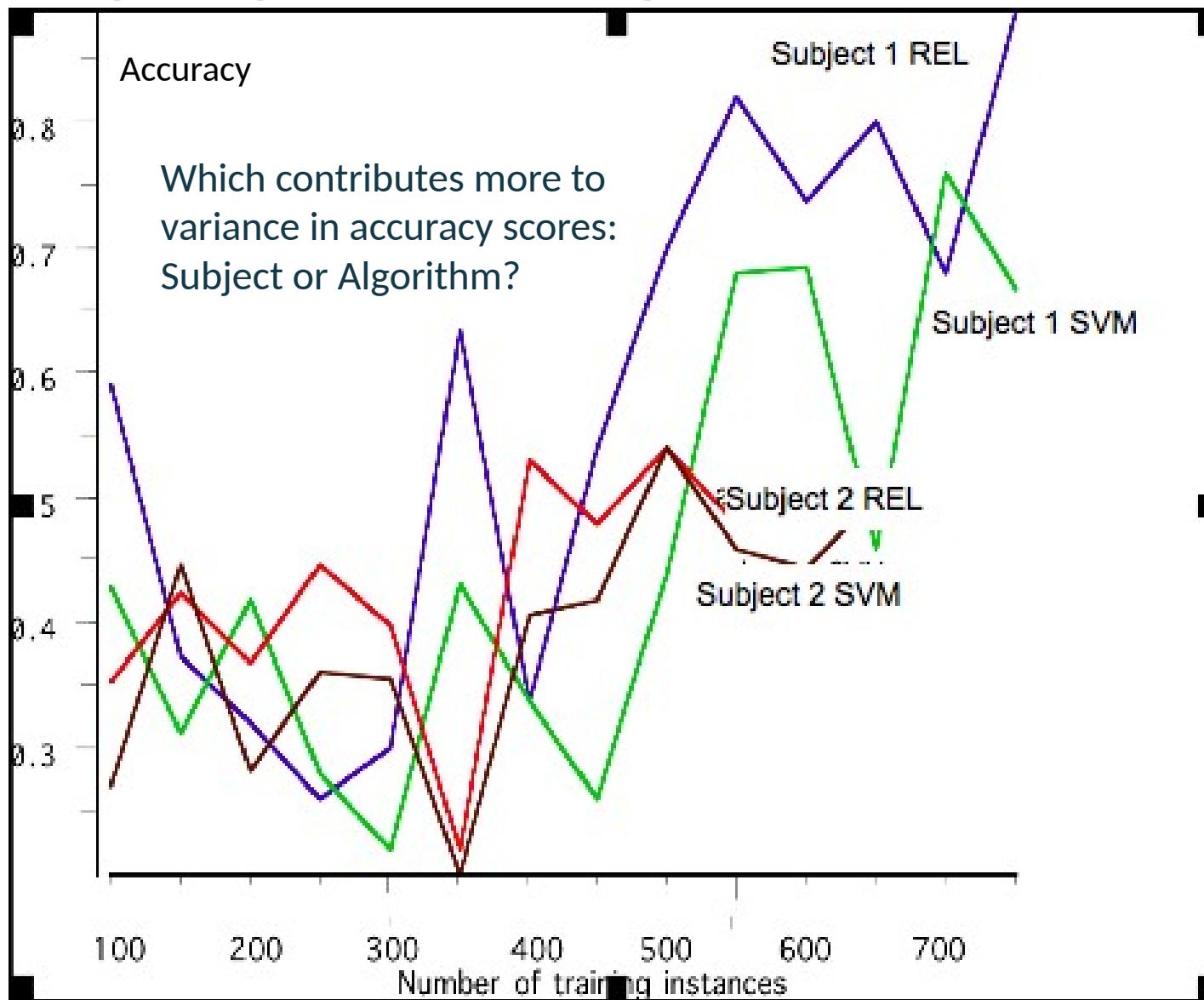
Means are far apart but variance is higher

Why is variance high? Your experiment looks at  $X_1$ , the algorithm, and  $Y$ , the score, but there is usually an  $X_2$  lurking which contributes to variance

Lesson 2: The task of empirical science is to explain/reduce variability. Find and control  $X_2$ !

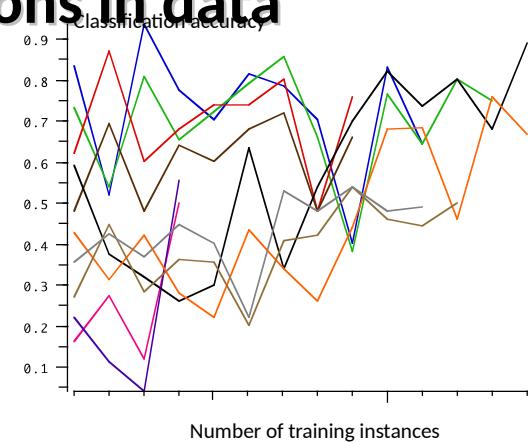


## Lesson 7: Exploratory Data Analysis means your eyes to look for explanations in data



## 7) EDA: use your eyes to look for explanations in data

- Three categories of “errors” identified
  - Mis-foldered (drag-and-drop error)
  - Non-stationary (wouldn’t have put it there now)
  - Ambiguous (could have been in other folders)
- Users found that 40% – 55% of their messages fell into one of these categories



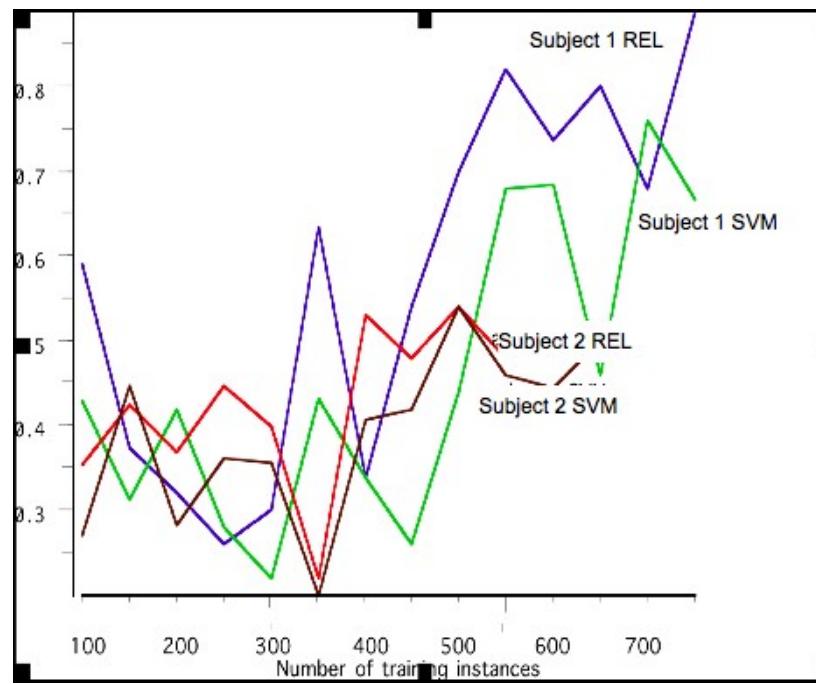
| Subject | Folders | Messages | Mis-     | Non-       |           |
|---------|---------|----------|----------|------------|-----------|
|         |         |          | Foldered | Stationary | Ambiguous |
| 1       | 15      | 268      | 1%       | 13%        | 42%       |
| 2       | 15      | 777      | 1%       | 24%        | 16%       |
| 3       | 38      | 646      | 0%       | 7%         | 33%       |

EDA tells us the problem: We're trying to find differences between algorithms when the gold standards are themselves errorful – but in different ways, increasing variance!

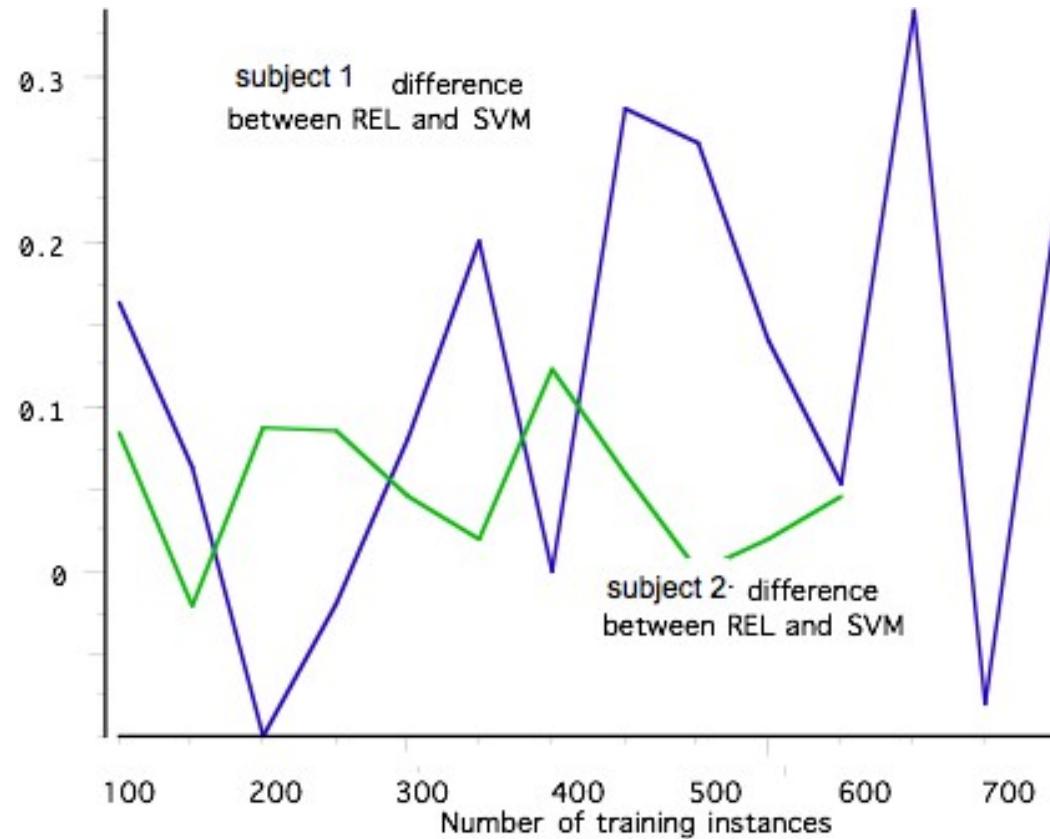
## Lesson 4: Of sample variance, effect size, and sample size, control the first before touching the last. (You cannot control effect size!)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{N}}}$$

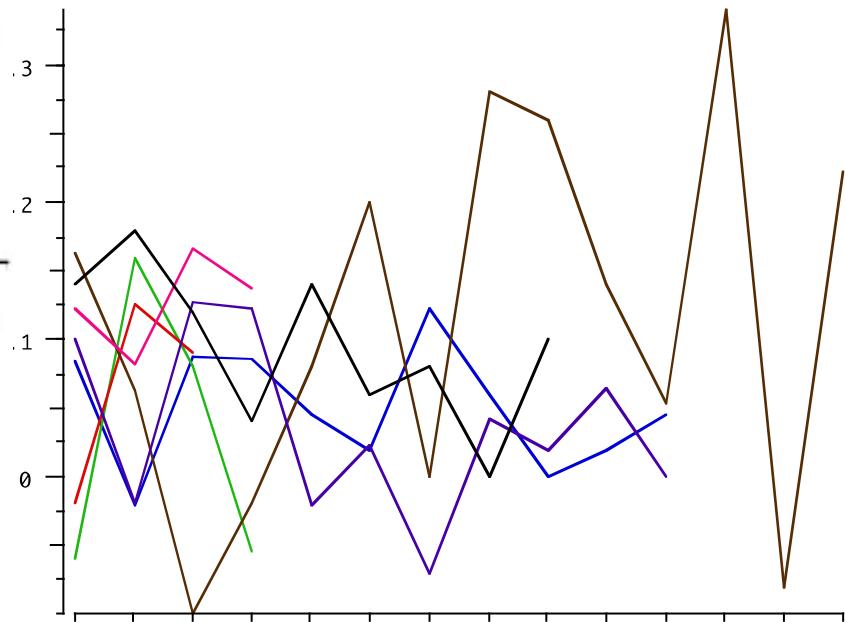
Legend: Blue bar, Red bar, Green bar, Brown bar



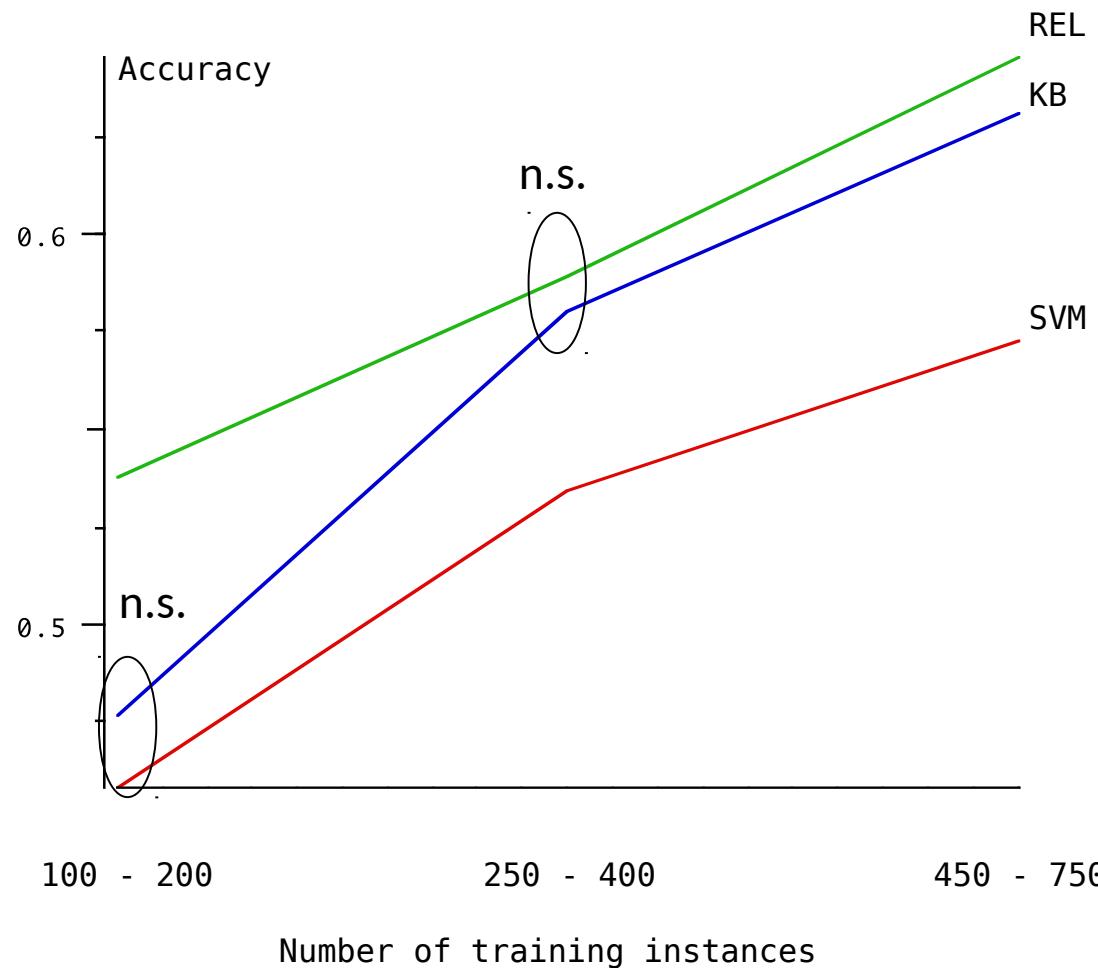
## Lesson 4: Of sample variance, effect size, and sample size, control the first before touching the last



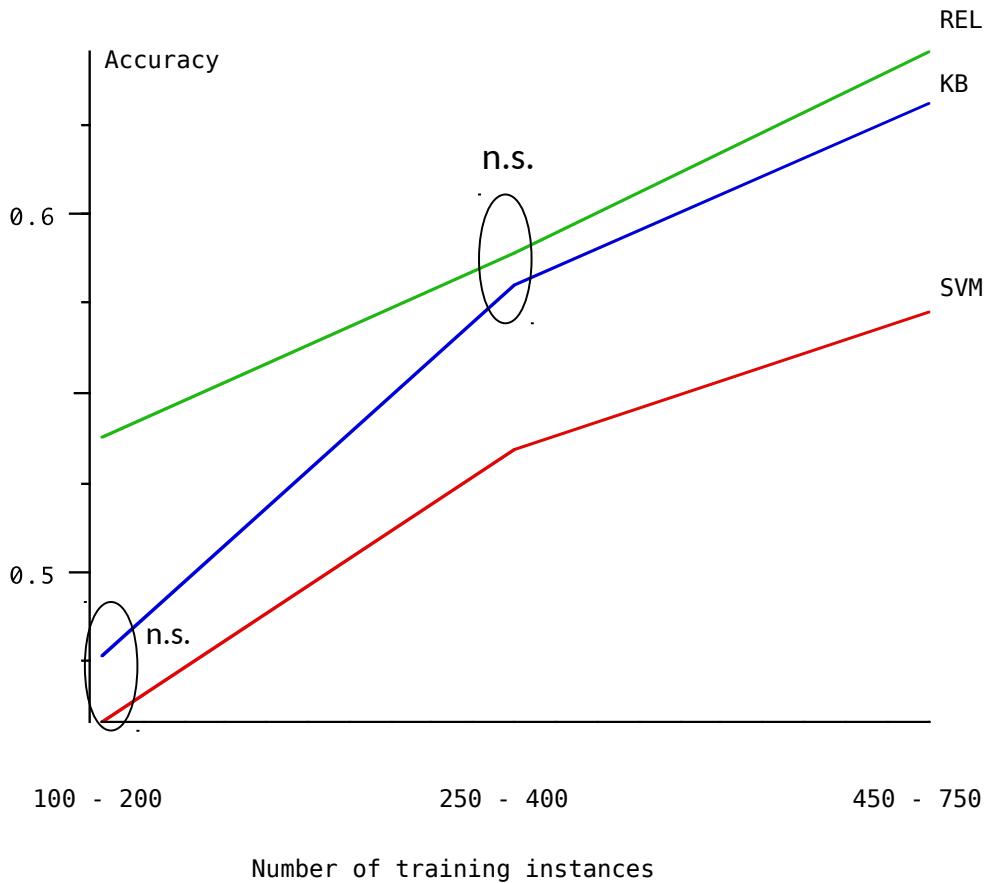
Subtract Alg1 from Alg2 for each subject, i.e., look at difference scores, correcting for variability of subjects  
"matched pair" test



# SIGNIFICANT DIFFERENCE HAVING CONTROLLED VARIANCE DUE TO SUBJECTS



## Lesson 5: Demonstrations are nice; explanations better



Having demonstrated that one algorithm is better than another we still can't explain:

- Why is it better? Is it something to do with the task or a general result?
- Why is it not better at all levels of training? Is it an artefact of the analysis or a repeatable phenomenon?
- Why does the REL curve look straight, unlike conventional learning curves?

These and other questions tell us we have demonstrated but not explained an effect; we don't know much about it.

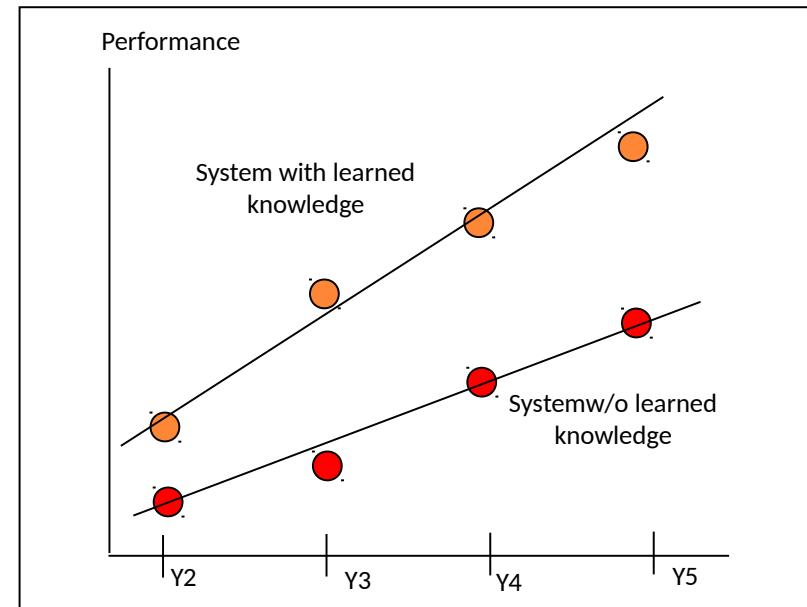
## Lesson 6: Most interesting science is about interaction effects, not main effects

System's performance improves at a greater rate when learned knowledge is included than when only engineered knowledge is included.

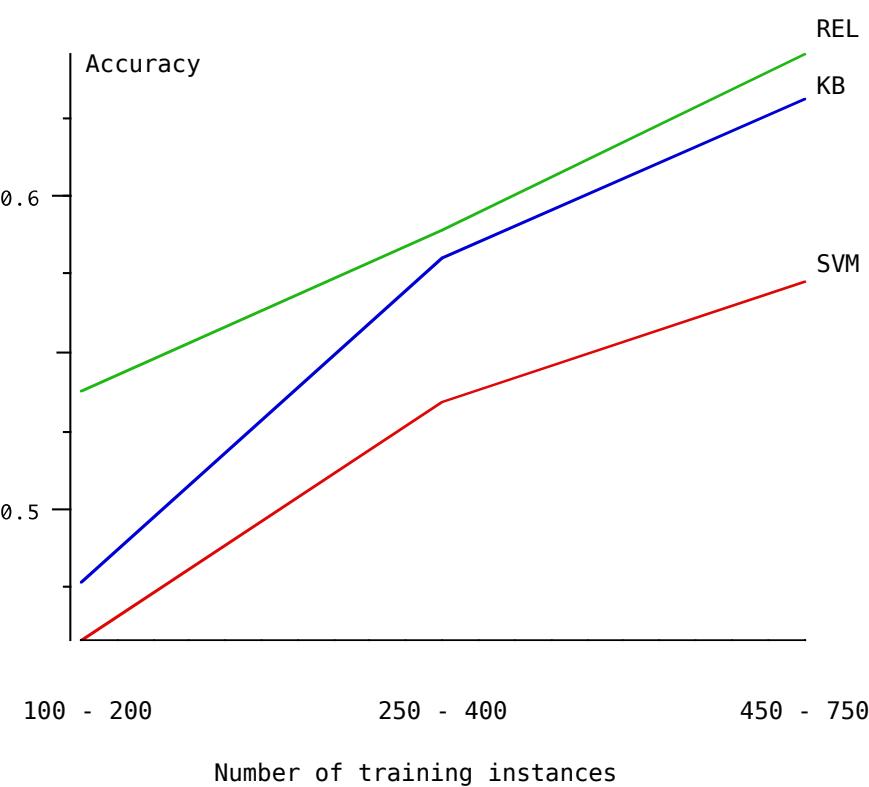
Learned knowledge begets learned knowledge

The lines aren't parallel: The effect of development effort (horizontal axis) is different for the learning system than for the nonlearning system.

Interaction effect!



# Lesson 8: Not all studies are experiments, not all analyses are hypothesis testing



The purpose of the study might have been to *model* the rate of learning

Modeling also involves statistics, but a different kind: Degree of fit, percentage of variance accounted for, linear and nonlinear models...

## Lesson 9: Significant and meaningful are not synonyms

$$\omega^2 = \frac{\sigma_1^2 - \sigma_{1|\text{Algorithm}}^2}{\sigma_1^2} \quad \text{Reduction in uncertainty due to knowing Algorithm}$$

$$\hat{\omega}^2 = \frac{t^2 - 1}{t^2 + N - 1} \quad \text{Estimate of reduction in variance}$$

| $\hat{\omega}^2$ |            |
|------------------|------------|
| .192             | KB vs SVM  |
| .336             | REL vs SVM |
| .347             | REL vs KB  |

For "fully trained" algorithms ( $\geq 500$  training instances)

# Review of lessons every DARPA program manager needs to know

1. Evaluation begins with claims; metrics without claims are meaningless
2. The task of empirical science is to explain/reduce variability
3. Humans are large sources of variability
4. Of sample variance, effect size, and sample size, control the first before touching the last
5. Demonstrations are good, explanations are better
6. Most explanations involve additional factors; most interesting science is about interaction effects, not main effects
7. Exploratory Data Analysis: use your eyes to look for explanations in data
8. Not all studies are experiments, not all analysis hypothesis testing;
9. Significant and meaningful are not synonyms

# **Experimental Design**

# Experiments

- Studies in which conditions are controlled so that one or more independent variable(s) can be manipulated to test a hypothesis about a dependent variable

# **Experimental Research**

- Claim: Manipulation of A treatment variable (x), followed by observation of response variable (y)
- Experiment must be designed to control for other variables to establish causal relationship
- Experiments often can only reject hypothesis.

# **Descriptive Research**

- Descriptive research studies a situation or some aspect of it that stem at a specific point in time.
- No real hint of causal insight can be obtained from descriptive data. Well written may convince people of incorrect causal links. You may

# What Constitutes Causality?

- A change in one variable will produce a change in another
- Concept of a precondition influencing a variable of interest
- Sequence in time
- No other possible explanation

Attitude —→ Behavior

# Direction of Causation Issue

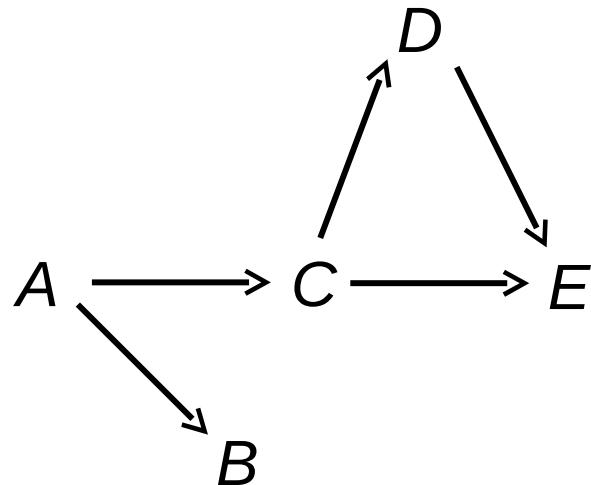
## Determining the direction of causation

- Draw on logic and previous theory
  - Whether one of the variables is relatively fixed and unalterable
- If a time lag exists between cause and effect then the causal variable should have a positive association with the effect variable lagged in time
- Temporal Correlation != Casual (but people often assume it is)

# What is a causal diagram?

## Components

- Variables
- Unidirectional arrows



# **Rules: displaying variables**

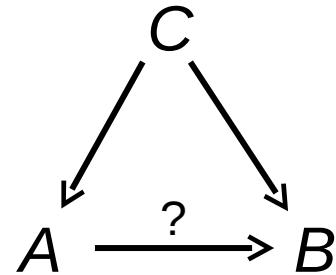
- Called “nodes” or “vertices”
- Should be clearly understood by others
- Displayed along the time axis (left to right or top to bottom)
  - but sometimes we ignore this rule if too complex
- **Variables, not values of variables**
  - “complexity” or “parallelism” is okay; “ $O(N)$ ” or 128 nodes is not

# Rules: drawing arrows

- An arrow
  - From a postulated cause to its postulated effect
- No bidirectional arrows
- An arrow with a question mark
  - The research question at hand
- An arrow without a question mark
  - Background theory or axiomatic

$$A \longrightarrow B$$

$$A \cancel{\longleftrightarrow} B$$



# Conditions for valid Causal Inference

Types of evidence relevant to evaluating causal relationships:

## *Condition of concomitant variation*

- Evidence that a strong association exists between an action and an observed outcome

## *Condition of time order of occurrence*

- Evidence that the action preceded the outcome

## *Absence of competing causal explanations*

- Evidence that there is no strong competing explanation for the relationship – that a high level of internal validity exists

# Issues In Experimental Research

- What type of experimental design should be used?
  - We start with designs involving humans somewhere, then pure computational experiments
- Should the experiment be performed in a “laboratory” setting or in the “field”?
- What are the internal and external threats to the validity of the experiment, and how can we control for the various threats to the experiment’s internal and external validity?

# Basic Symbols and Notations

- $O$  denotes a formal *observation* or measurement
- $X$  denotes *exposure* of test units participating in the study to the experimental manipulation of treatment
- $EG$  denotes an *experimental group* of test units that are exposed to the experimental treatment.
- $CG$  denotes a *control group* of test units participating in the experiment but not exposed to the experimental treatment
- $R$  denotes *random* assignment of test units and experimental treatments to groups. Increases reliability
- $M$  denotes that both the experimental group and the control group are *matched* on the basis of some relevant characteristics

# Types of Experimental Designs

## *Classical*

- Considers only one treatment level of an independent variable at a time

## *Statistical*

- Allows for examining the impact of different treatment levels of an independent variable and the impact of two or more independent variables

## Preexperimental Designs

- One-group, After-Only Design
- One group, Before-After Design
- Nonmatched Control Group Design
- Matched Control Group Design

## True Experimental Designs

- Two-group, Before-After Design
- Two group, After-Only Design
- Solomon Four Group Design

## Quasi-Experimental Designs

- Time Series Design
- Continuous Panel Design

## Classical Designs

## Experimental Designs

## Statistical Designs

## Completely Randomized Design

- Randomized-Block Design
- Latin Square Design
- Factorial Design

# **Classical Designs**

## **- Pre-experimental Designs**

### *One Group, After-only Design*

- Apply the experimental treatment to a subject or group and measure the results

EG X O

Avoid using this if at all possible. Why

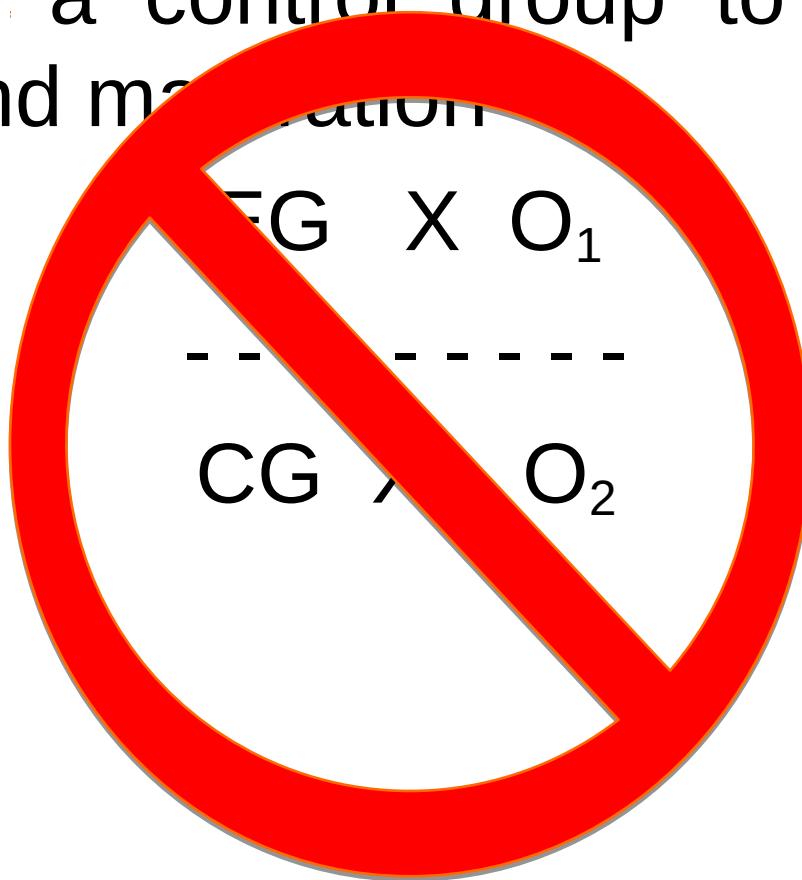
Leaves open the possibility that the results could be explained by events external to the design

# **Classical Designs**

## **- Pre-experimental Designs (Contd.)**

### *Nonmatched Control Group*

- Introduce a control group to control for history and maturation.



# **Classical Designs**

## **- Pre-experimental Designs (Contd.)**

### *Matched Control Group Design*

- Matches experimental and control groups to reduce selection bias

EG   M   X   O<sub>1</sub>

-----

CG   M   X   O<sub>2</sub>

# **Classical Designs**

## **- Pre-experimental Designs (Contd.)**

*One-group, Before - After Design  
(also called pre-post)*

- 改善控制 by adding before measure

EG    O<sub>1</sub>    X    O<sub>2</sub>

Before measure adds sensitivity by adding another method to control for confounding variables

Do we need control Group?

# **Classical Designs**

## **- Pre-experimental Designs (Contd.)**

### Threats to Experiment Validity

#### *Before Measure Effect*

- May change response because respondents know that they are being studied
- Results in more “socially desirable” behavior

#### *Mortality Effect*

- Some subjects may stop participating in the experiment

#### *Instrumentation Effect*

- Results from a change in the measuring instrument

# **Classical Designs**

## **- True-experimental Designs**

True experimental designs adopt random assignment procedure and use one or more control groups

### *Random Assignment*

- For any given assignment to a treatment, every member of the universe has an equal probability of being chosen for that assignment

# **Classical Designs - True-experimental Designs (Contd.)**

## *Two-group, Before-after Design*

- Adds a control group to one-group, before - after design
- Helps control for history and maturation
- Controls for reactive effect of  $O_1$  and  $O_2$

EG    R     $O_1$      $\times$      $O_2$

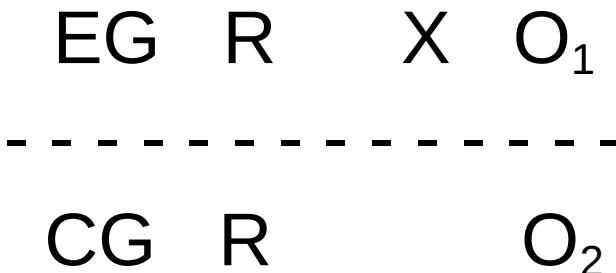
-----

CG    R     $O_3$          $O_4$

# **Classical Designs - True-experimental Designs (Contd.)**

## *Two Group, After-only Design*

- Randomization can *match* test and control groups on all dimensions simultaneously, given a sufficient sample size



- There is no interaction effect of testing as there are no pretest requirements

# **Pretest-Posttest (one group)**

- Quasi-experimental
- One set of measures taken before and after treatment or intervention
- Compare pretest and posttest scores
- Analysis
  - paired *t* test
- Weakness
  - No comparison or control group

# Pretest-Posttest (control group)

- Experimental design - random assignment
- Two groups
  - Control
  - Experimental
- Measures on dependent variable made on *both* groups pre- and posttest
- Significant differences in experimental group not found in control group attributable to treatment
- Analysis
  - difference scores compared with independent *t* test
  - ANCOVA pretest score as covariate

# **Classical Designs - Quasi-experimental Designs**

- Offer some degree of control but there is no random assignment of variables
- Provide more measurements and more information than pre-experimental design

## *Time Series Designs*

- Series of measurements are employed during which an experimental treatment occurs

EG O<sub>1</sub> O<sub>2</sub> O<sub>3</sub> O<sub>4</sub> X O<sub>5</sub> O<sub>6</sub> O<sub>7</sub> O<sub>8</sub>

# **Classical Designs - Quasi-experimental Designs (Contd.)**

## *Trend Studies*

- Measures over time come from succession of separate random samples from the same population

## *Continuous Panel Studies*

- Collect a series of measurements on the same sample of test units over an extended period of time

# Statistical Designs

## *Completely Randomized Design*

- Any number of treatments can be assigned to test units on a random basis

|        |   |       |       |
|--------|---|-------|-------|
| $EG_1$ | R | $X_1$ | $O_1$ |
| -----  |   |       |       |

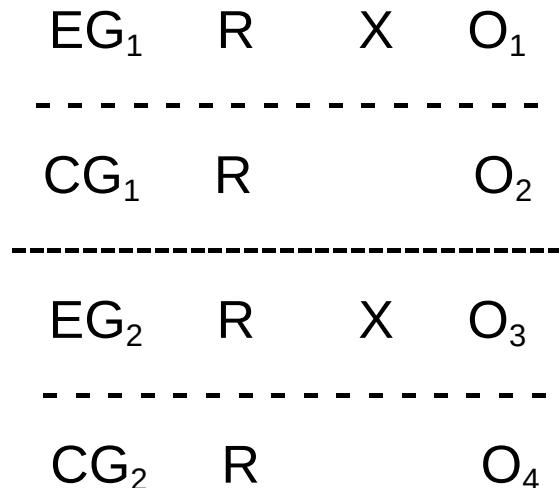
|        |   |       |       |
|--------|---|-------|-------|
| $EG_2$ | R | $X_2$ | $O_2$ |
| -----  |   |       |       |

|        |   |       |       |
|--------|---|-------|-------|
| $EG_3$ | R | $X_3$ | $O_3$ |
|--------|---|-------|-------|

# Statistical Designs (Contd.)

## *Randomized Block Design*

- Employs the randomization process for all variables
- Matching ensures that there are no differences between test samples on matched variables
- Matching and randomization are combined in randomized block design



# **Statistical Designs (Contd.)**

## *Latin Square Design*

- Reduces number of groups involved when interaction between the treatment levels and control variables are unimportant
- Requires same number of rows, columns, and treatment levels
- Cannot be used to determine interaction effects

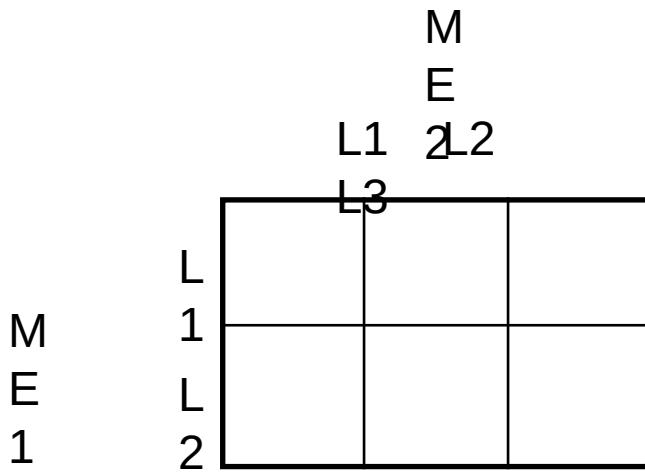
# Latin Square Design

- Minimizes order effects

|           | Test session |   |   |
|-----------|--------------|---|---|
|           | 1            | 2 | 3 |
| Subject 1 | A            | B | C |
| Subject 2 | B            | C | A |
| Subject 3 | C            | A | B |

# Two-Way Factorial Design

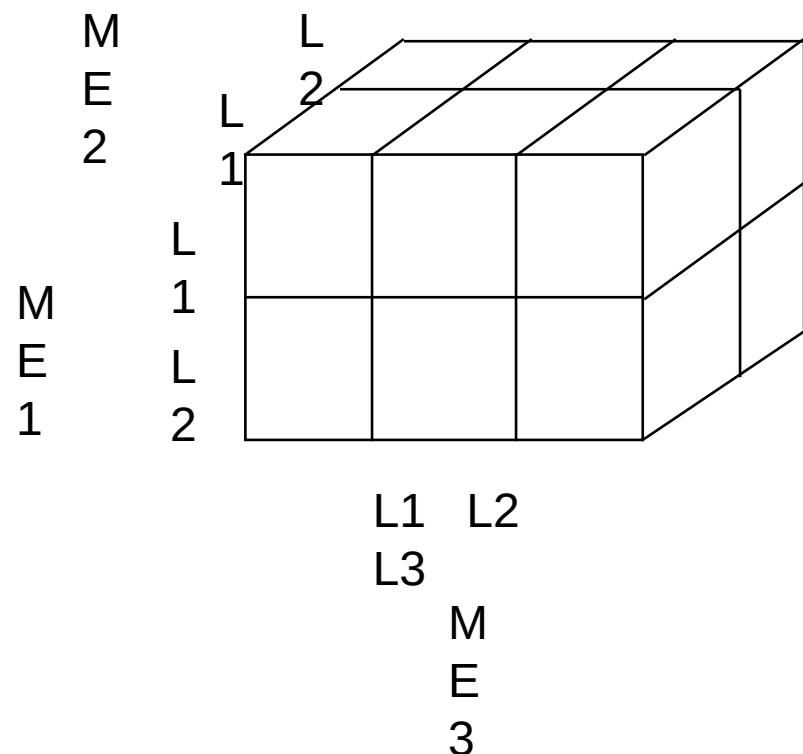
- Studies multiple independent variables
    - Main effects (ME)
    - Each with a number of levels (L)
  - Permits study of interactions
  - Analysis
    - ANOVA
- Example:  $2 \times 3$



# Three-Way Factorial Design

Example  $2 \times 2 \times 3$

- Studies multiple independent variables
  - Main effects (ME)
  - Multiple levels (L)
  - Interactions effects
- Analysis
  - ANOVA
  - Post hoc pairwise comparisons



# Statistical Designs (Contd.)

## Factorial Designs

- Two or more experimental variables are considered simultaneously
- Each combination of the experimental treatment levels applies to randomly selected groups

|                 |                |                |
|-----------------|----------------|----------------|
| EG <sub>1</sub> | X <sub>1</sub> | O <sub>1</sub> |
| EG <sub>2</sub> | X <sub>2</sub> | O <sub>2</sub> |
| .               | .              | .              |
| EG <sub>n</sub> | X <sub>n</sub> | O <sub>n</sub> |

- Provides the ability to determine interactive effects of pairs of experimental variables and the main effect

# **PURE COMPUTATIONAL DESIGNS FOR STATIC DATA WITH NO PEOPLE.**

- If one can run the algorithms on “static” data without human interaction, then can do perfectly matched pairs. But often hard to show anything causal in such testing. For causality testing, or if dynamic data, must use (fully) factorial designs on control variables.
- Must avoid tuning on test data and contamination of test data into training processes.
- Datasets always have bias. Unfair to select just datasets where you do well without analysis/discussion of why.

# **EXAMPLES FOR DISCUSSION**

- Comparing two algorithms for UW Communication?
- Evaluating a models for predicting battery charge?
- Evaluating Adversarial examples?
- Comparing two processes for improving Security?

# **Issues in Experimental Research**

- What type of experimental design should be used?
- Should the experiment be performed in a "laboratory" setting or in the "field"?
- What are the internal and external threats to the validity of the experiment?

# Laboratory Experiments

- Experiments in which the experimental treatment is introduced in an artificial or laboratory setting
- Laboratory experiments with people tend to be *artificial*
- *Testing effect* exists as respondents are aware of being in a test and may not respond naturally
- Results may not have external validity
- Least costly and allow experimenter greater control over the experiment
- Alternative explanations of results are reduced, increasing internal validity

# **Field Experiments**

- Research study in which one or more independent variables are manipulated by the experimenter under carefully controlled conditions as the situation will permit
- Experimental treatment or intervention introduced in a completely natural setting
- Response tends to be natural
- Tend to have much greater external validity
- Difficult to control
- Competing explanations for results exist

# Threats to Experimental Validity

## *Threats to Internal Validity*

- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Selection Bias
- Mortality
- Selection - Maturation Interaction

# **Threats to Experimental Validity (Contd.)**

## *Threats to External Validity*

- Reactive or interaction effect of testing
- Interaction effect of selection bias and experimental variable
- Reactive effects of experimental arrangements
- Multiple treatment interference

# Limitations of Experiments

- Time and Cost
- Security
- Implementation Problems
  - Difficult to gain cooperation within the organization
  - Contamination may occur in experiments involving human subject due to inability to confine the treatment to designated experimental area
  - Variability in behavior across test units can be so large that it is difficult to detect experimental effects

# Conclusion

- Experimental CS is a fundamental underpinning of the information age
- Synthetic: studies phenomena that are entirely the product of human creation
- But they may be models of and depend on real data thus its studying the world or people.
- Information artefacts are extremely complex and can only be understood via empirical observation, especially if people use them
- Complexity often precludes direct theoretical analysis
- Even with proofs, need to validate assumptions
- **Your Project Needs Experimental Analysis**

# Checklist for evaluation design

- What are the claims? What are you testing, and why?
- What is the experiment *protocol* or procedure? What are the factors (independent variables), what are the metrics (dependent variables)? What are the conditions, which is the control condition?
- Sketch a sample data table. Does the protocol provide the data you need to test your claim? Does it provide data you don't need? Are the data the right kind (e.g., real-valued quantities, frequencies, counts, ranks, etc.) for the analysis you have in mind?
- Sketch the data analysis and representative results. What will the data look like if they support / don't support your conjecture?

## **Checklist for evaluation design, cont.**

- Consider possible results and their interpretation.  
For each way the analysis might turn out, construct an interpretation. A good experiment design provides useful data in "all directions" – pro or con your claims
- Ask yourself again, what was the question? It's easy to get carried away designing an experiment and lose the big picture
- Is everyone satisfied? Are all the stakeholders in the evaluation going to get what they need?
- Run a pilot experiment to calibrate parameters

# ANOVA: ANALYSIS OF VALUE

## IS YOUR RESEARCH WORTH ANYTHING?

Developed in 1912 by geneticist R.A. Fisher, the Analysis of Value is a powerful statistical tool designed to test the significance of one's work.



am i  
wasting  
my time?

Significance is determined by comparing one's research with the **Dull Hypothesis**:

$$H_0 : \mu_1 = \mu_2 ?$$

where,

$H_0$  : the Dull Hypothesis

$\mu_1$  : significance of your research

$\mu_2$  : significance of a monkey typing randomly on a typewriter in a forest where no one hears it.

The test involves computation of the  $F'd$  ratio:

$$F'd = \frac{\text{sum(people who care about your research)}}{\text{world population}}$$

This ratio is compared to the F distribution with  $I-1$ ,  $N_r$  degrees of freedom to determine a  $p(\text{in your pants})$  value. A low  $p(\text{in your pants})$  value means you're on to something good (though statistically improbable).

### Type I/II Errors

The Analysis of Value must be used carefully to avoid the following two types of errors:

Type I: You incorrectly believe your research is not Dull.

Type II: No conclusions can be made. Good luck graduating.

Of course, this test assumes both Independence and Normality on your part, neither of which is likely true, which means *it's not your problem*.

# CS 6000 Research Methods

---

Terrance E. Boult

El Pomar Prof. of Innovation and Security

# How To Read a Research Paper

---

CS 6000 Computer Science Research methods  
Terrance E. Boult

The way I read a letter's this:  
'Tis first I lock the door,  
And push it with my fingers next,  
To transport it be sure.

And then I go the farthest off  
To counteract a knock  
Then draw my little letter forth  
And softly pick its lock.

....

Emily Dickinson

# Why Read?

- Learn to do research
- Learn to think critically about quality of research and research papers
  - ◆ Someone will be thinking critically about your own work!
  - ◆ In any discipline, there are fads and there are lasting ideas... learn to tell the difference!
- Gain perspective
- Key issue: what are the questions to ask/answer?

# Research Papers

- Primary form in which research results are disseminated in computer science
- Conference papers (shorter)
- Journal papers (longer)
  - ◆ Often the complete version of a conference paper
  - ◆ May come out several years after the conference paper, may combine multiple conf. papers
- **Peer review** is the cornerstone of the scientific publishing process -- quality and Impact matters. We'll talk about that again when we discuss writing this

# How to Read a CS Research Paper

- five-pass method:
  - ◆ Pass 1: Browse
  - ◆ Pass 2: Scan to get General ideas
  - ◆ Pass 3: Critical Read
  - ◆ Pass 4: Creative Read
  - ◆ Pass 5: In-depth understanding

Source: S. Keshav

# Searching is NOT enough.

- We've discussed searching for papers.
- But it is very important to also keep aware, especially when starting out, so need to just look though stuff.
- Need to be able to “consider” 100's of papers per year. Big conferences 200~700 papers and journals have 100-600 papers per year. With 2-3 big conferences per area per year, plus 4-5 tier-1 journals 4 or 5- tier 2 journals there is a lot to consider in your field.
- Keep a “reading list”.. add to it frequently

# The browse

- You need to go through 100's or 1000's of papers a year The "Browse" is to help be efficient.
- You "browse" through many many papers to see what is interesting/relevant. One approach is to first read abstract, scan figures/captions, This is a 1-2 min "gut check" based on content.
- Venue check? Author Check? Citation check
  - ◆ papers in places by people or that cite stuff you already know are relevant are more likely relevant.
- Err on being positive. Add to “to read” list if even slightly positive

# The Skim

- if Browse is positive, time to skim positively. Believe the author and assume its all true, decide if the paper is actually interesting and worth more exploration. You'll be surprised how many papers fail at this stage.
- If in 2-3 sentences you cannot say what is the problem addressed and what is new about the papers solution, and how significant was the experimentation/theory, you skimmed too lightly. If you have too many ideas to express it in only 2-3 sentences each you

# After the scan you should be able to answer the “six Cs”:

## 1. Category

- What type of paper?

## 2. Context

- What other papers is it related to? How related to your work?

## 3. Contributions

- Main contributions?

## 4. Credible

- Does it appear to have any credibility?

## 5. Care (i.e. the Who care’s test)

- Given its valid, will anyone care? will it change anything?

## 6. Cost:

- How much time and effort will it cost you to read it carefully

# Critical read: 3<sup>rd</sup> pass

- ~ 1 hour
  - Read carefully, but ignore details (proofs, for example).
1. Note assumptions and question them
  2. Analyze experiments/conclusions
  3. Analyze Figures, diagrams, illustrations, graphs.
    - ◆ Properly labeled? Error bars? Etc...
  4. Mark relevant unread references
- After you should be able to summarize main thrust & point out likely weakness in assumptions, methodology and/or

# Reading a Paper Critically

- Understand the problem
- Understand the proposed solution
- Understand competing approaches / designs
- Understand the Evaluation Methodology
- Understand the comparison to state of the art
- Evaluate the paper's claims. Do not assume the paper is correct, even if published in a prestigious peer-reviewed venu

# Critically Evaluating Sources

## ➤ When reading sources critically.

- ◆ Ask questions about the credentials and reputation of the author and the place of publication.
- ◆ What do you think is the writer's purpose and the audience whom the author is addressing?
- ◆ Ask questions about the ideas you read: an easy way to do highlight sections, write your annotations in the margins or a reading journal and/or if you get a sense of doubt, make a note of what troubles you.
- ◆ Be on the lookout for assumptions that may be faulty. If you are reading an article on computer-aided home-schooling and the writer favors home-schooling because it avoids subjecting students to violence in schools, the unstated (but untrue) assumption is that all schools are violent places.

- ◆ Make sure the writer's evidence is adequate and accurate. For example, if the writer is making a generalization about all Chinese students based on a study of only 10, you have cause to challenge the generalization as resting on inadequate evidence.
- ◆ Note how the writer uses language. Which terms does the writer use with positive—or negative---connotations, signaling the values the writer holds? Does the writer flamboyantly denigrate and dismiss the views of others with such phrases as “a ridiculous notion” or “laughably inept policies?”
- ◆ Be alert for sweeping generalizations, bias, and prejudice: “Women want to stay home and have children.” “Men love to spend Sundays watching sports.”

# Critical Questions

- Is the problem carefully stated/formulated?
- Is it a meaningful/real problem?
- Is it the RIGHT problem (formulation vs description)?
- Can you make a list of assumptions, explicit and more importantly implicit about the problem?
- What, if anything, is novel about the problem formulation?
- What well known problems are related or the same?
- Are there simple solutions the authors do

# Critical Questions

- Can you make a list of assumptions, explicit and more importantly implicit about the technique? Consider every divide, matrix op, function inversion, minimization (maximization) ? Are the assumptions justified?
- What are the limitations of the solution (including limitations the authors might not have noticed or clearly admitted)?
- Is the logic of the paper clear and justifiable, given the assumptions, or is there a flaw in the reasoning?

# Critical Questions

- Do all the pieces of their work fit together logically?
- Has the right theorem been proven?
- Are Theorems Logically supported? Are proofs given? Cited? If cited, how credible a source?
- If the authors present data, did they gather the right data to substantiate their argument, and did they appear to gather it in the correct manner?
- Did they have enough data to make a statistically sound decision?

# Critical Questions

- Did they interpret the data in a reasonable manner?
- Would other data be more compelling?
- Did they compare with actual state of the art performance?
- Was the same data used for training and testing?
- WereParms selected on Test Results?
- What were the results? Did they do what they set out to do?
- On what dimension(s) did they advance the art?

# Critical Questions

- Have the authors been cutting corners (intentionally or unintentionally)?
- Would results be reproducible?
- Problematic experimental setup?
- Confounding factors?
- Unrealistic or artificial benchmarks?
- Comparing apples and oranges?
- Methodological misunderstanding?
- Do the numbers add up?
- Are the generalizations valid?
- Are the claims modest enough?

# 4<sup>th</sup> pass: Creative Read

- Fourth read is to read creatively. Reading a paper critically is easy, in that it is generally easier to tear something down than to build it up. Reading creatively involves harder, more positive thinking and more "creative" thinking.
- What is the Novelty of the paper? Does it open up new directions?
- What are the good ideas in this paper? Do these ideas have other applications or extensions that the authors might not have thought of? Can they be generalized further?

# 4<sup>th</sup> pass: Creative Read

- Could their ideas be combined with other approaches? Combined with your work?
- If there were other assumptions made, could that improve the approach? Could you ensure those assumptions were true for some problem?
- Are there better numerical methods than what they are doing?
- Are there possible improvements that might make important practical differences?
- If you were going to start doing research from this paper, what would be the next thing you would do?

# Assignment #2 Due in 2 weeks

- Choose a “top” journal or conference in your field with  $\geq 45$  papers one 2017/2018 issues. (DO NOT search for papers with keywords.. but can search for journals/conf title)
- Browse at least 45 papers in that issue, in order. Time yourself on each, aim for no more than 2 min.
- Scan 8+ papers, write 2-3 sentence summary
- Critically & Creatively read best 2 or 3.

Upload your notes as a latex document

In you journal discuss your process and your learning about the process (not the papers).

# Where: Sources and Publication Venues

- Why do we cite papers?
  - ◆ Relate our problem to prior work
  - ◆ To establish what is the current state-of-the-art
  - ◆ To add credibility to our arguments/story
  - ◆ To show readers/reviewers we know what is important in prior work, what is solved and what is open.

# Why Publish?

- To go to exotic location and fancy hotels
- To impress your mother with your name in print
- To become rich and famous
  
- To graduate
- To get a job
  
- To satisfy research contract requirements
- To help others learn from your experience
- To grow the body of useful knowledge.

# Recognize scholarly articles

- Scholarly articles are not usually found in magazines in a dentist's office.
- Scholarly articles are peer reviewed—that is, other scholars read all the articles and approve them for publication.
- These articles have section headings, abstracts, and “summary” and/or “conclusion” headings. They determine the author’s main idea.
- They refer to works of other scholars (Reference Page, in-text citations, author credentials, notes, in depth analysis, uses academic or technical language for informed readers, appears in journals that don’t include colorful advertisements, etc.

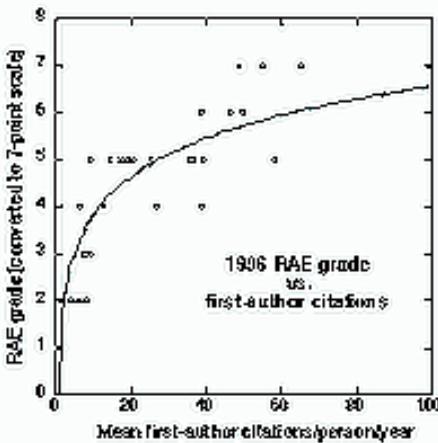
# Evaluating sources: Developing junk antennae

- If you find an article in a subscription database (IEEEExplore, *InfoTrac*, *LexisNexis*, etc.) you will know that the article has been published in print.
- If the article has been published in a reputable periodical or in an online journal sponsored by a professional organization or university, you can assume that it is a valid source for a research paper citation But beware junk publications.
- For works devised specifically for the Web, need strong checking o separate the information from the junk.

# Where to publishes and where to look for papers

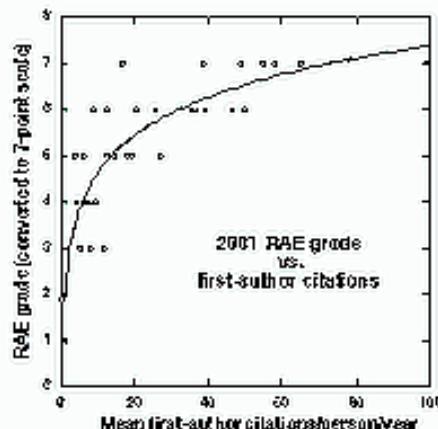
- IEEE Xplore
- Google Scholar (Demo/features later)
- Researchgate (Demo later)
- Other DBs in library
- Signup for mailing lists
  
- Before we get into how to read let look at understanding venue quality and metrics

# Research Assessment, Research Funding, and Citation Impact



**"Correlation between RAE ratings and mean departmental citations +0.91 (1996) +0.86 (2001) (Psychology)"**

**"RAE and citation counting measure broadly the same thing"**



**"Citation counting is both more cost-effective and more transparent"**

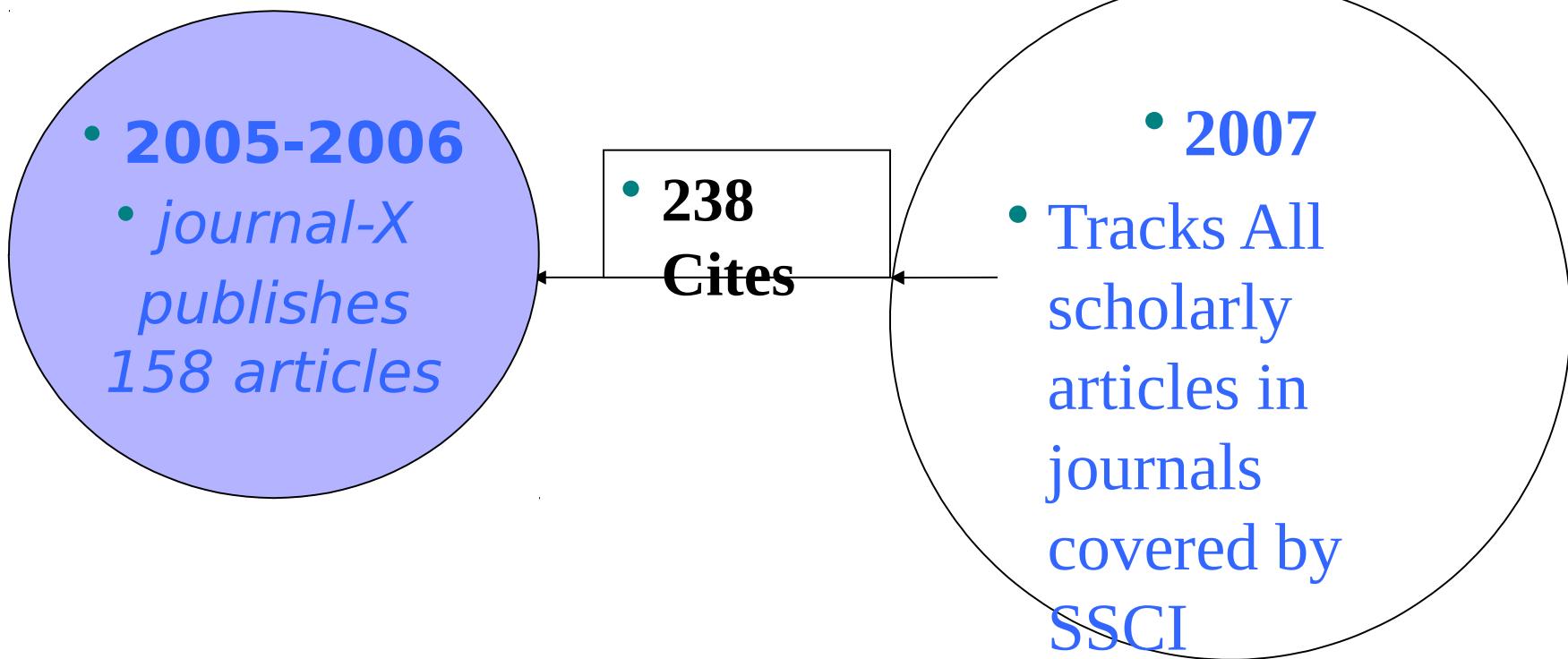
(Eysenck & Smith 2002)

<http://psyserver.pc.rhbnc.ac.uk/citations.pdf>

# Assessing Pub. Venue Quality

- Amount of Influence
- # readers
- Journal Impact Factors
- Leading Conferences (really only in CS, most other fields conferences don't matter)
  - ◆ Acceptance Rates (Selectivity)
  - ◆ Total Citations
  - ◆ H-index

- The Classic Journal Impact Factor



$$\bullet \text{2007 Impact Factor} = \frac{\text{238 2007 citations}}{\text{158 2005-06 articles}} = 1.506$$

# So what?

- JIF is a measure of relative currency – 2 year window.
- JIF is a GROSS average. Average article in Nano Letters cited 10.371 times,
- But the citation RANGE = 0 - 319 times (14 articles cited zero times!).
- Never ever intended to measure quality of an individual article or author, even Thomson Scientific says that.

# A Better Citation Metric

- h-Index (Hirsch Index)
- An h-Index of 45 means a person (or dept. or a journal) has 45 articles cited at least 45 times.
- Can be calculated from Web of Science  
<http://library.buffalo.edu/libraries/e-resources/webofscience.html>
- Estimated in Google Scholar and MS Academic Index (but both include self-citations in it and can be somewhat gamed.. so check for self-citations in people..)

# Critique of h-index

- Rewards longevity, but not least-publishable-unit or sheer quantity. (Though quantity does matter)
- Recent and old work rewarded equally
- ± Does not reward highly cited papers more
- Many variants (g-index, m-index, etc. proposed to weight age, recent work, & highly cited papers, # of coauthors)
- Relatively insensitive to manipulation. (but some demonstrated cases of manipulation)

# Variants of h-index

- **g-index** = g number of papers that received (collectively)  $g^2$  citations [Rewards highly cited papers]
- **m-index** = h-index / no. of years a researcher has published [normalizes for longevity]
- **hX-index** = h-index based on citations in the most recent X years (but includes citations to older papers). (e.g. h5-index in google scholar)
- **iY-index** number of papers with at least Y citations (ie.g i10-index in scholar)

# Pub Indexes – Many more players – 1

- Google Scholar/Harzing's POP
- SciFinder
- NASA Astrophysics Data System (ADS)
- Amazon (Search inside this book)
- Scitation/Spin Web/PROLA
- Citation Bridge (US Patents)
- USPTO
- Optics InfoBase

# Pub Indexes – Many more players - 2

- CiteSeer (primarily computer & info sci)
- ScienceDirect
- PsycInfo
- IEEE Xplore
- Spires (High Energy Physics)
- IOP Journals
- CrossRef

# My Take

- Top quality venues looked at more favorably for hiring (since time for pubs to get citations is short.)
- For an individual or department:
  - ◆ h-index plus
  - ◆ Total cites to all published articles plus
  - ◆ List of top citation papers
- This gives a pretty good take on the impact of one's articles within the limits of available citation data.
- Demonstrably superior to Journal Impact Factor in terms of perceived quality (by experts)

# CS 6000 Research Methods Publishing Venues, OA and Predatory Publishers

---

Terrance E. Boult  
El Pomar Prof. of Innovation and Security

As a scientists/professor you are  
a professional writer;

you get paid, in large part, to produce papers!

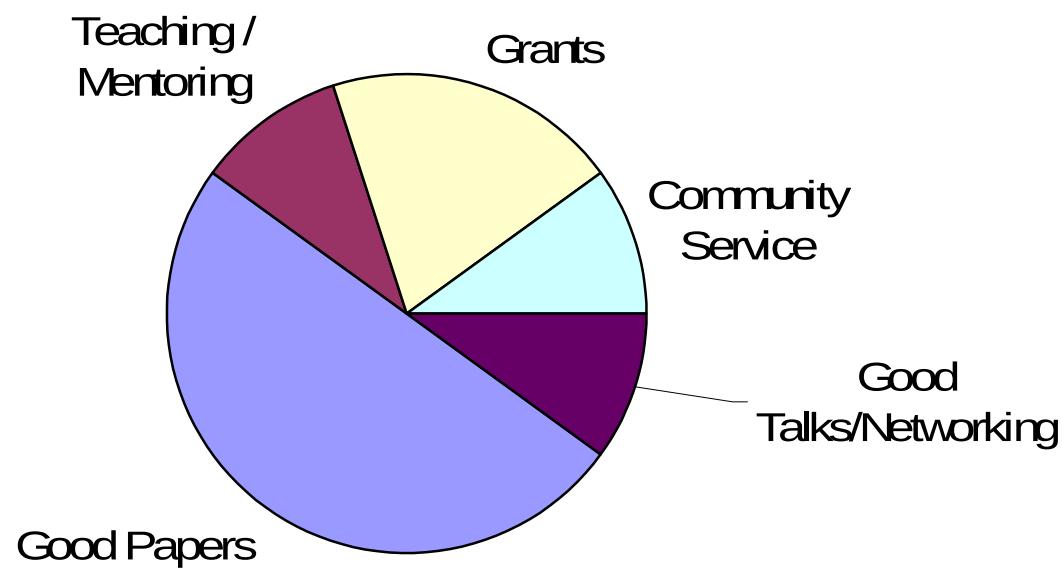
# What is a scientific paper

- A scientific paper is a written and published report describing original research results.
- 1. It must be the first publication of original research results,
- 2. In a form whereby peers of the author can repeat the experiments and test the conclusions, and
- 3. In a journal or other source document readily available within the scientific community

# Definition of Scientific paper

- An accepted original scientific publication containing scientific information to enable peers:
  1. To assess observations
  2. To repeat experiments
  3. To evaluate intellectual processes
  4. Must have an impact
  5. Available to scientific community without restriction
  6. Available for regular screening by one or more of the major recognized secondary services

# THE FACETS OF AN ACADEMIC CAREER (US) in Research University



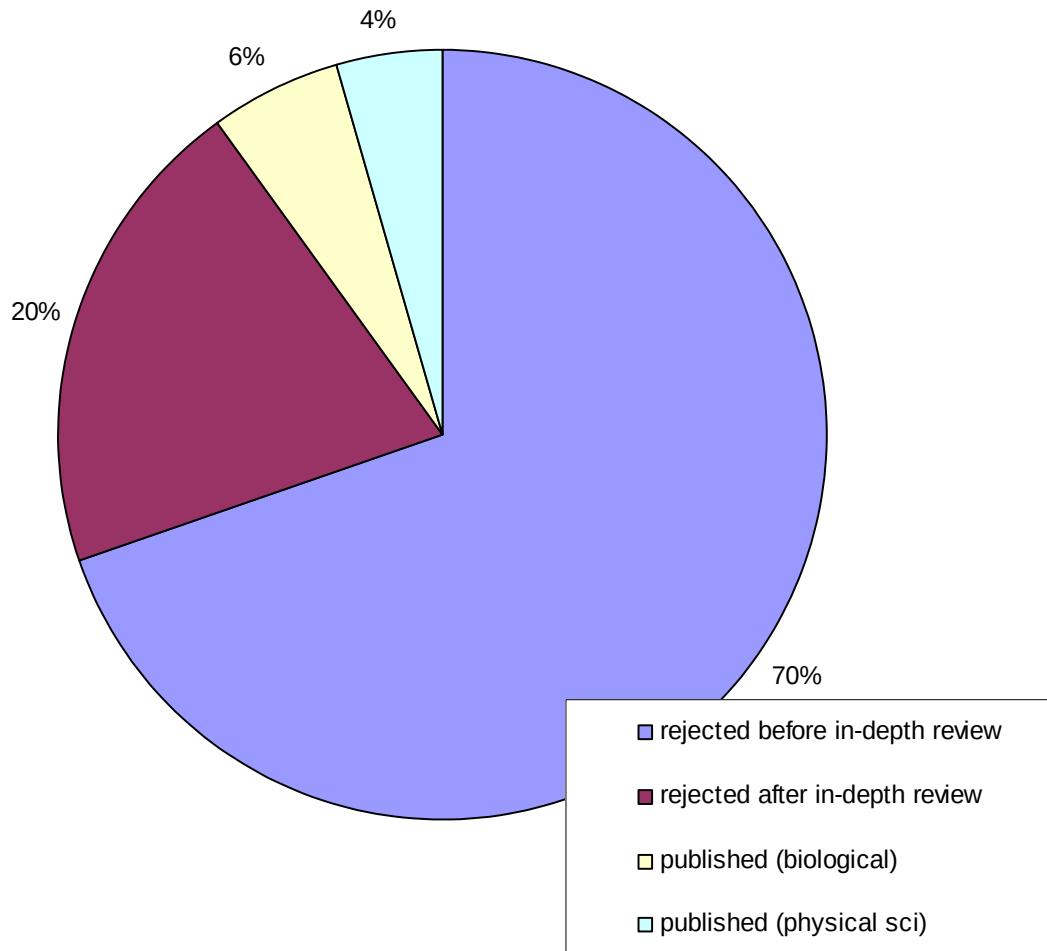
➤ Remember what you put in the literature is your scientific legacy after all else is gone

# ALWAYS LOOK FROM A DISTANCE

- See your paper as the reviewer will see it – believe me the reviewer will not see it as the best thing since sliced bread even if you do!

➤ Understand that the best scientists get rejected and/or have to make major revisions

# The rejection rate is tough.



# Choosing a Venue

- Fit to your topic
- Quality
  - ◆ Various “Tiers” of publications & Formal rating in some countries.
- Acceptance Rate
  - ◆ Generally in the publication (message from chairs, etc). Top venues are often summarized in websites.
- H5-Index in google or Journal Impact Factor
- Timelines
- Other dimensions of impact (e.g. OA, publisher etc).

# Finding Venues

- Google “top conferences in xyz” top journals in XYZ” (e.g.  
<http://www.guide2research.com/topconf/>)
- Wikicfp.com
- Make a calendar
- Search to make sure you are a good fit.
- Ask someone else to make sure you are a good fit.
- Read some papers in the venue
- Read what you can about their review process (more later on that).

# Finding Venues

- Most good venues do not allow double submission. Check before you submit!
- Plan maybe submit
  - tier1->tier2->tier3
  - ◆ As your plan.. Or maybe
    - tier1->tier1->tier[23].
- With most fields in CS, can do 2-3 conference rounds within a year, with mixed tiers.
- Understand your commitment (e.g. reviewing if you submit)
- Understand your costs and how you might reduce them, e.g. volunteer.

# Primer on Open Access (OA)

- OA simply means free-to-read.
- OA is fully compatible with rigorous peer review.
- OA does not necessarily mean author-pay  
(there are many models being tested).
- OA journals can be low or high quality, just like subscription journals.

# Can OA have Prestige?

- PLOS One (Maybe the largest science journal in the world) – est. 8,000 articles
  - ◆ JIF= 4.351 (10<sup>th</sup> out of 76 gen. biology journals)
- PLOS Biology
  - ◆ JIF=12.916 (8<sup>th</sup> out of 283 biochem journals)
  - ◆ Started in October 2003
- CVPR, most cited CS publication has been IEEE+“open access’ since 2011 and seen substantial increases in citations.

# OA – a flash in the pan?

- Directory of Open Access Journals (DOAJ)
  - ◆ [www.doaj.org](http://www.doaj.org)
  - ◆ More than 10500 fully OA, peer reviewed journals with over 1.9 Million articles.
  - ◆ new titles per day
- OA journal not the whole story
- Most non-OA journals allow authors to deposit their articles in an IR/DR (e.g. arxiv)

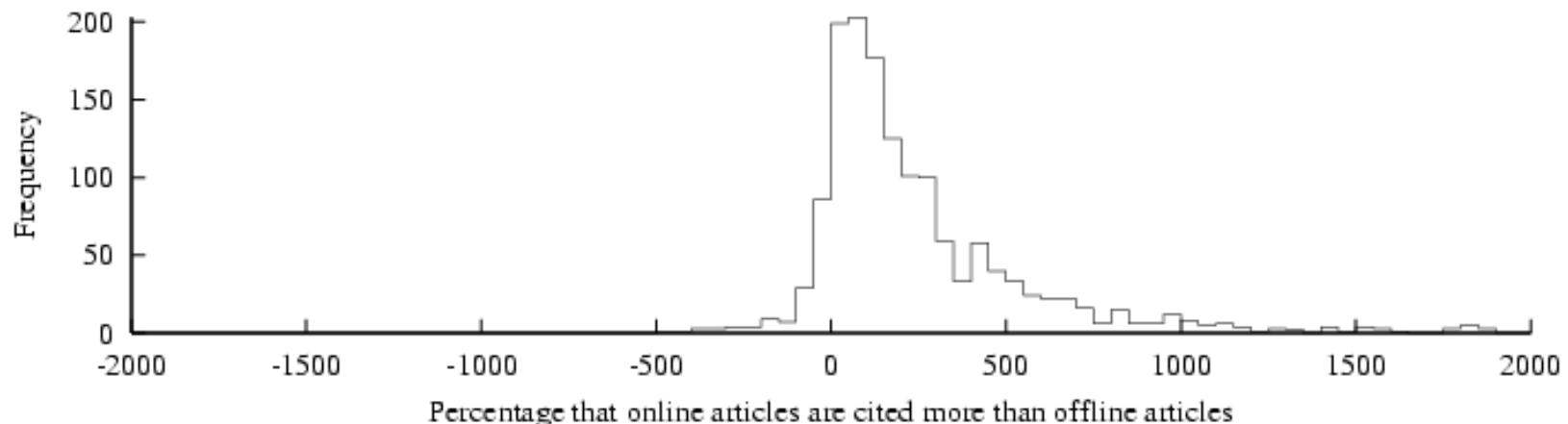
# SO WHAT?

- We publish for prestige, but we also publish to be read & cited.
- Some research that shows OA articles in good OA venues are cited 25-250% more than toll access (TA) articles?
- Lawrence S (2001) Free online availability substantially increases a paper's impact. Nature, 31 May 2001. <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- Open Access Citation Advantage: An Annotated Bibliography – A. Ben Wagner, Issues in Science & Technology Librarianship, Winter 2010.
  - ◆ <http://www.istl.org/10-winter/article2.html>

# The OA Advantage

- As scholar, enlarge your audience/impact.
- As reader, enjoy free online access to the literature.
- As teacher, your students have free, liability-free access (fair use, course pack).
- For all of us, moving away from an unsustainable journal publishing system.

# ***“Online or Invisible?” (Lawrence 2001)***



**“average of 336% more citations to online articles compared to offline articles published in the same venue”**

Lawrence, S. (2001) Free online availability substantially increases a paper's impact  
Nature 411 (6837): 521.

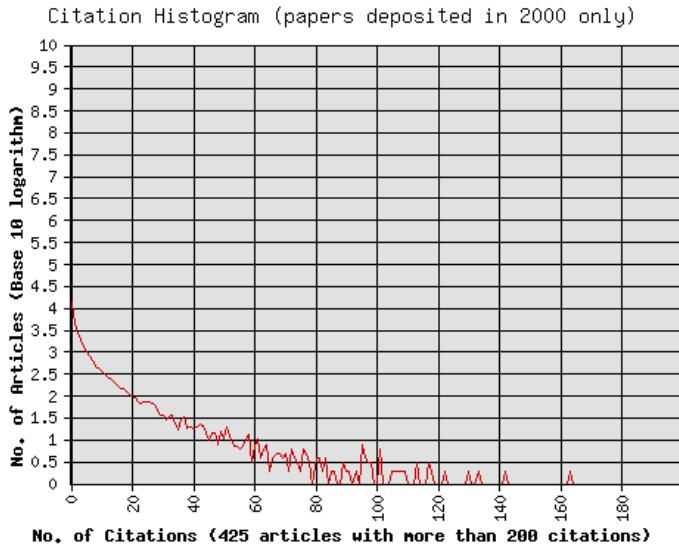
<http://www.neci.nec.com/~lawrence/papers/online-nature01/>

# Research Impact

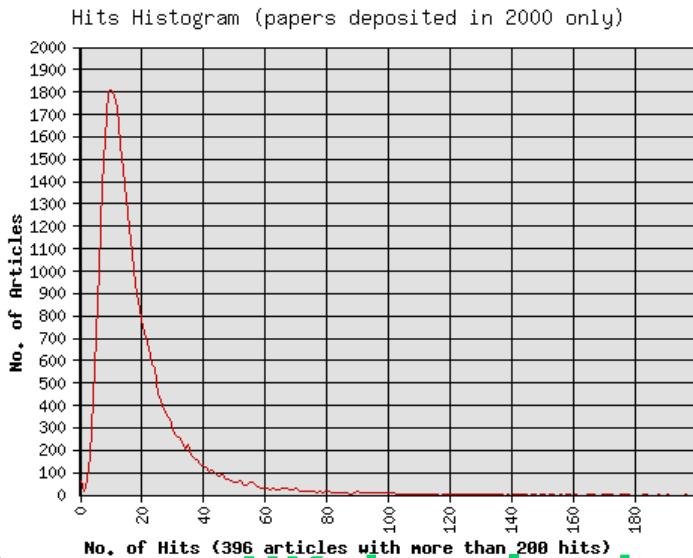
- measures the size of a research contribution to further research (“publish or perish”)
- generates further research funding
- contributes to the research productivity and financial support of the researcher’s institution
- advances the researcher’s career
- promotes research progress

**The author/institutional self-archived version  
is a supplement to -- not a substitute for --  
the publisher's official version**

1. Link the self-archived author/institution supplement to the publisher's official website
2. Pool and credit download counts for the self-archived supplement with downloads counts for the official published version
3. (All citation counts of course accrue to the official published version)



**Papers are not cited at all**



- **Average UK downloads per paper: 10**
- **(UK site only: 18 mirror sites in all)**

## Usage Impact (downloads)

is correlated with **Citation Impact**

(Physics ArXiv: **hep**, astro, cond, quantum; math, comp)  
<http://citebase.eprints.org/analysis/correlation.php>

**downloads from first 6 months after publication predict citations 2 years after publication**

(Quartiles Q1 (lo) - Q4 (hi))

**All**  $r=.27$ ,  $n=219328$

Q1 (lo)  $r=.26$ ,  $n=54832$

Q2  $r=.18$ ,  $n=54832$

Q3  $r=.28$ ,  $n=54832$

**Q4 (hi)**  $r=.34$ ,  $n=54832$

**hep**  $r=.33$ ,  $n=74020$

Q1 (lo)  $r=.23$ ,  $n=18505$

Q2  $r=.23$ ,  $n=18505$

Q3  $r=.30$ ,  $n=18505$

**Q4 (hi)**  $r=.50$ ,  $n=18505$

(correlation is highest for high-citation papers/authors)

# Publishing models for scholarly journals

- Traditional (subscription) model
- Gold open access = free to reader, author pays a fee
- Platinum open-access = free to author, free to reader
- Delayed open access = subscription model but OA after some time
- **Green Open Access: Open-Access Self-Archiving (Traditional + Arxiv)**

Slide adapted from Predatory Publishers are Poisoning Scholarly Communication by Jeffrey Beall UC Denver

# Some economic aspects of open access

- The change from subscription-financed journals to author-financed journals has many negative, unanticipated effects
- New business models focus on authors as revenue source
- Conflict of interest: more papers accepted = more income
- Not all subscription journals are good; not all OA journals are bad

# Predatory publishers

- Predatory publishers (journals) are those that exploit the gold open-access model for their own profit
- They take advantage of, exploit, and pander to scholarly authors
- They pretend to be legitimate, copying established and respected journals' websites and practices
- Many do a poor or fake peer review
- While Beall address journals, There are also predatory conferences..

*International Journal of Science  
and Advanced Technology  
(IJSAT)*



eISSN: 2221-8386  
Print ISSN:  
2304-652X

**CALL FOR PAPERS**

**October ISSUE**

Submissions due  
**October 27, 2014**

Notification of Acceptance  
**October 29, 2014**

Publishing Date  
**October 31, 2014**

The "International Journal of Science and Advanced Technology (IJSAT)" is an open access peer-reviewed international online journal. We publish high quality and refereed papers monthly. Papers reporting original research or extended versions of already published



Why IJSAT?  
IJSAT maintains blind peer review for all submitted manuscripts.

IJSAT indexed in DOAJ, Index Copernicus, New Jour, Cite Seer, Get Cited, Google Scholar etc.

OTHER NEWS

Volume 4 Issue 9, September 2014 is in Progress at Archive section

# Top Red Flags for Predatory Conferences

- Email to you inviting you to apply, especially with your name
- Not in google scholar under metrics
- Unrealistic review timelines
- Part of large multi-topic conference
- No listing of program committee
- Old topic with new conference.
- Insufficient Contact Information
- Unclear Author Fee structure
- False Index Claims
- No ISSN/ISBN

Table 1: Beall's Criteria for Identification of Predatory Journals and Publishers\*

Editor and Staff

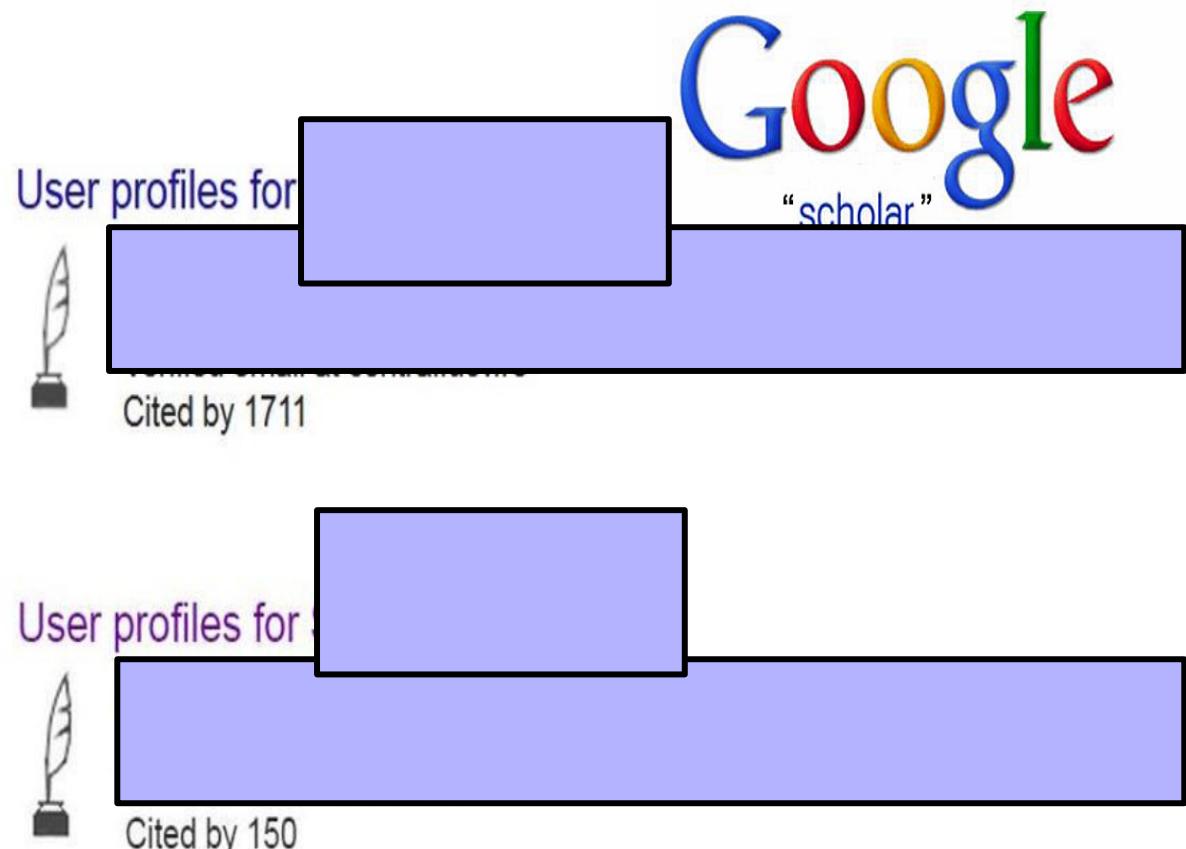
- The publisher's owner is identified as the editor of each and every journal published by the organization.
- No single individual is identified as any specific journal's editor.
- The journal does not identify a formal editorial / review board.
- No academic information is provided regarding the editor, editorial staff, and/or review board members.
- Evidence exists showing that the editor and/or review board members do not possess academic expertise to reasonably qualify them to be publication gatekeepers in the journal's field.
- Two or more journals have duplicate editorial boards (i.e., same editorial board for more than one journal).
- The journals have an insufficient number of board members (e.g., 2 or 3 members), have concocted editorial boards (made up names), name scholars on their editorial board without their knowledge or permission or have board members who are prominent researchers but exempt them from any contributions to the journal except the use of their names and/or photographs.
- There is little or no geographical diversity among the editorial board members, especially for journals that claim to be international in scope or coverage.
- The editorial board engages in gender bias (i.e., exclusion of any female members).

# Breakdown of research cultures

- Undoubtedly some have earned tenure and promotion through easy articles in predatory journals
- The role of merit in academic advancement is disappearing
- Never before has so much pseudo-science been published that looks like real science
- Some academic databases are filled with junk science
- Many researchers now expect cheap, easy, and fast publishing
- Open access advocates are in denial about the problems OA has caused
- It is possible for articles to be published open access yet still hidden

# Impact on Scholarly metrics

- Metrics can be somewhat gamed using predatory publishers
- Alternative metrics (ISI) gamed too
- Backlash against the impact factor



After a blog article by Beall brought up this case, google change scholar resulting in the drop.

And people are still playing their games  
But lawsuits & pressure on employers have reduced  
publicity



- <https://www.insidehighered.com/quicktakes/2015/01/21/scholar-accused-bogus-boost-google-rating>
- <http://www.biochemia-medica.com/2017/27/273>

# How predatory publishers damage science

- They've increased published research misconduct, such as plagiarism
- The pseudo-science they publish gets indexed in Google Scholar and other academic indexes
- They threaten demarcation, the division between science and pseudo-science, the cumulative nature of research
- They feed bogus research to societal institutions that depend on authentic science
- Entrepreneurs are using predatory publishers to make products appear efficacious

# Fake Impact Factors

- Companies make up and sell impact factors to open-access journal publishers
- Many publish articles mistakenly believing that their work is in an reasonable impact factor journal
- Predatory journals advertise their bogus impact factors in spam email and on their websites



COSMOS IMPACT FACTOR



International Institute For Research  
Impact Factor Journals (IFJ)



INTERNATIONAL JOURNAL  
IMPACT FACTOR (IJIF)

Scientific Impact Factor Information Services



INTERNATIONAL  
Scientific Indexing

# Resources on Predatory Pub.

- <https://libguides.gwumc.edu/PredatoryPublishing>
- <https://www.councilscienceeditors.org/resource-library/editorial-policies/cse-policies/approved-by-the-cse-board-of-directors/predatory-deceptive-publishers-recommendations-caution/>
- <https://libguides.usask.ca/predatorypublishers>

# Author services companies

- Provide services such as copyediting, pre-publication peer-review, statistical, methodological review, research promotion
- Most are not predatory but quality varies.
- Advantages those with money, who submit polished manuscripts
- The costs of scholarly publishing on the authors, favoring those with money (I say it has for a long time)





# **HOW TO GIVE A GREAT RESEARCH TALK**

Based in part on material from Simon.Jones, J.Huges, J. Launchbury,  
Stephanie Pfirman, Tom Carsey



**Vision And Security Technology**  
University of Colorado **Colorado Springs**

# GIVING A GOOD TALK

This presentation is about how to give a good research talk

- What your talk is for
- What to put in it (and what not to)
- How to present it

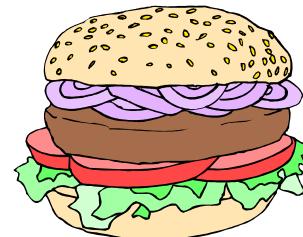


# WHY YOU SHOULD LISTEN TO THIS TALK

- Because many research talks are poor...
- ...and quite simple things can make *your* talks much better
- Because everyone benefits from good talks
  - Your audience is happier
  - You get promoted
- Because a research talk gives you access to the world's most priceless commodity: the time and attention of other people. Don't waste it!

# WHAT YOUR TALK IS FOR

Your paper = **The beef**



Your talk = **The sizzle,  
advertising the  
beef**



*Do not  
confuse the  
two*

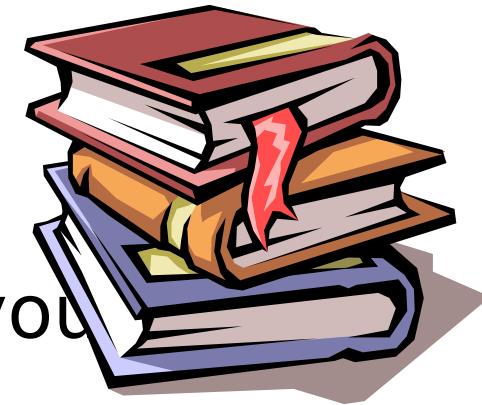
# WHICH "SCALE" TALK ARE YOU GIVING?

- 1-3Min
  - Advertising to get them to look at paper.
- 5Min
  - Why+Overview of what is most important
- 10Min
  - Why,What (what more) & How
- 20Min
  - Why,What (what more) & How &How it fails
- 30min
  - Why,What (what more) & How + Details
- Job talk
  - Who,Why,What & Why can do other things

# THE PURPOSE OF YOUR TALK...

The purpose of your talk **is not**:

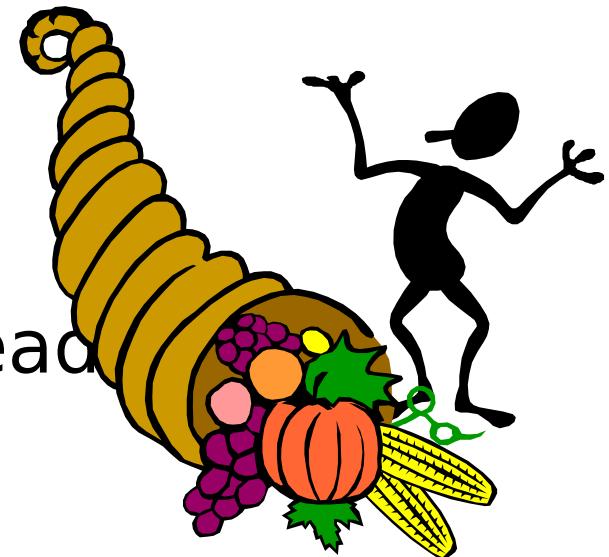
- To impress your audience with your brainpower
- To tell them all you know about your topic
- To present all the technical details



# THE PURPOSE OF YOUR TALK...

The purpose of your talk **is**:

- To give your audience an intuitive feel for your idea
- To make them foam at the mouth with eagerness to read your paper
- To engage, excite, provoke them
- To make them glad they came



# YOUR AUDIENCE...

The **audience** you might think you would like

- Have read all your earlier papers
- Thoroughly understand all the relevant theory of cartesian closed endomorphic bifunctors
- Are all agog to hear about the latest developments in your work
- Are fresh, alert, and ready for action

# BUT THAT IS PROBABLY A VERY SMALL AUDIENCE



And they will read their paper anyhow

# YOUR ACTUAL AUDIENCE...

The audience you get

- Have never heard of you
- Have heard of your subarea,  
but probably wish they hadn't
- Have just had lunch and are ready for a doze

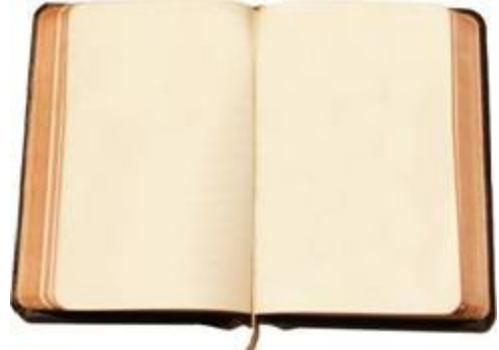


Your mission is to

## WAKE THEM UP

And make them glad you did

# TELL A STORY



- Prepare your material so that it tells a story logically
  - Use examples, anecdotes, and significant details
  - Create continuity so that your slides flow smoothly
    - Guide the audience through your story
    - Your last point on one slide can anticipate the next slide
- 
- [http://www.cgd.ucar.edu/cms/agu/scientific\\_talk.html](http://www.cgd.ucar.edu/cms/agu/scientific_talk.html)

# WHAT TO PUT IN



# WHAT TO PUT IN

1. Motivation (20%)
2. Part of your key idea (70%)
3. Impact 10%

# STRUCTURE

## Basic rule

- Say what you are going to say
  - 1-3 main points in the Motivation
- Say it
  - Give the talk's key ideas
- Then say what is the impact of what you said
- Don't try to build suspense and then unveil a surprise ending



▪ [http://www.safetyoffice.uwaterloo.ca/  
hspm/tools/images/scaffold\\_stair.png](http://www.safetyoffice.uwaterloo.ca/hspm/tools/images/scaffold_stair.png)

# MOTIVATION

**You have 15 seconds for  
they judge and 2  
minutes**

*to engage your audience  
before they start to doze*

They are asking

- Why should I tune into this talk?
- What is the problem?
- Why is it an interesting problem?
- Does this talk describe a worthwhile advance?

# The Motivation should end with a 1-sentence story.

## THE **BARE BONES** VERSION OF THE **EIGHT ESSENTIAL ELEMENTS**

\_\_\_\_\_ (CHARACTER)

**NEEDED** \_\_\_\_\_

**BECAUSE** \_\_\_\_\_

**BUT** \_\_\_\_\_

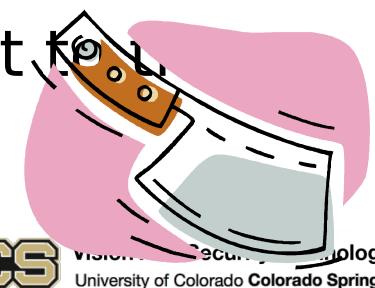
**SO,** \_\_\_\_\_

**FINALLY** \_\_\_\_\_

# YOUR KEY IDEA

If the audience remembers only one thing from your talk, what should it

- You must identify a key idea. “What I did this summer” is Not Good.
- Be specific. Don’t leave your audience to figure it out what is important for themselves.
- Be absolutely specific. Say “If you remember nothing else, remember this.” Or “The key idea is..”
- Organize your talk around this specific goal. Ruthlessly prune material that is irrelevant to the goal.



# NARROW, DEEP BEATS WIDE, SHALLOW



Avoid shallow overviews at all costs

Cut quickly to the chase: the technical “meat”

Best is quick context, the meat, and broad impact at end

# Examples & stories are your main weapons

- To motivate the work
- To convey the basic intuition
- To illustrate The Idea in action
- To show extreme cases
- To highlight shortcomings

When time is short, omit the details,  
not the example

# PAPER RUBERIC

- 90 total
- 15 Story
- 30 Comparative Analysysis
- 20 Number and appropriate References
- 25 Writing, style grammar.

# WHAT TO LEAVE OUT



# SKIP THE “OUTLINE OF MY TALK” (UNLESS >30MIN)

- Background
- The FLUGOL system
- Shortcomings of FLUGOL
- Overview of synthetic epimorphisms
- $\pi$ -reducible decidability of the pseudo-curried fragment under the Snezkovwski invariant in FLUGOL
- Benchmark results
- Related work
- Conclusions and further work



# NO OUTLINE!

“Outline of my talk”: conveys near zero information at the start of your talk

Worse, since your audience only gives you 2 minutes before dozing, you've just lost them

- But in long talks can put up an outline for orientation after your motivation
- ...and if >30min can have signposts at pause points during the talk

## PAST WORK

- [Pav83] The seminal paper
- [PPZ88] First use of epimorphisms
- [N93] Application of epimorphisms  
wiblification
- [XX98] Lacks full abstraction
- [XFB99] Only runs on Sparc, no  
integration with GUI

# **DO NOT PRESENT RELATED WORK**

But

- You absolutely must know the related work; respond readily to questions
- Acknowledge co-authors (title slide), and pre-cursors (as you go along)
- Don't forget acknowledgements (cite on slide), always give proper credit
  - Tip: Everyone in the audience has come to listen to your lecture with the secret hope of hearing their work mentioned
- Praise the opposition

“X’s very interesting work does Y; I have extended it to do Z”

# TECHNICAL DETAIL

$$\begin{array}{c}
 \frac{}{\Gamma \vdash k : \tau_k} \\
 \frac{\Gamma \cup \{x : \tau\} \vdash e : \tau'}{\Gamma \vdash \lambda x.e : \tau \rightarrow \tau'} \\
 \frac{\Gamma \vdash e_1 : \text{ST } \tau^\circ \tau \quad \Gamma \vdash e_2 : \tau \rightarrow \text{ST } \tau^\circ \tau'}{\Gamma \vdash e_1 >>= e_2 : \text{ST } \tau^\circ \tau'}
 \end{array}$$

$$\begin{array}{c}
 \frac{\Gamma \vdash e : \tau}{\Gamma \vdash \text{returnST } e : \text{ST } \tau^\circ \tau} \\
 \frac{\Gamma \vdash e : \tau}{\Gamma \vdash \text{newVar } e : \text{ST } \tau^\circ (\text{MutVar } \tau^\circ \tau)} \\
 \frac{\Gamma \vdash e : \text{MutVar } \tau^\circ \tau}{\Gamma \vdash \text{readVar } e : \text{ST } \tau^\circ \tau}
 \end{array}$$

$$\begin{array}{c}
 \frac{\Gamma \vdash e_1 : \text{MutVar } \tau^\circ \tau \quad \Gamma \vdash e_2 : \tau}{\Gamma \vdash \text{writeVar } e_1 e_2 : \text{ST } \tau^\circ \text{Unit}} \\
 \frac{}{\Gamma \cup \{x : \forall \alpha_i.\tau\} \vdash x : \tau[\tau_i/\alpha_i]}
 \end{array}$$

$$\frac{\Gamma \vdash e : \tau' \rightarrow \tau \quad \Gamma \vdash e' : \tau'}{\Gamma \vdash e e' : \tau}$$

$$\frac{\Gamma \vdash e : \text{ST } \alpha^\circ \tau \quad \alpha^\circ \notin FV(\Gamma, \tau)}{\Gamma \vdash \text{runST } e : \tau}$$

$$\frac{\forall j. \Gamma \cup \{x_i : \tau_i\}_i \vdash e_j : \tau_j \quad \Gamma \cup \{x_i : \forall \alpha_{j_i}.\tau_i\}_i \vdash e' : \tau'}{\Gamma \vdash \text{let } \{x_i = e_i\}_i \text{ in } e' : \tau'} \quad \alpha_{j_i} \in FV(\tau_i) - FV(\Gamma)$$

**Figure 1.** Typing Rules

# OMIT TECHNICAL DETAILS

- Even though every line is **drenched** in your **blood** and **sweat**, dense clouds of notation will send your audience to sleep
- Present specific aspects only; refer to the paper for the details (and poster if you have one)
- By all means have backup slides to use in response to questions



# PRESENTING YOUR TALK



# HOW TO PRESENT YOUR TALK

Your most potent weapon, by far, is  
YOUR

enthusias



# ENTHUSIASM

- If you do not seem excited by your idea, why should the audience be?
- It wakes 'em up
- Enthusiasm makes people dramatically more receptive
- It gets you loosened up, breathing, moving around

# **WRITE YOUR TALK BEFORE YOU GIVE IT**

Identical things will occur to you during the conference, as you witness on your talk and other people's talks. Make notes, fresh and update them before you give your talk.

But your talk absolutely must be fresh in your mind. So do review it the day before

# TECHNOLOGY

- Borrow a laser pointer, but avoid using it. Better to use mouse on the laptop.
- Even better don't need to point!
- Consider borrowing a wireless slide changer
- Test that your laptop works with the projector, in advance and just before presentation session.
- Laptops break: leave a backup copy on the web; bring a backup copy on a disk or USB key

# DO NOT APOLOGIZE

- “I didn’t have time to prepare this talk properly”
- “My computer broke down, so I don’t have the results I expected”
- “I don’t have time to tell you about this”
- “I don’t feel qualified to address this audience”

# THE JELLY EFFECT

If you are anything like I was when I starting, you may experience apparently-severe pre-talk symptoms

- Inability to breathe
- Inability to stand up (legs give way)
- Inability to operate brain

# WHAT TO DO ABOUT IT

- Deep breathing during previous talk
- *Script your first few sentences precisely*  
(=> no brain required)
- Use time talks/videos.
- Practice, Practice (maybe with flying objects;-)
- Move around a lot, use large gestures, wave your arms, stand on chairs
- Go to the bathroom before the session

You are not a wimp.  
Most People feels this way.

# Be prepared!



# BEING SEEN, BEING HEARD

- DO NOT turn to the screen. Point using mouse and keep facing the audience.
- Speak to someone at the back of the room, even if you have a microphone on
- Make eye contact; identify a **nodder**, and speak to him or her (better still, more than one).
- Don't focus on a friend.
- Watch audience for body-language and questions...

# QUESTIONS

- Questions are not a problem

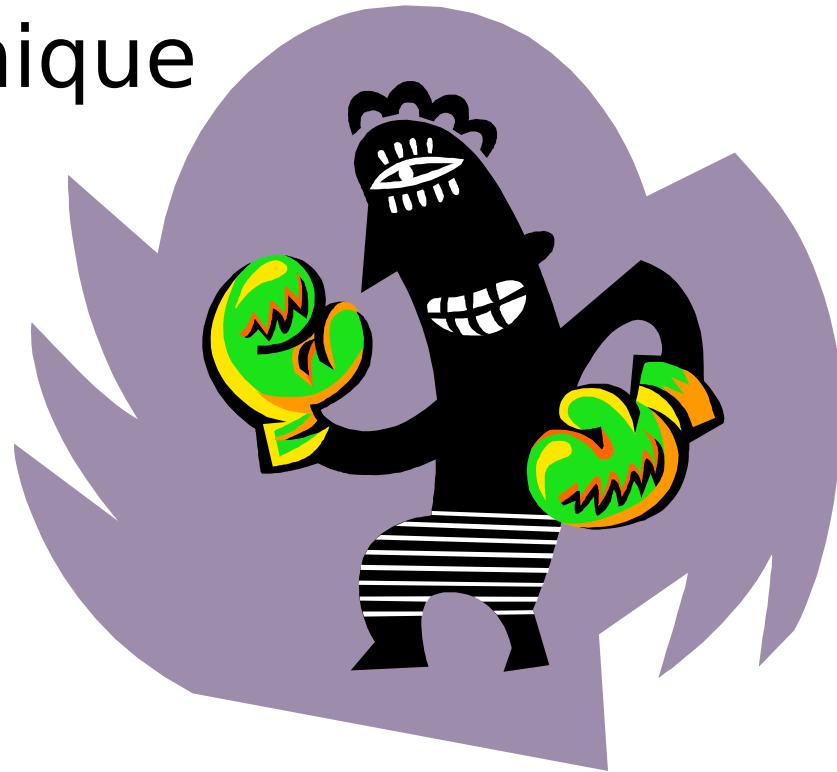
Questions are a **golden  
golden golden** opportunity  
to connect with your audience

- Specifically encourage questions during your talk: pause briefly now and then, ask for questions if appropriate. If there is after session send questions there.
- Be prepared to truncate your talk if you run out of time. Better to connect, and not to present all your material

# PRESENTING YOUR SLIDES

A very annoying technique

- is to reveal
- your points
- one
- by one
- by one, unless...
- there is a punch line



# PRESENTING YOUR SLIDES

Use animation effects

very

very

very

very

very

very

very

sparingly

# FINISHING

Absolutely without fail,  
finish on time

- Audiences get restive and essentially **stop listening** when your time is up. Continuing is very counter productive
- Simply truncate and conclude
- Do **not** say “would you like me to go on?”  
(it’s hard to say “no thanks”)

# CONCLUSION: THERE IS HOPE

The general standard is often low.

You don't have to be outstanding to stand out

You will attend 50x as many talks as you give.

Watch other people's talks intelligently, and pick up ideas for what to do and what to avoid.

# DO IT! DO IT! DO IT!

Good papers and talks are a fundamental part of research excellence

- Invest time
- Learn skills
- Practice

Write a paper, and give a talk, about  
**any idea,**  
no matter how weedy and  
insignificant it may seem to you

# RESEARCH IS COMMUNICATION

The greatest ideas are worthless if you  
keep them to yourself

Your papers and talks

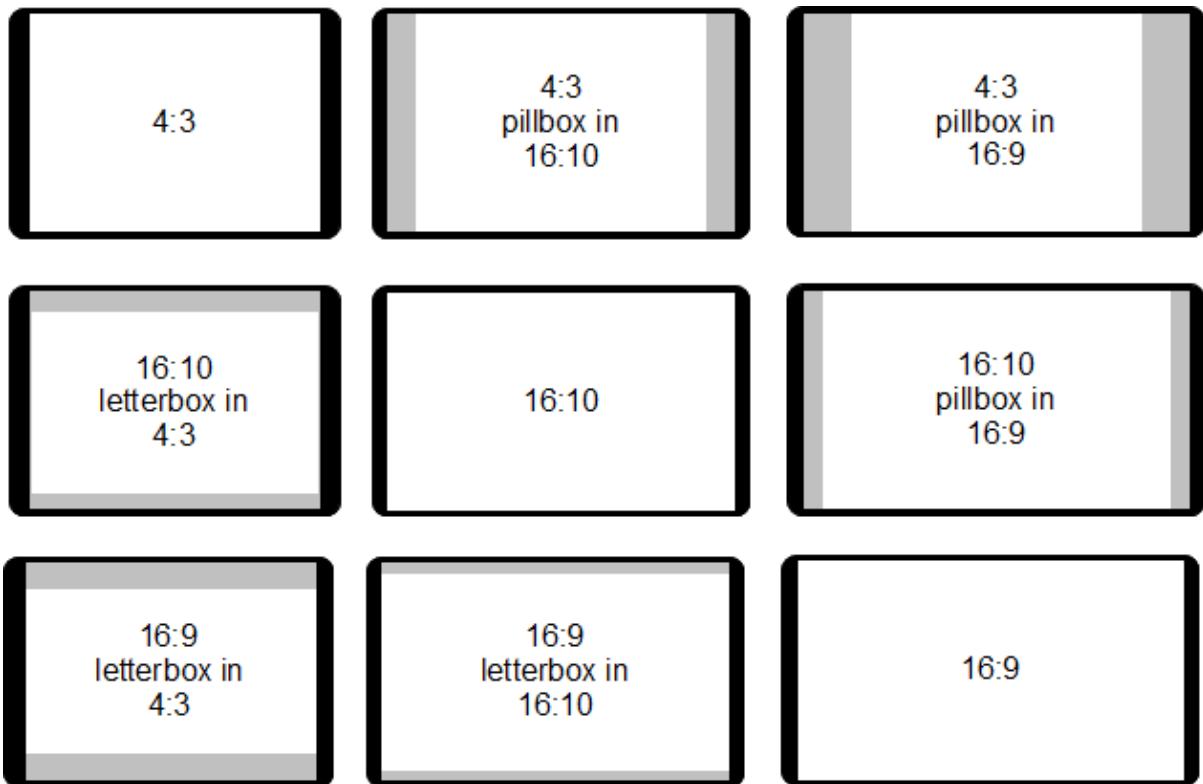
- Crystalise your ideas
- Communicate them to others
- Get feedback
- Build relationships
- (And garner research brownie points)

# TALK PREP TIPS

- 1) Know your presentation tool
- 2) Choose aspect ratio
- 3) Choose right font
- 4) Choose colors
- 5) Timing

# WHAT ASPECT? PLATFORM? CONNECTOR

- Ask about ratio



- Changing aspect in PPT scales things, often oddly.
- Bring connectors/converters for VGA/HDMI
- Test well before your talk.

# WHAT FONT TO USE

- Type size should be 18 points or larger:
  - 18 point
  - 20 point
  - 24 point
  - 28 point
  - 36 point
- AVOID USING ALL CAPITAL LETTERS  
BECAUSE IT'S MUCH HARDER TO  
READ
- \* References can be in 12-14 point font

## COLOR

- Dark letters against a light background work
- Dark letters against a light background are best for smaller rooms, especially when the lights are on for teaching. Best for any “dim” projector

- <http://www.fw.msu.edu/orgs/gso/documents/GSOWorkshopDocsSp2006/PresentationTipsinPowerPointBasics1-Whatfonttouse.pdf>



# Color

---

Light letters against a dark background also work

Many graphics experts feel that a dark blue or black background works best for talks in a large room, best “perceived” color & image differences. Greater chance of “sleeping”

Colors on your screen are NOT how they project Test color pallets on a projector. Check again on the actual projector

- [http://www.fw.msu.edu/orgs/gso/documents/GSOWorkshopDocsSp2006/PresentationTipsinPowerPoint.ppt#302,5,Powerpoint\\_basics:1---What font to use](http://www.fw.msu.edu/orgs/gso/documents/GSOWorkshopDocsSp2006/PresentationTipsinPowerPoint.ppt#302,5,Powerpoint_basics:1---What font to use)

# TALK TIMING

- The shorter the talk the more important it is you have the timing down very tight.
- I recommend auto-timed slides or a video for talks < 10min, or any length if you are nervous but will NOT be interrupted with questions.
- Feels awkward at first, as they will “go on without you”. You may feel everyone is staring at you, but its less noticeable than if you have very long pauses and look lost.
- Good to practice with distractions. Having things tossed at you while practicing is good.
- If you record a video with narration you can “practice” via listening to help know the material cold. But beware of losing energy in final talk!

## COPYRIGHT & TALKS

- If for an academic venue, “Fair use” applies. Good to have reference on source (at least in notes). Ideal choose images with “reuse rights” in google search
- Images & relevant Clip art can greatly improve talk. Weakly related images/art can distract but can still work.
- A talk where you can connect your story and visualize it as all the pictures on every slide is way easier to remember, for you and your audience.
- Finding/creating good images takes times. But its worth it.

## GET INSPIRED

- Lots of conferences have their talks on line. Watch some with some friends and discuss what works and what does not.
- Watch some (technically oriented) TED talks. (Some TedX are good, but you don't get to TED unless you are a good speaker).

## 7 TALK TIPS (FROM INSPIRING TED SPEAKERS)

1. Immerse your audience in a story.
2. Tell an emotional or personal story.
3. Create suspense
4. Bring characters to life.
5. Show. Don't tell.
6. Build up to a S.T.A.R. moment.
7. End with a positive takeaway.

DUARTE, NANCY. “RESONATE.” DUARTE. 2014.

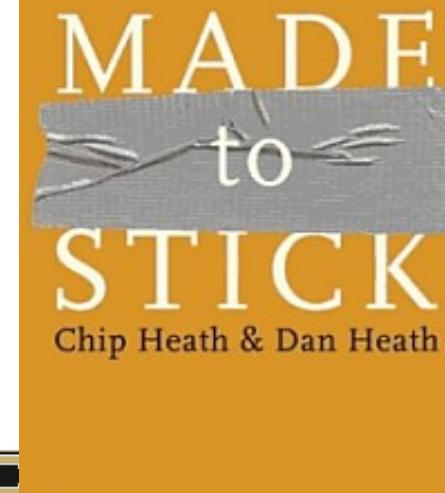
- A successful presentation creates a connection with the audience. In other words, it has to have “Something They'll Always Remember.” The **STAR moment** doesn't have to be particularly big or flashy, but it needs to be awe-inspiring. According to Duarte, it's all about creating a “significant, sincere, and enlightening moment...that helps magnify your big idea.”
- **1.) Memorable Dramatization**
- **2.) Repeatable Sound Bites**
- **3.) Evocative Visuals**
- **4.) Emotive Storytelling**
- **5.) Shocking Statistics**

## #5 STICKY STORIES

- Think about everything you say in terms of its impact on your “story” and its stickiness



Why Some Ideas Survive  
and Others Die



# SUCCESS: STICKY STORIES

**S**imple

**U**nexpected

**C**oncrete

**C**redible

**E**motional

**S**tories

**S**tick!

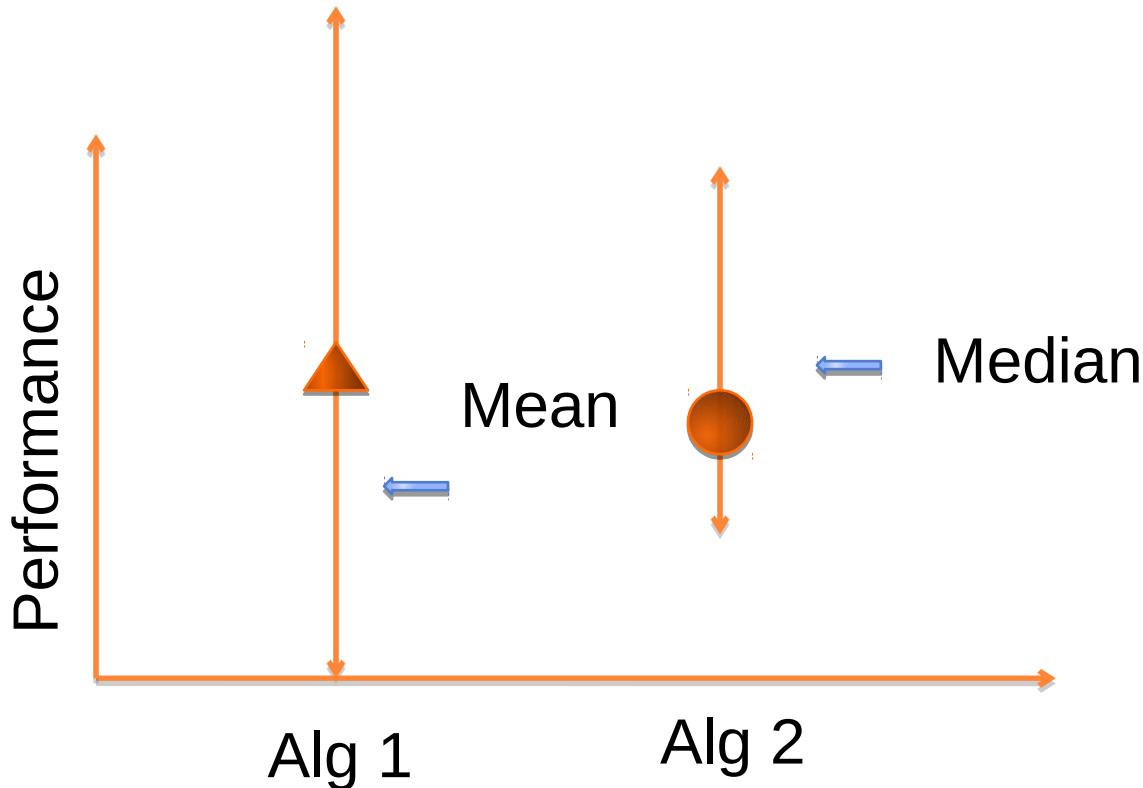
**■ PLAN FOR  
■ SUCCESS**

# Stats for Computer Science Experiments

by Jef Mallett

May 08, 2006





Parameter estimates or algorithm comparisons without realistic error bars are nearly useless. As a reviewer call people on it.. as an author expect reviewers to call you on it.

Always take the standard error into account when quoting summary statistics. Quoting a silly number of decimal places/significant figures is a sure indicator of statistical and/or scientific illiteracy.

mean of 10 runs of a program on single input with times  
3.01, 3.10 ... 3.19, 3.23  
don't report run time of 3.14159

Report  $3.1 \pm .05$  (where .05 is standard deviation),

# WHY STATISTICS

- Almost all experimental CS of interest involves noisy data.
- Need to separate our hypothesized meaningful variations from random variations caused by noise.
- Ideally, we want to show the difference in performance is *statistically significant*

# Experimental Hypothesis

- Any scientific research should state a hypothesis, then
  - Provide evidence for/against
  - Conclude whether it is supported or refuted
- Should be:
  - Precise, explicit statement
  - Falsifiable
- Run experiment to confirm or refute
- Ensure that the experiment really tests the hypothesis

# Elaborating the Hypothesis

- Hypothesis might be that system/theory/technique/ parameter P is:
  - Good for task X
  - Better than rivals Q and R for task X
- According to:
  - Behavior - correctness or quality of solution
  - Coverage - range of problems to which it applies
  - Efficiency - resources consumed
- Evidence can be theoretical, experimental or both
  - Theoretical evidence - theorem based
  - Experimental evidence – statistical testing on data

# Nonstatistical Hypothesis Testing...

A criminal trial is an example of hypothesis testing without the statistics.

In a trial a jury must decide between two hypotheses. The null hypothesis is

$H_0$ : The defendant is innocent

The alternative hypothesis or research hypothesis is

$H_1$ : The defendant is guilty

The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

# Nonstatistical Hypothesis Testing...

In the language of statistics convicting the defendant is called *rejecting the null hypothesis in favor of the alternative hypothesis*. That is, the jury is saying that there is enough evidence to conclude that the defendant is guilty (i.e., there is enough evidence to support the alternative hypothesis).

If the jury acquits it is stating that *there is not enough evidence to support the alternative hypothesis*.

Notice that the jury is not saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis.

# Nonstatistical Hypothesis Testing...

- There are two possible errors.
- A Type I error occurs when we reject a true null hypothesis. That is, a Type I error occurs when the jury convicts an innocent person. We would want the probability of this type of error [maybe 0.001 – beyond a reasonable doubt] to be very small for a criminal trial where a conviction results in the death penalty, whereas for a civil trial, where conviction might result in someone having to “pay for damages to a wrecked auto”, we would be willing for the probability to be larger [0.49 – preponderance of the evidence ]
- $P(\text{Type I error}) = \alpha$  [usually 0.05 or 0.01]

# Nonstatistical Hypothesis Testing...

- A Type II error occurs when we don't reject a false null hypothesis. That occurs when a guilty defendant is acquitted.
- While less critical in law, in experiments this type of error is by far the most serious mistake we normally make. For example, if we test the hypothesis that the amount of medication in a heart pill is equal to a value which will safely cure your heart problem and “accept” the null hypothesis that the amount is ok. Later on we find out that the average amount is WAY too large and people die from “too much medication” [I wish we had rejected the hypothesis and threw the pills in the trash can], it’s too late because we shipped the pills to the public.

# Nonstatistical Hypothesis Testing...

- The probability of a Type I error is denoted as  $\alpha$  (Greek letter *alpha*). The probability of a type II error is  $\beta$  (Greek letter *beta*).
- The two probabilities are inversely related -- for a fixed sample size, decreasing one increases the other.
- In other words, you can't force  $\alpha$  and  $\beta$  to both be real small for any given sample size. You may have to take a much larger sample size, or in the court example, you need much more evidence.

# Hypothesis Testing...

- The critical concepts are theses:
  - 1. There are two hypotheses, the null and the alternative hypotheses.
  - 2. **The procedure begins with the assumption that the null hypothesis is true.**
  - 3. The goal is to determine whether there is enough evidence to infer that the alternative hypothesis is true, or **the null is not likely to be true.**
  - 4. There are two possible decisions:
    - Conclude that there is enough evidence to support the alternative hypothesis. **Reject the null.**
    - Conclude that there is *not* enough evidence to support the alternative hypothesis. **Fail to reject the null.**

# Concepts of Hypothesis Testing (1)...

- The **two** hypotheses are called the *null hypothesis* and the other the *alternative* or *research hypothesis*. The usual notation is:

pronounced  
H “nought”

- $H_0$ : — *the ‘null’ hypothesis*
- $H_1$ : — *the ‘alternative’ or ‘research’ hypothesis*
- The null hypothesis ( $H_0$ ) will always state that the **parameter equals the value** specified in the alternative hypothesis ( $H_1$ )

# Concepts of Hypothesis Testing (4)...

- The **two** possible decisions that can be made:
  - ➔ Conclude that there ***is enough evidence*** to support the alternative hypothesis  
(also stated as: **reject the null hypothesis** in favor of the alternative)
  - ➔ Conclude that there ***is not enough evidence*** to support the alternative hypothesis  
(also stated as: **failing to reject** the null hypothesis in favor of the alternative)

NOTE: we **do not** say that we **accept** the null hypothesis especially **if a statistician is around...**

# Statistical Hypothesis testing

- Tests of statistical significance assuming some model of and random noise
- Data not consistent with  $H_0$ :
  - $H_0$  can be rejected in favour of some alternative hypothesis  $H_1$  (the objective of our study).
- Data are consistent with the  $H_0$  :
  - $H_0$  cannot be rejected
    - You cannot say that the  $H_0$  is true.
    - You can only decide to reject it or not reject it.

# **p value**

**p value** = probability that our result (e.g. a difference between proportions or a RR) or more extreme values could be observed under the null hypothesis

$H_0$  rejected using reported *p* value

## ***p* values – practicalities**

**Low *p* values = low degree of compatibility between  $H_0$  and the observed data:**

- association unlikely to be by chance
- you reject  $H_0$ , the test is significant

**High *p* values = high degree of compatibility between  $H_0$  and the observed data:**

- association likely to be by chance
- you don't reject  $H_0$ , the test is not significant

# Levels of significance – practicalities

We need of a cut-off !

1%      5%      10%

p value > 0.05 =  $H_0$  not rejected (**non significant**)

p value  $\leq 0.05$  =  $H_0$  rejected (**significant**)

BUT:

Better to give the p-value rather than “significant”  
vs. “non-significant”.

# Examples from the literature

- "The limit for statistical significance was set at p=0.05."
- "There was a strong relationship ( $p<0.001$ )."
- "..., but it did not reach statistical significance (ns)."
- „The relationship was statistically significant ( $p=0.0361$ )”

$p=0.05 \sqsubseteq$  Agreed convention  
Not an absolute truth

*“Surely, God loves the 0.06 **nearly as much as the 0.05**” (Rosnow and Rosenthal, 1991)*

# $p = 0.05$ and its errors

- Level of significance, usually  $p = 0.05$
- $p$  value used for decision making

**But still 2 possible errors:**

$H_0$  should not be rejected, but it was rejected :

▫ Type I or alpha error

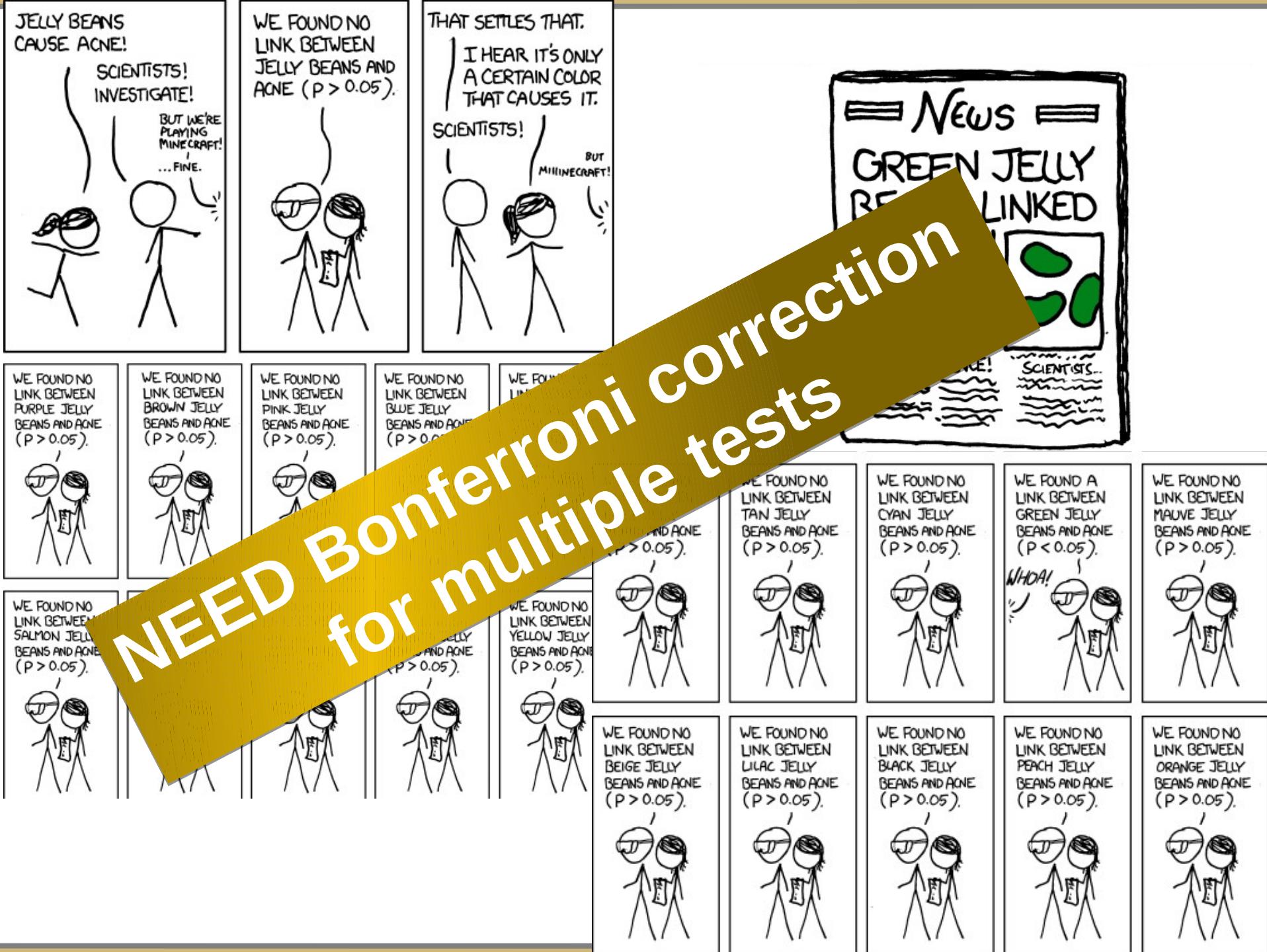
$H_0$  should be rejected, but it was not rejected :

▫ Type II or beta error

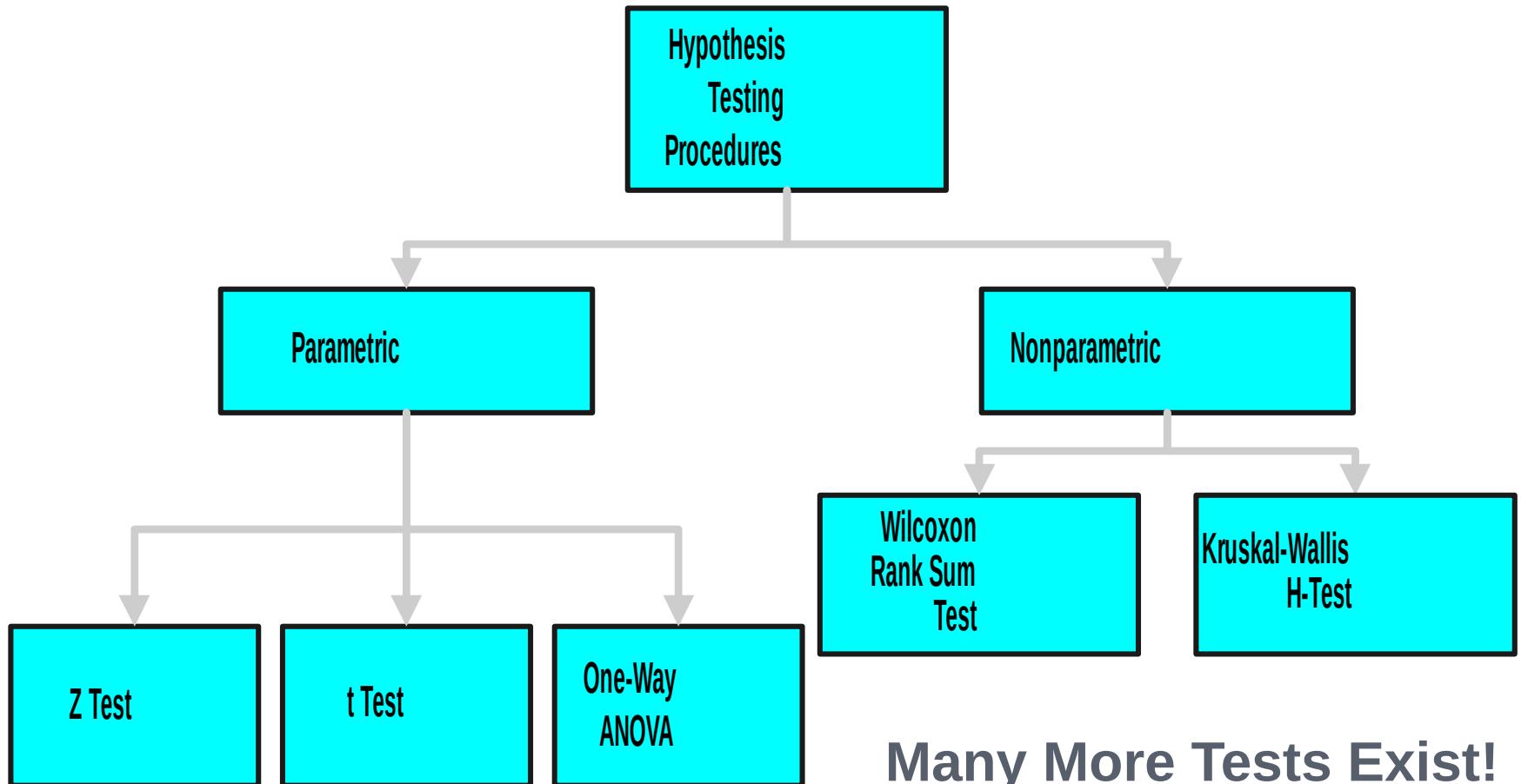
# Types of errors

|                                      |   | Truth                             |   |
|--------------------------------------|---|-----------------------------------|---|
|                                      |   | No diff                           | Diff  |
| Decision based on the <i>p</i> value | H <sub>0</sub> not rejected               | H <sub>0</sub> to be not rejected | H <sub>0</sub> to be rejected (H <sub>1</sub> ) |
|                                      | No diff                                   | Right decision<br>1- $\alpha$     | $\beta$<br><b>Type II error</b>                 |
|                                      | H <sub>0</sub> rejected (H <sub>1</sub> ) | $\alpha$<br><b>Type I error</b>   | Right decision<br>1- $\beta$                    |
|                                      | Diff                                      |                                   |   |

- H<sub>0</sub> is “true” but rejected: Type I or  $\alpha$  error
- H<sub>0</sub> is “false” but not rejected: Type II or  $\beta$  error



# Hypothesis Testing Procedures



# Parametric Test Procedures

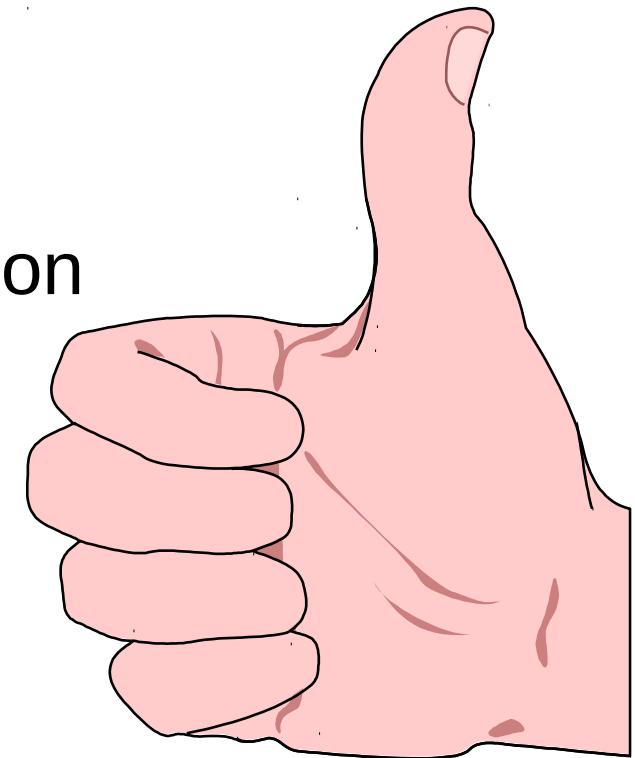
1. Involve Population Parameters (Mean)
2. Have Stringent Assumptions  
(Normality)
3. Examples: Z Test, t Test,  $\chi^2$  Test,  
F test

# Nonparametric Test Procedures

1. Do Not Involve Population Parameters  
Example: Probability Distributions, Independence
2. Data Measured on Any Scale (Ratio or Interval, Ordinal or Nominal)
3. Example: Wilcoxon Rank Sum Test

# Advantages of Nonparametric Tests

1. Used With All Scales
2. Easier to Compute
3. Make Fewer Assumptions
4. Need Not Involve Population Parameters
5. Results May Be as Exact as Parametric Procedures

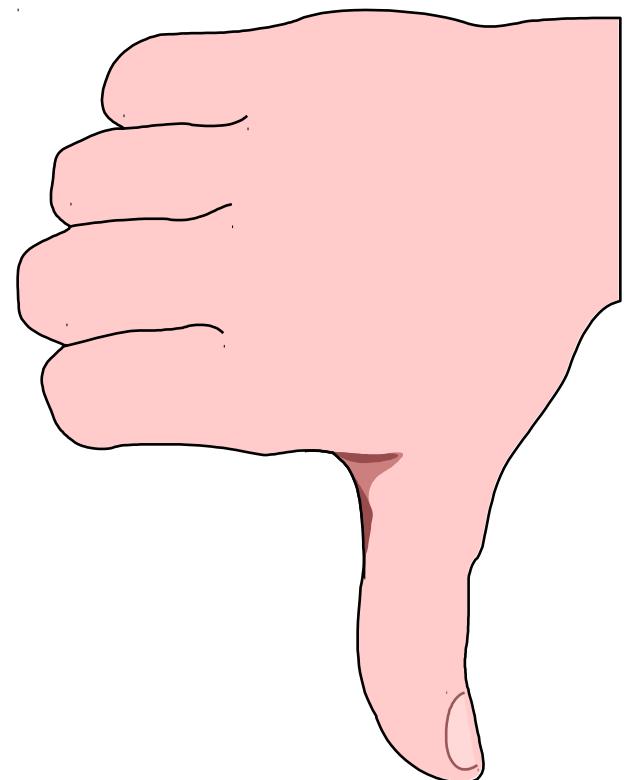


© 1984-1994 T/Maker Co.

# Disadvantages of Nonparametric Tests

1. May Waste Information  
Parametric model more efficient  
if data Permit
2. Difficult to Compute by  
hand for Large Samples  
(But easy to compute ;-)
3. Less widely known

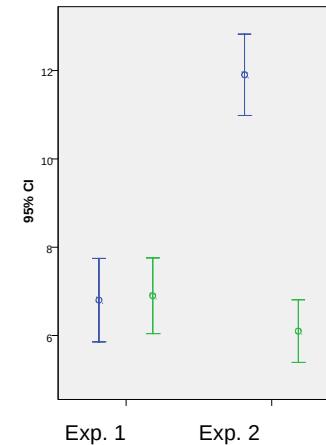
© 1984-1994 T/Maker Co.



# Parametric tests-nonparametric equivalent

- Paired t-test – Wilcoxon signed rank
- Two sample t-test – Wilcoxon rank sum
- ANOVA – Kruskal-Wallis test
  - When you have two or more independent samples and the assumptions of ANOVA are not met, you can use the Kruskal-Wallis test. This is a rank based test.

# T-TESTS



- Compare the **mean** between 2 samples/ conditions
- if 2 samples are taken from the same population, then they should have fairly similar means
  - if 2 means are statistically different, then the samples are likely to be drawn from 2 different populations, i.e., **they really are different**

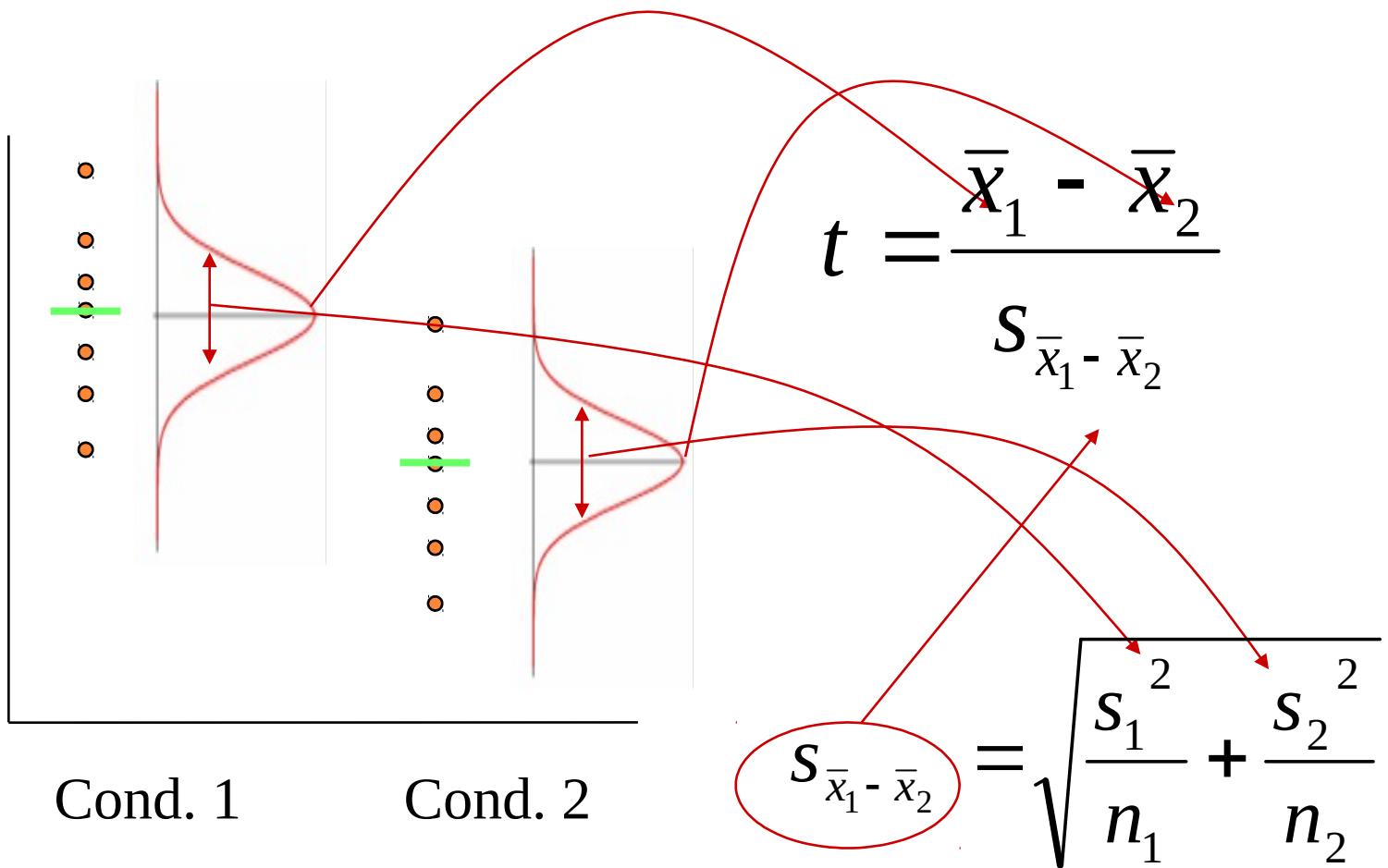
# FORMULA

Difference between the means divided by the pooled **standard error of the mean**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

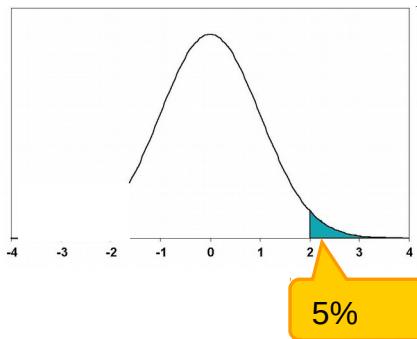
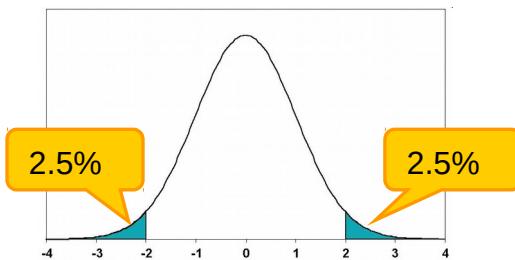
**Reporting convention: t= 11.456, df= 9, p< 0.001**

# FORMULA CONT.



# TYPES OF T-TESTS CONT.

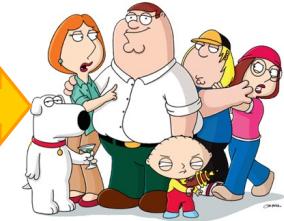
■ 2-tailed tests vs one-tailed tests



■ 2 sample t-tests vs 1 sample t-tests



Mean



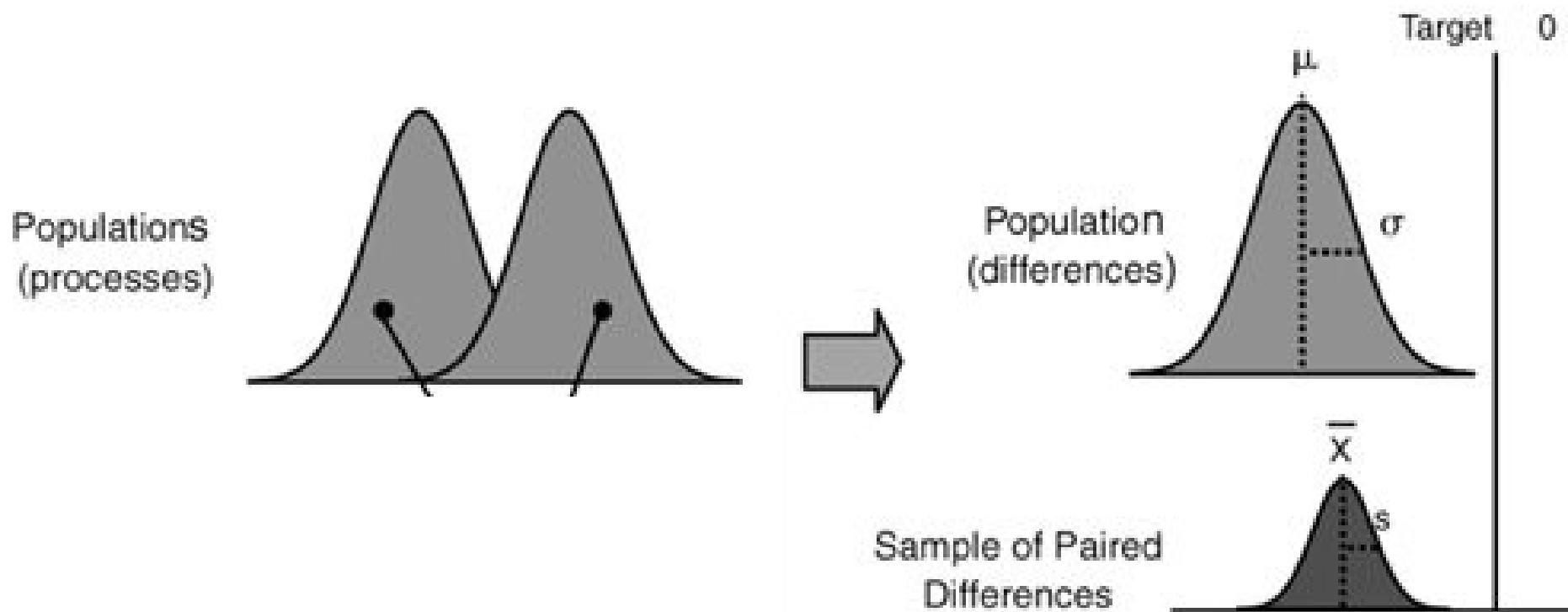
Mean



Mean

A known value

# Why Paired Test



# TYPES OF T-TESTS

|                                      | <b>Independent Samples</b>  | <b>Related Samples</b><br>also called dependent means test   |
|--------------------------------------|---|--|
| <b>Interval measures/ parametric</b> | Independent samples t-test*<br>  | Paired samples t-test**<br> |
| <b>Ordinal/ non-parametric</b>       | Mann-Whitney U-Test   | Wilcoxon test  |

\* 2 experimental conditions and different participants were assigned to each condition

\*\* 2 experimental conditions and the same participants took part in both conditions of the experiments

# Excel: TTEST function

- TTEST (array1, array2, tails, type)
  - array1 = the cell address for the first set of data
  - array2 = the cell address for the second set of data
  - tails: 1 = one-tailed, 2 = two-tailed
  - type: 1 = a paired t test; **2 = a two-sample test (independent with equal variances)**; 3 = a two-sample test with unequal variances

# Excel: TTEST function

- It does not compute the t value
- It returns the likelihood that the resulting t value is due to chance (the possibility of the difference of two groups is due to chance)

# Wilcoxon signed rank test

- To test difference between paired data when data is not “normal” (guassian) e.g. it is ordinal (e.g. survey data)

# **STEP 1**

- **Exclude any differences which are zero**
- **Put the rest of differences in ascending order**
- **Ignore their signs**
- **Assign them ranks**
- **If any differences are equal, average their ranks**

## STEP 2

- Count up the ranks of +ives as  $T_+$
- Count up the ranks of -ives as  $T_-$

## STEP 3

- If there is no difference between Treatment( $T_+$ ) and baseline ( $T_-$ ), then  $T_+$  &  $T_-$  would be similar
- If there were a difference one sum would be much smaller and the other much larger than expected
- The smaller sum is denoted as  $T$
- $T = \text{smaller of } T_+ \text{ and } T_-$

## STEP 4

- Compare the value obtained with the critical values (5%, 2% and 1% ) in appropriate table
- N is the number of differences that were ranked (**not the total number of differences**)
- So the zero differences are excluded

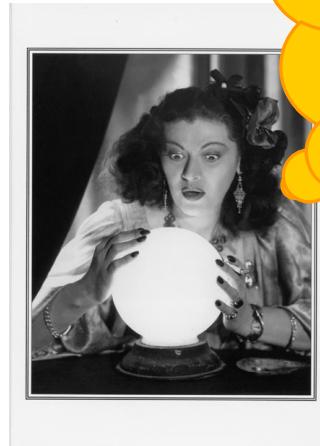
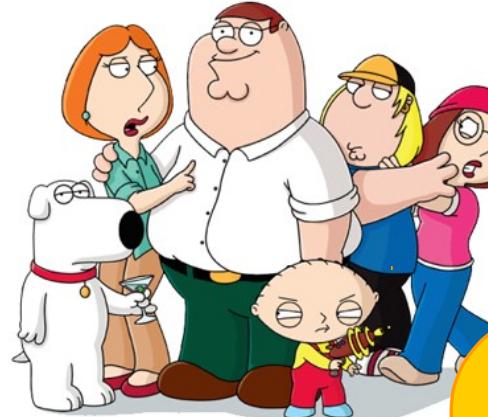
| Subject | Hours of sleep |          | Difference | Rank Ignoring sign |
|---------|----------------|----------|------------|--------------------|
|         | Treatment      | Baseline |            |                    |
| 1       | 6.1            | 5.2      | 0.9        | 3.5*               |
| 2       | 7.0            | 7.9      | -0.9       | 3.5*               |
| 3       | 8.2            | 3.9      | 4.3        | 10                 |
| 4       | 7.6            | 4.7      | 2.9        | 7                  |
| 5       | 6.5            | 5.3      | 1.2        | 5                  |
| 6       | 8.4            | 5.4      | 3.0        | 8                  |
| 7       | 6.9            | 4.2      | 2.7        | 6                  |
| 8       | 6.7            | 6.1      | 0.6        | 2                  |
| 9       | 7.4            | 3.8      | 3.6        | 9                  |
| 10      | 5.8            | 6.3      | -0.5       | 1                  |

3<sup>rd</sup> & 4<sup>th</sup> ranks are tied hence averaged

T= smaller of T<sub>+</sub> (50.5) and T<sub>-</sub> (4.5)

Here T=4.5 significant at 2% level indicating the Treatment is effective

# COMPARISON OF MORE THAN 2 GROUPS



Tell me the difference between these groups...  
Thank God I have ANOVA

# Multiple paired comparisons

- Let  $a_c$  be the probability of an error in a single comparison
  - alpha = the probability of incorrectly rejecting null hypothesis
- $1-a_c$ : probability of making **no error** in a single comparison
- $(1-a_c)^m$ : probability of no error in  $m$  comparisons (experiment)
- $a_e = 1-(1-a_c)^m$ : probability of an error in the experiment
  - Under assumption of independent comparisons

$a_e$  quickly becomes large as  $m$  increases

3. **One-Way ANOVA** – Similar to a t-test, except that this test can be used to compare the means from **THREE OR MORE** groups (t-tests can only compare TWO groups at a time, and for statistical reasons it is generally considered “illegal” to use t-tests over and over again on different groups from a single experiment).
4. **Two-Way ANOVA** – A very useful statistical test, because it’s the only one that allows you to compare the means of **TWO OR MORE** groups in response to **TWO DIFFERENT INDEPENDENT VARIABLES**. With this test available, you can set up an experiment in which each member of your sample is exposed to a varying level of *two different treatments!!!* In a field study, this test allows you to compare a mean Response Variable relative to *two different* environmental conditions.

# Types of analysis-independent samples

| Outcome       | Explanatory | Analysis                             |
|---------------|-------------|--------------------------------------|
| Continuous    | Dichotomous | t-test, Wilcoxon test                |
| Continuous    | Categorical | ANOVA, linear regression             |
| Continuous    | Continuous  | Correlation, linear regression       |
| Dichotomous   | Dichotomous | Chi-square test, logistic regression |
| Dichotomous   | Continuous  | Logistic regression                  |
| Time to event | Dichotomous | Log-rank test                        |

# Example

- A recent study compared the hypointensity of gray matter structures on MRI in normal controls, benign MS patients and secondary progressive MS patients
- Increased hypointensity is a marker of disease
- Question: Is there any difference among these groups?

- The null hypothesis is that all of the groups have the same hypointensity on average
  - Categorical predictor
  - Continuous outcome
- You could compare each of the groups to each of the other groups which would be 3 pair wise comparisons at the 0.05 level, but what happens to the overall alpha level?
- What is  $\alpha$ ?
  - $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$  so in this case  $\alpha = P(\text{one difference} \mid \text{all are equal})$
- Also,  $P(\text{fail to reject } H_0 \mid H_0 \text{ is true}) = 1 - \alpha$

# Overall $\alpha$ level

- Now, if we completed each of the 3 pair wise tests at the 0.05 level and all of the tests were independent,  $P(\text{fail to reject all 3 hypotheses } | H_0 \text{ is true}) = (1-0.05)^3 = 0.857$
- Therefore,  $P(\text{reject at least 1 } | H_0 \text{ is true}) = 1 - 0.857 = 0.143 = \alpha = \text{ type I error}$
- Type I error is greater than 0.05. This gets worse as number of comparisons increases
- What can we do?
  - ANOVA

# Analysis of variance (ANOVA)

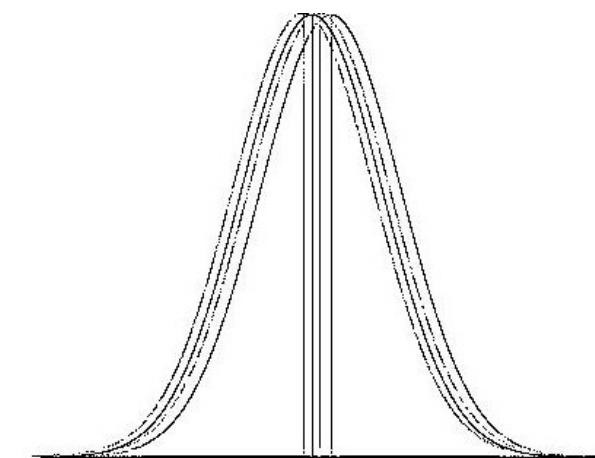
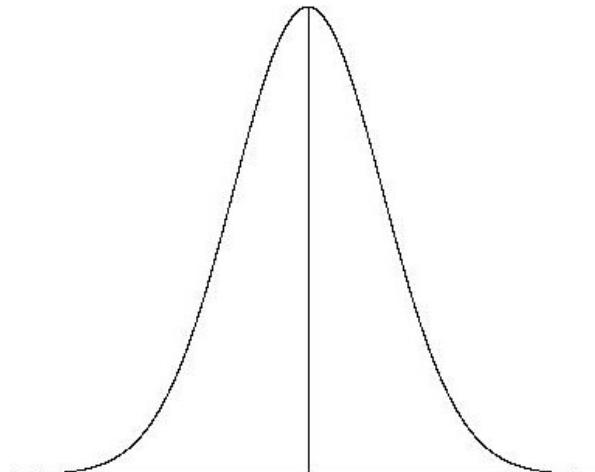
- Null hypothesis is  $\mu_1 = \mu_2 = \dots = \mu_n$
- We are testing if the mean is equal across groups
- The alternative hypothesis is that at least one of the means is different (but we will not be able to determine which one using this test)
- The name tells us that we are going to be using the variance, but the goal is to use the variance to compare the means (this is a common source of confusion)

# How does this work?

- As with the t-test, we have a **continuous outcome**, but now we have multiple groups, which is a **categorical variable**
- Before we begin, we must consider the assumptions required to use ANOVA
  - The underlying distributions of the populations are normal
  - The variance of each group is equal (This is critical for ANOVA), **homoscedastic**
- These are similar to the two sample t-test

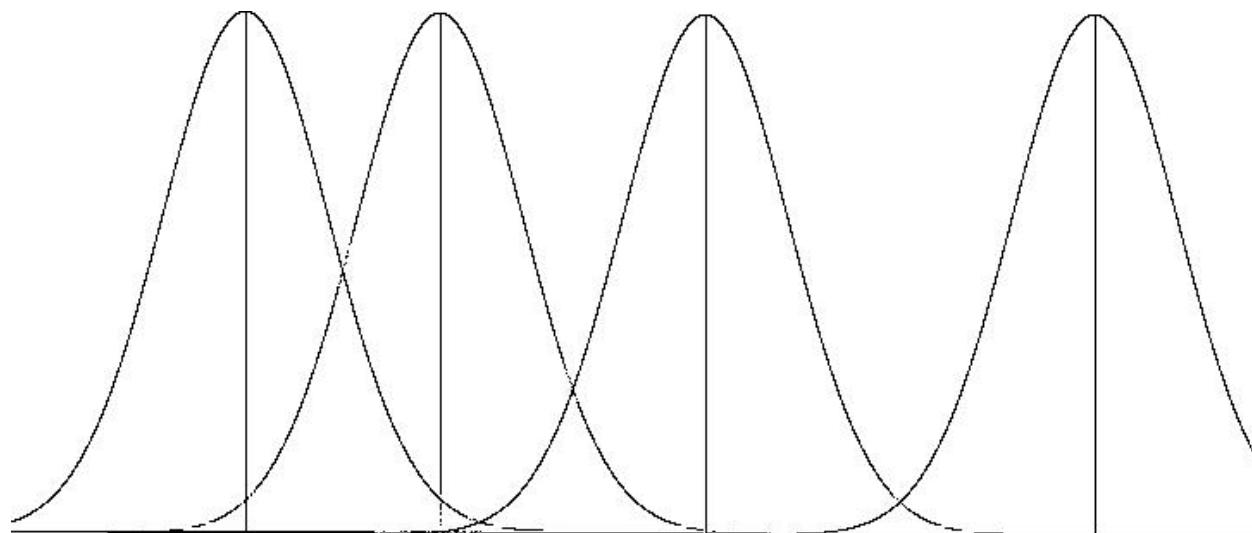
# Picture

- If all of the groups had the same means, the distributions for all of the populations would look exactly the same (overlaid graphs)



## Picture II

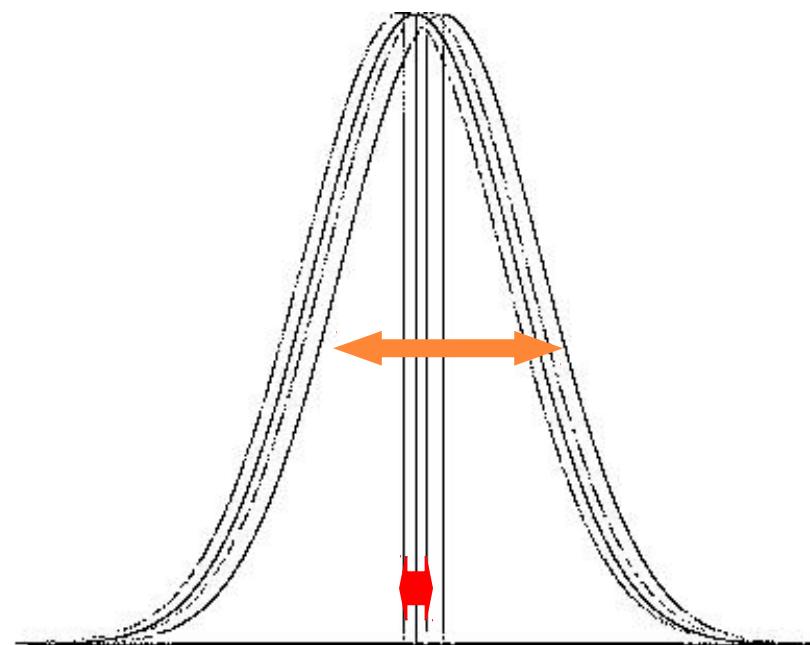
- Now, if the means of the populations were different, the picture would look like this. Notice that the variability between the groups is much greater than within a group



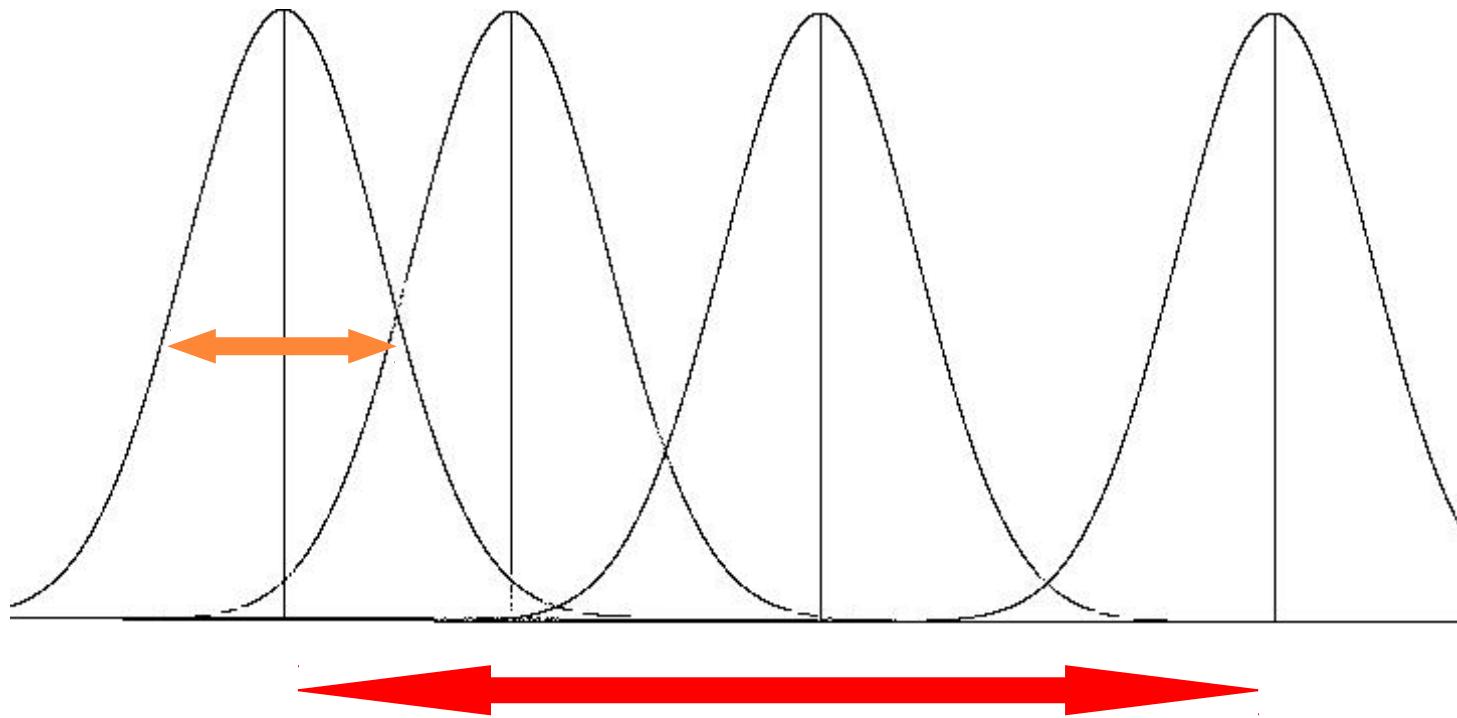
# Sources of variance

- When we take samples from each group, there will be two sources of variability
  - Within group variability - when we sample from a group there will be variability from person to person in the same group
  - Between group variability – the difference from group to group
    - If the between group variability is large, the means of the two groups are likely not the same

- We can use the two types of variability to determine if the means are likely different
- How can we do this?
- Look again at the picture
- Blue arrow: within group, red arrow: between group



- Blue arrow: within group, red arrow: between group
- Notice that when the distribution are separate, the between group variability is much greater than the within group



# Notation

- First we will define

$x_{ij}$  = observation from student i from group j

$\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$  mean of group j

$\bar{\bar{x}} = \frac{\sum_j n_j \bar{x}_j}{\sum_j n_j}$  grand mean over all of the groups

- How could we express the different forms of variability?

# Sources of variability

- The distance of each observation from the grand mean can be broken into two pieces

$$x_{ij} - \bar{x} = x_{ij} - \bar{x} + \bar{x}_j - \bar{x}_j = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x})$$



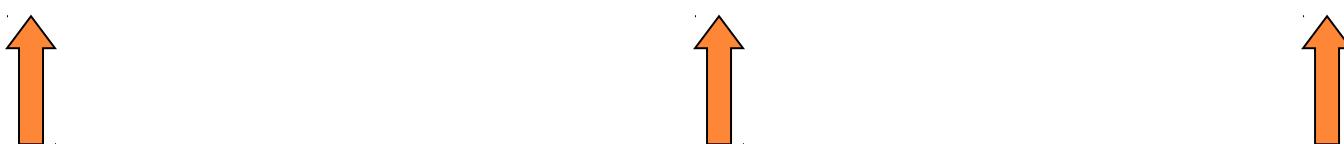
Within group  
variability



Between group  
variability

- Like the calculation of the variance, we are interested in the square of the deviation
- What does the squared deviation look like?

- The final squared deviation simplifies to

$$\sum_{j=1}^3 \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = \sum_{j=1}^3 \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^3 \sum_{i=1}^{n_j} (\bar{x}_j - \bar{x})^2$$


Total sum of squares (SS<sub>T</sub>)

Within group sum of squares (SS<sub>w</sub>)

Between group sum of squares (SS<sub>B</sub>)

- As we discussed earlier, we are going to compare the two errors to determine if the group means are equal

- The within group variability can be written in terms of the individual group standard deviations,  $s_i$

$$SS_W = \sum_{j=1}^3 \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^3 (n_j - 1)s_j^2$$

- The result is called the within group mean square error, which is the combined estimate of the within group variance

$$MS_W = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n_1 + n_2 + n_3 - 3}$$

- Note the denominator is the total sample size minus the number of groups

- The between group variability can be broken into pieces from the summary statistics as well

$$SS_B = \sum_{j=1}^3 \sum_{i=1}^{n_j} \left( \bar{x}_j - \bar{x} \right)^2 = \sum_{j=1}^3 n_j \left( \bar{x}_j - \bar{x} \right)^2$$

- The between group mean square error can be written as

$$MS_B = \frac{\sum_{j=1}^3 n_j \left( \bar{x}_j - \bar{x} \right)^2}{3 - 1}$$

- The denominator of the  $MS_B$  is the number of groups minus 1 because we are considering the group means as the observations and the grand mean as the mean

# F-statistic

- Now that we have estimates of the between group and within group variation, we can use an F-statistic

$$F_{k-1, n-k} = \frac{MS_B}{MS_W} = \frac{SS_B / (k - 1)}{SS_W / (n - k)}$$

where k is the number of groups and n is the total sample size

- This test statistic is compared to an F-statistic with k-1 and n-k degrees of freedom

## ANOVA table

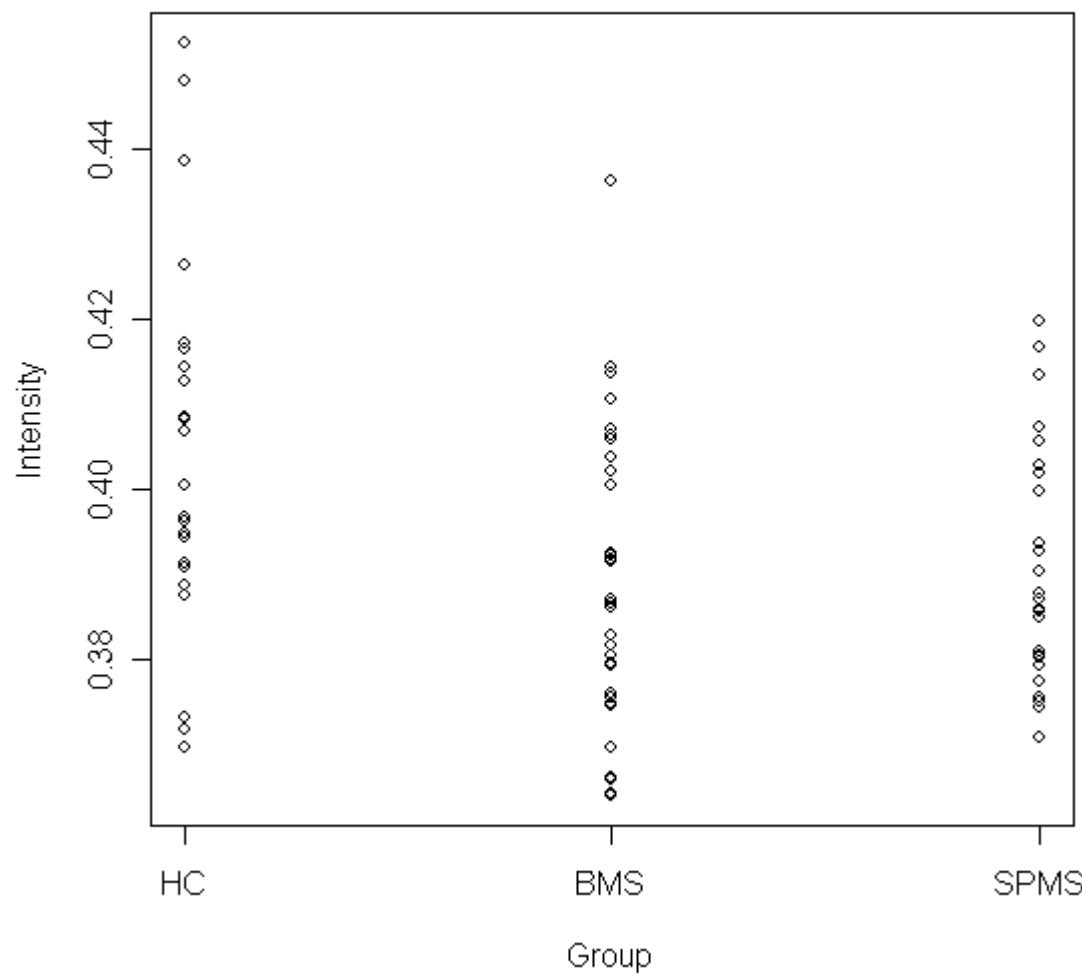
- To complete the analysis, we need to calculate the SS's, MS's and the F-statistic
- A specific display of this data is often used called the ANOVA table
- Standard software may provide results in this form

| Source of variation | SS     | df    | MS     | F           | p-value |
|---------------------|--------|-------|--------|-------------|---------|
| Between             | $SS_B$ | $k-1$ | $MS_B$ | $MS_B/MS_W$ |         |
| Within              | $SS_W$ | $n-k$ | $MS_W$ |             |         |
| Total               | $SS_T$ |       |        |             |         |

# Example

- Let's perform an ANOVA test for the hypointensity
- Here are the summary statistics

|                    | Healthy | BMS   | SPMS  |
|--------------------|---------|-------|-------|
| Mean               | 0.404   | 0.389 | 0.391 |
| Standard deviation | 0.022   | 0.017 | 0.014 |
| Sample size        | 24      | 35    | 26    |



# Hypothesis test

- 1)  $H_0: \mu_1 = \mu_2 = \mu_3$
- 2) Continuous outcome/categorical predictor
- 3) ANOVA
- 4) Test statistic:  $F=5.42$
- 5)  $p\text{-value}=0.0062$
- 6) Since the  $p$ -value is less than 0.05, we can reject the null hypothesis
- 7) We conclude that the mean is different in at least one group

# ANOVA table

- Here is the ANOVA table for this data

| Source of variation | SS     | df | MS      | F    | p-value |
|---------------------|--------|----|---------|------|---------|
| Between             | 0.0035 | 2  | 0.0017  | 5.42 | 0.0062  |
| Within              | 0.026  | 82 | 0.00032 |      |         |
| Total               |        |    |         |      |         |

## An even smaller example

Suppose we have three groups

- Group 1: 5.3, 6.0, 6.7
- Group 2: 5.5, 6.2, 6.4, 5.7
- Group 3: 7.5, 7.2, 7.9

We get the following statistics:

| SUMMARY  |       |      |          |          |  |
|----------|-------|------|----------|----------|--|
| Groups   | Count | Sum  | Average  | Variance |  |
| Column 1 | 3     | 18   | 6        | 0.49     |  |
| Column 2 | 4     | 23.8 | 5.95     | 0.176667 |  |
| Column 3 | 3     | 22.6 | 7.533333 | 0.123333 |  |

# COMPUTING ANOVA F STATISTIC

|          |       | <b>WITHIN</b>                            |       | <b>BETWEEN</b>                           |       |                |  |
|----------|-------|--|-------|--|-------|----------------|--|
|          |       | difference:<br>group - data - group mean |       | difference:<br>group mean - overall mean |       |                |  |
| data     | group | mean                                     | plain | squared                                  | plain | squared        |  |
| 5.3      | 1     | 6.00                                     | -0.70 | 0.490                                    | -0.4  | 0.194          |  |
| 6.0      | 1     | 6.00                                     | 0.00  | 0.000                                    | -0.4  | 0.194          |  |
| 6.7      | 1     | 6.00                                     | 0.70  | 0.490                                    | -0.4  | 0.194          |  |
| 5.5      | 2     | 5.95                                     | -0.45 | 0.203                                    | -0.5  | 0.240          |  |
| 6.2      | 2     | 5.95                                     | 0.25  | 0.063                                    | -0.5  | 0.240          |  |
| 6.4      | 2     | 5.95                                     | 0.45  | 0.203                                    | -0.5  | 0.240          |  |
| 5.7      | 2     | 5.95                                     | -0.25 | 0.063                                    | -0.5  | 0.240          |  |
| 7.5      | 3     | 7.53                                     | -0.03 | 0.001                                    | 1.1   | 1.188          |  |
| 7.2      | 3     | 7.53                                     | -0.33 | 0.109                                    | 1.1   | 1.188          |  |
| 7.9      | 3     | 7.53                                     | 0.37  | 0.137                                    | 1.1   | 1.188          |  |
| TOTAL    |       |  |       | 1.757                                    |       | 5.106          |  |
| TOTAL/df |       |  |       | <b>0.25095714</b>                        |       | <b>2.55275</b> |  |

overall mean: 6.44    $F = 2.5528/0.25025 = 10.21575$

# EXCEL ANOVA OUTPUT

| ANOVA               |          |    |          |          |          |          |
|---------------------|----------|----|----------|----------|----------|----------|
| Source of Variation | SS       | df | MS       | F        | P-value  | F crit   |
| Between Groups      | 5.127333 | 2  | 2.563667 | 10.21575 | 0.008394 | 4.737416 |
| Within Groups       | 1.756667 | 7  | 0.250952 |          |          |          |
| Total               | 6.884    | 9  |          |          |          |          |

1 less than number of groups

1 less than number of individuals  
(just like other situations)

number of data values -  
number of groups  
(equals df for each group added together)

# Notes

- Remember the assumption of normality of data and equal variance across groups is required
- We were able to conclude that one of the means is different, but we do not know which of the means is different. ANOVA is often considered a first step
- We can do pair wise comparisons to determine which specific means are different, but we must still take into account the problem with multiple comparisons

# Bonferroni correction

- The simplest way to handle the multiple comparisons is to correct the alpha level to allow the overall alpha level to be closer to the desired 0.05 level
- The Bonferroni correction takes the observed p-values and multiplies it by the number of comparisons
  - If we have 3 groups and we would like to complete all pair wise comparison, we multiply the p-values by 3
- In addition, we assume that the variance is equal in the pairwise t-tests

# MULTIPLE TESTING

- Say that you perform a statistical test with a 0.05 threshold, but you repeat the test on twenty different observations. Assuming that all of the observations are explainable by the null hypothesis, what is the chance that at least one of the observations will receive a p-value less than 0.05?
- $\text{Pr}(\text{making a mistake}) = 0.05$
- $\text{Pr}(\text{not making a mistake}) = 0.95$
- $\text{Pr}(\text{not making any mistake}) = 0.95^{20} = 0.358$
- $\text{Pr}(\text{making at least one mistake}) = 1 - 0.358 = 0.642$
- **There is a 64.2% chance of making at least one mistake.**

# Pairwise t-test

- Here are the pairwise t-test results

| Group 1 | Group 2 | p-value | Adjusted p-value |
|---------|---------|---------|------------------|
| HC      | BMS     | 0.0022  | 0.0065           |
| HC      | SPMS    | 0.014   | 0.042            |
| BMS     | SPMS    | 0.62    | 1.0              |

We conclude that there is a significant difference between the healthy controls and both groups of MS patients, but no difference between the two groups of MS patients.

# More on Bonferroni correction

- For three groups, we have three pairwise comparisons
- What if we were only interested in comparing each MS group to the healthy controls? How many comparisons would we need to correct for?
  - Two comparisons
  - Multiply each p-value by 2

JELLY BEANS  
CAUSE ACNE!

SCIENTISTS!  
INVESTIGATE!

BUT WE'RE  
PLAYING  
MINECRAFT!  
...FINE.

WE FOUND NO  
LINK BETWEEN  
JELLY BEANS AND  
ACNE ( $P > 0.05$ ).

THAT SETTLES THAT.  
I HEAR IT'S ONLY  
A CERTAIN COLOR  
THAT CAUSES IT.

SCIENTISTS!  
BUT  
MINICRAFT!

= News =

GREEN JELLY  
BEANS LINKED  
TO ACNE!

95% CONFIDENCE



ONLY 5% CHANCE  
OF COINCIDENCE!



SCIENTISTS...

WE FOUND NO  
LINK BETWEEN  
PURPLE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BROWN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
PINK JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BLUE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TEAL JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
GREY JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TAN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
CYAN JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND A  
LINK BETWEEN  
GREEN JELLY  
BEANS AND ACNE  
( $P < 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
MAUVE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
SALMON JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
RED JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
TURQUOISE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
MAGENTA JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
YELLOW JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BEIGE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
LILAC JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
BLACK JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

WE FOUND NO  
LINK BETWEEN  
ORANGE JELLY  
BEANS AND ACNE  
( $P > 0.05$ ).

What is the  
Bonferroni correction?

# BONFERRONI CORRECTION

- Assume that individual tests are *independent*. (Is this a reasonable assumption?)
- Divide the desired p-value threshold by the number of tests performed.
- For the previous example,  $0.05 / 20 = 0.0025$ .
- $\Pr(\text{making a mistake}) = 0.0025$
- $\Pr(\text{not making a mistake}) = 0.9975$
- $\Pr(\text{not making any mistake}) = 0.9975^{20} = 0.9512$
- $\Pr(\text{making at least one mistake}) = 1 - 0.9512 = 0.0488$

# TYPE I AND TYPE II ERRORS

- | **Definition:** A *Type I error ( $\alpha$ )* corresponds to the error of rejecting  $H_0$ , the null hypothesis, when it is, in fact, true. A *Type II error ( $\beta$ )* corresponds to the error of failing to reject  $H_0$  when it is false.
- | **Definition:** The power of a test is the probability of rejecting  $H_0$  given that it is false. Power =  $1 - \beta$

# ***WHY DOES POWER MATTER? I***

- All the hypothesis tests described in the previous three sections are only concerned about reducing the Type I error.
- i.e., they try to ascertain the conditions under which we are rejecting a hypothesis rightly.
- They are not at all concerned about the case where the null hypothesis is really false, but we do not reject it.

# **WHY DOES POWER MATTER? II**

- In the case of Machine Learning, reducing the type I error means reducing the probability of us saying that there is a difference in the performance of the 2 classifiers, when in fact, there isn't.
- Reducing the type II error means reducing the probability of us saying that there is no difference in the performance of the two classifiers, when, in fact, there is.
- Power matters because we do not want to discard a classifier that shouldn't have been discarded. If a test does not have enough power, then this kind of situation can arise

# WHAT IS THE EFFECT SIZE?

- The effect size measures how strong the relationship between two entities is.
- In particular, if we consider a particular procedure, in addition to knowing how statistically significant the effect of that procedure is, we may want to know what the size of this effect is.
- There are different measures of effect sizes, including:
  - Pearson's correlation coefficient
  - Odd's ratio
  - Cohen's d statistics
- Cohen's  $d$  statistic is appropriate in the context of a **t-test** on means. It is thus the effect size measure we concentrate on here.

[Wikipedia: [http://en.wikipedia.org/wiki/Effect\\_size](http://en.wikipedia.org/wiki/Effect_size)]

# COHEN'S D-STATISTICS

- Cohen's d-statistic is expressed as:  
$$d = (\bar{X}_1 - \bar{X}_2) / s_p$$
- Where  $s_p^2$ , the pooled variance estimate is:  
$$s_p^2 = \frac{((n_1-1)*s_1^2 + (n_2-1)*s_2^2)}{(n_1+n_2-2)}$$
- and  $s_p$ , its square root.
- [Note this is not exactly Cohen's d measure which is expressed in terms of parameters. This is a very usable estimate of d].

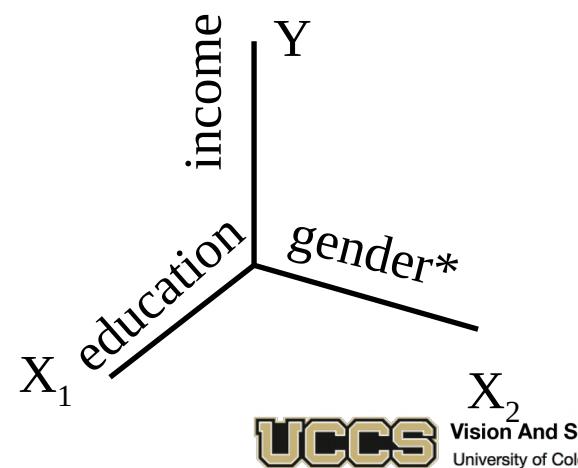
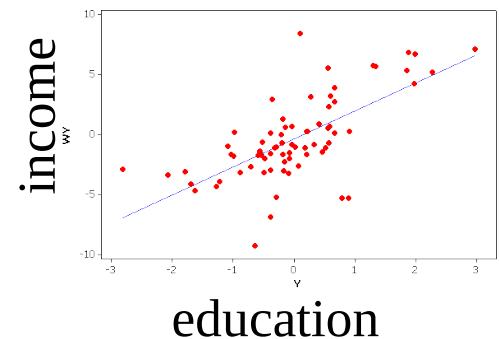
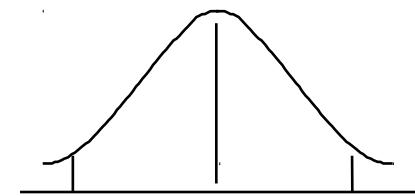
# USEFULNESS OF THE D STATISTIC

- d is useful in that it standardizes the difference between the two means. We can talk about deviations in terms of proportions of standard deviation points that are more useful than actual differences that are domain dependent.
- Cohen came up with a set of guidelines concerning d:
  - $d=.2$  has a small effect, but is probably meaningful;
  - $d= .5$  is a medium effect that is noticeable.
  - $d= .8$  shows a large effect size.

# Bivariate and Multivariate

- All measures so far have focused on one variable at a time
  - univariate
- Often, we are interested in the association or relationship between two variables
  - bivariate.
- Or more than two variables
  - multivariate

\*Gender = male or female



# CORRELATION AND REGRESSION

*THE MOST COMMONLY USED TECHNIQUES IN SCIENCE.*

- Review standard (non-spatial) approaches
- Correlation
- Regression

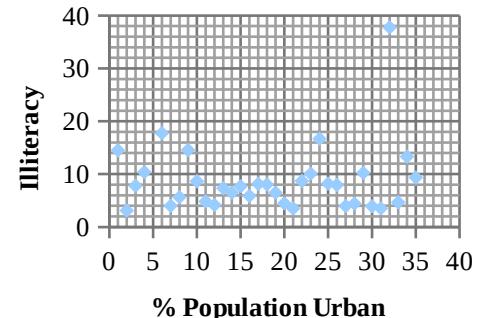
# Correlation and Regression

## *What is the difference?*

- Mathematically, they are identical.
- Conceptually, very different.

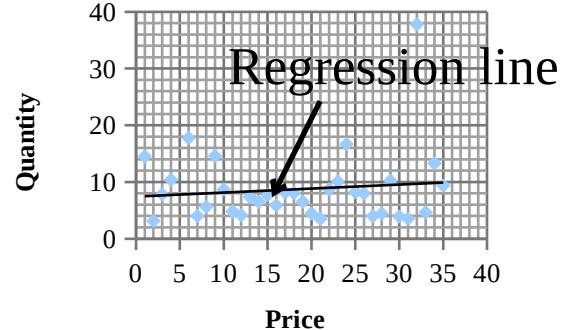
### Correlation

- Co-variation
- Relationship or association
- No direction or causation is implied
- $Y \longleftrightarrow X$        $X_1 \longleftrightarrow X_2$



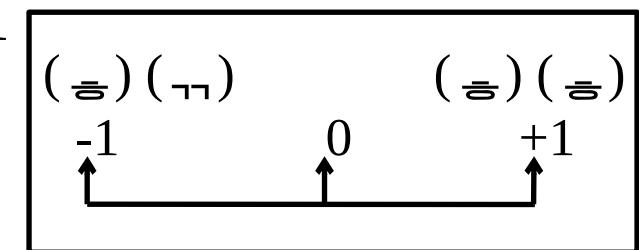
### Regression

- Prediction of Y from X
- Implies, but does not prove, causation
- X (independent variable) predicts Y (dependent variable)



# **CORRELATION COEFFICIENT ( $R$ )**

- The most common statistic in all of science
- measures the strength of the relationship (or “association”) between two variables e.g. income and education
- Varies on a scale from  $-1$  thru  $0$  to  $+1$



**+1** implies a perfect positive association

- As values go up ( $\uparrow$ ) on one, they also go up ( $\uparrow$ ) on the other
  - income and education

**0** implies no association

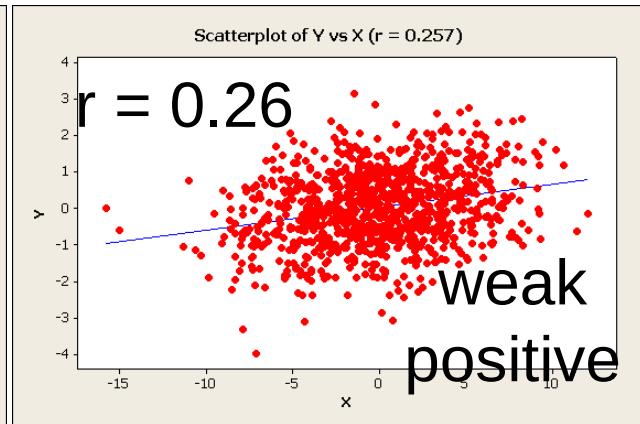
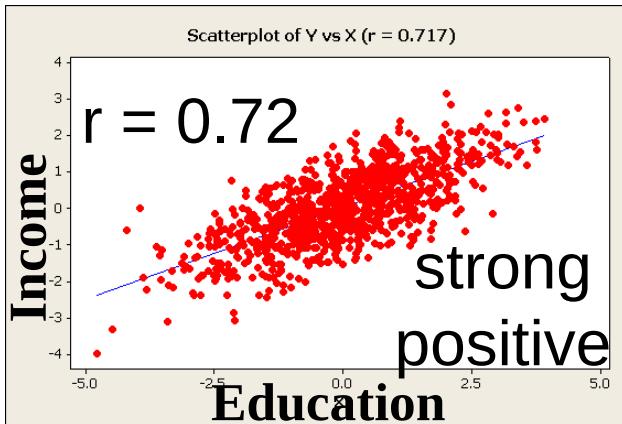
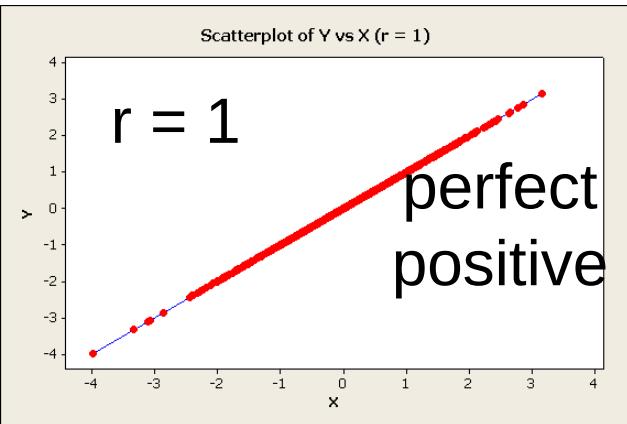
**-1** implies perfect negative association

- As values go up on one ( $\uparrow$ ), they go down ( $\downarrow$ ) on the other
  - price and quantity purchased

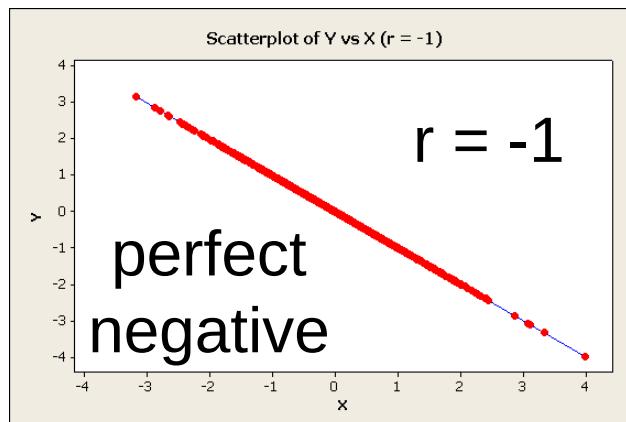
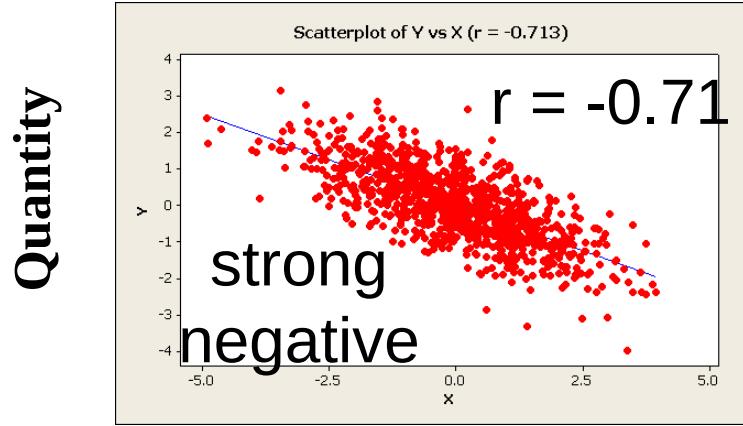
- Formal Full name is the

*Pearson Product Moment correlation coefficient,*

# Examples of Scatter Diagrams and the Correlation Coefficient Positive



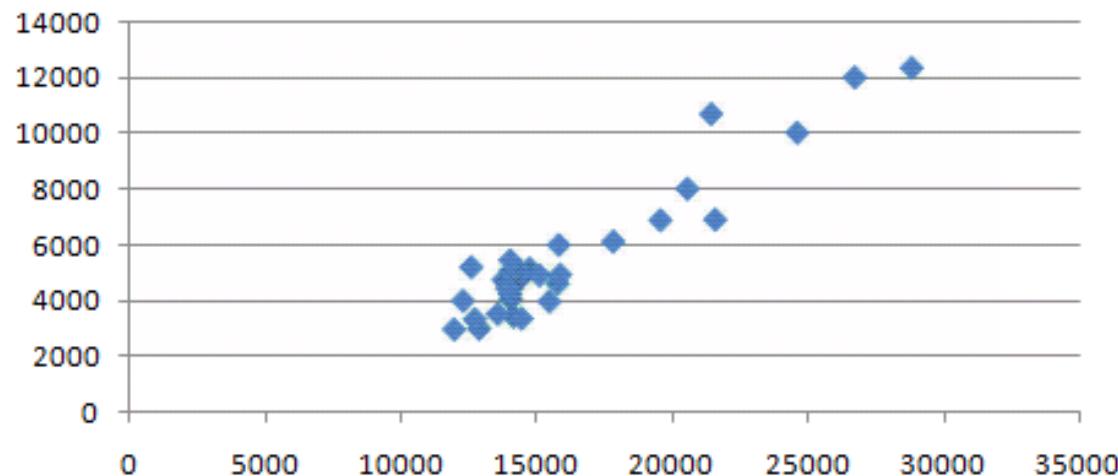
# Negative



# Correlation Coefficient: example

Rural Per capita Income (Y)  
related to Urban Income (X)

Correlation coefficient  
= 0.9458



China Provinces 29

excludes Xizang/Tibet, Macao, Hong Kong, Hainan, Taiwan, P'eng-hu

# Calculation Formulae for Correlation Coefficient ( $r$ )

*Before the days of computers, these formulae were easier to do “by hand.”*

$$r = \frac{\sum_{i=1}^n x_i y_i - \bar{X} \bullet \bar{Y}}{n S_x S_y},$$

$$S_x = \sqrt{\frac{\sum_{i=1}^n x^2}{n} - \bar{X}^2}$$

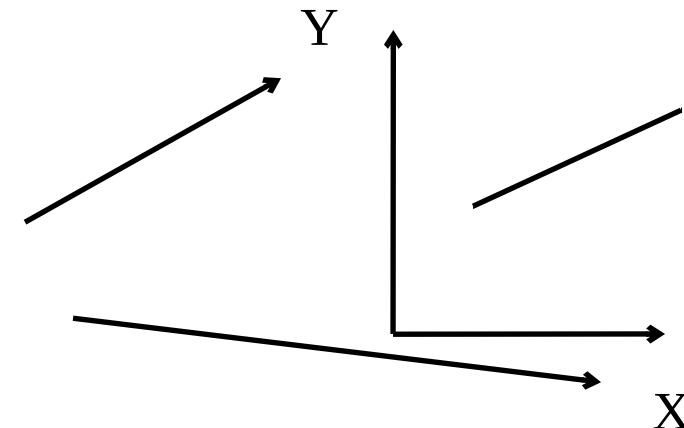
$$S_y = \sqrt{\frac{\sum_{i=1}^n y^2}{n} - \bar{Y}^2}.$$

*See next slide for example*

# Regression

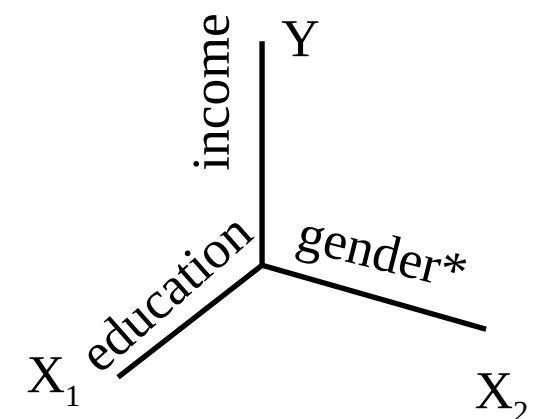
## Simple regression

- Between two variables
  - One dependent variable (Y)
  - One independent variable (X)



## Multiple Regression

- Between three or more variables
  - One dependent variable (Y)
  - Two or more independent variables ( $X_1, X_2, \dots$ )



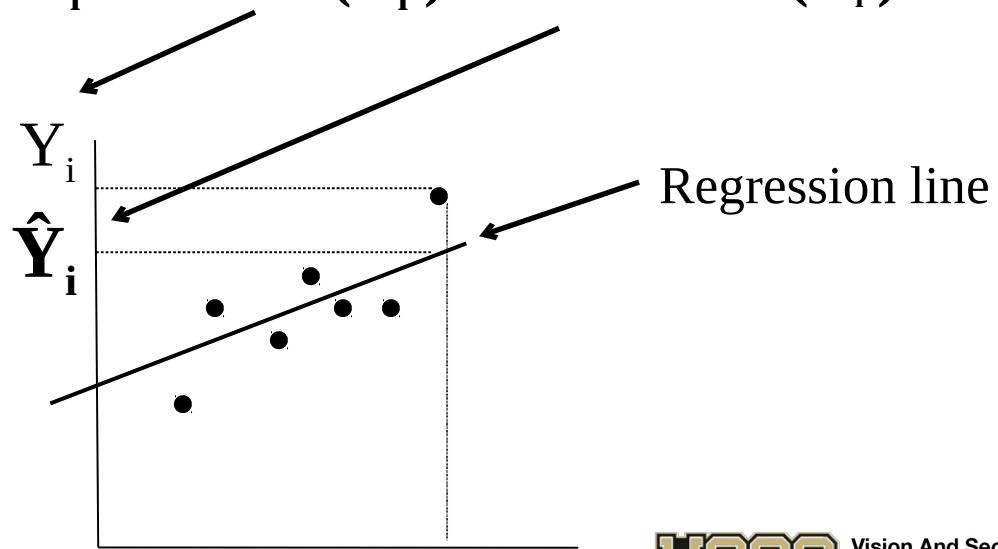
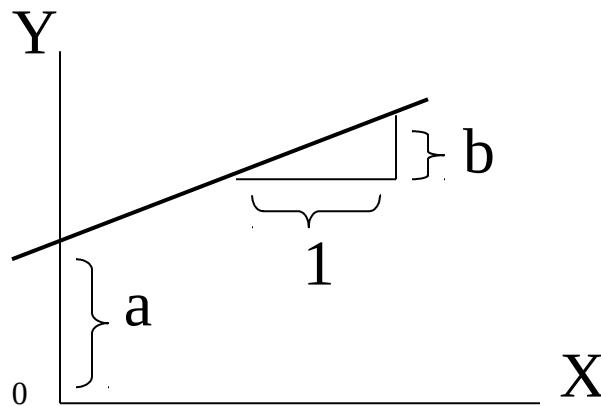
# Simple Linear Regression

- Concerned with “predicting” one variable ( $Y$  - the dependent variable) from another variable ( $X$  - the independent variable)

$$Y = a + bX + \varepsilon$$

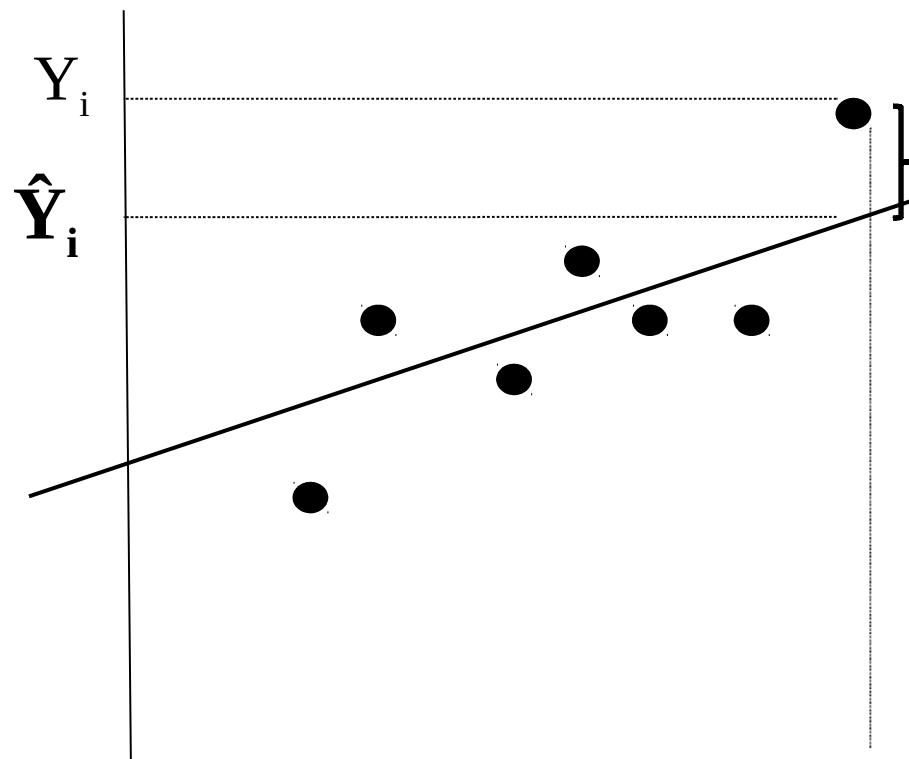
- a is the *intercept* —the value of  $Y$  when  $X = 0$
- b is the *regression coefficient* or slope of the line
  - the change in  $Y$  for a one unit change in  $X$

$$\varepsilon = \text{residual} = \text{error} = Y_i - \hat{Y}_i = \text{Actual } (Y_i) - \text{Predicted } (\hat{Y}_i)$$



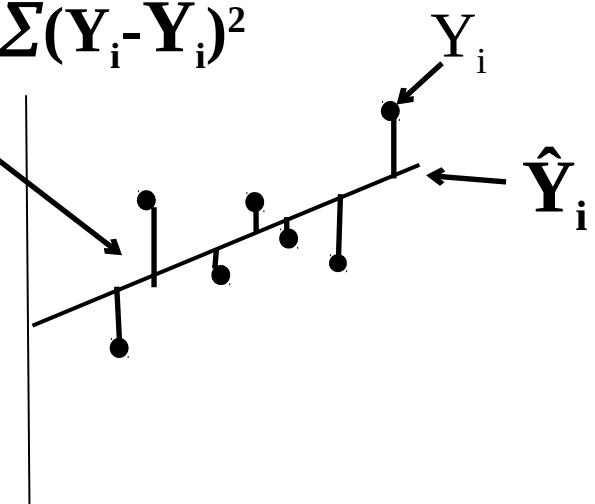
## Ordinary Least Squares (OLS)

--*the standard criteria for obtaining the regression line*



The regression line minimizes the sum of the squared deviations between actual  $Y_i$  and predicted  $\hat{Y}_i$

$$\text{Min } \sum(Y_i - \hat{Y}_i)^2$$



# Coefficient of Determination ( $r^2$ )

- The coefficient of determination ( $r^2$ ) measures the proportion of the variance in Y (the dependent variable) which can be predicted or “explained by” X (the independent variable). Varies from 1 to 0.
- It equals the correlation coefficient (r) squared.

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad \begin{matrix} \leftarrow & \text{SS Regression or Explained Sum of Squares} \\ \leftarrow & \text{SS Total or Total Sum of Squares} \end{matrix}$$

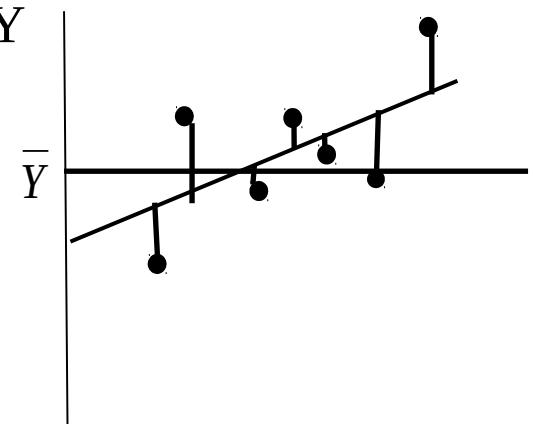
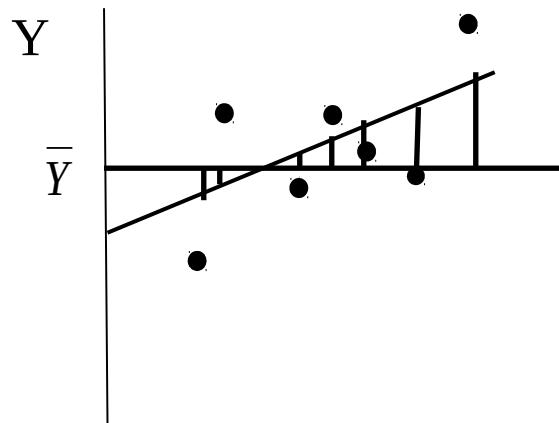
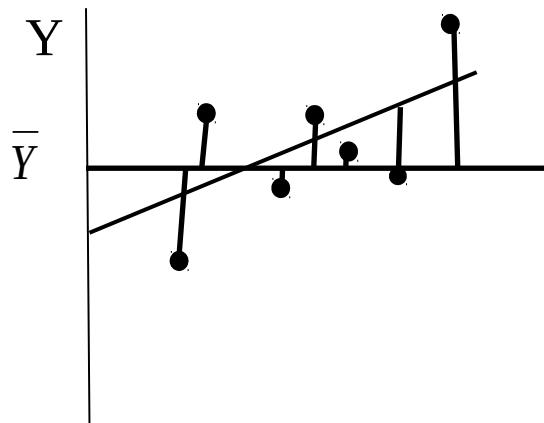
---

Note:

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

↑                              ↑                              ↑  
SS Total or            SS Regression or            SS Residual or  
Total Sum of            Explained Sum of            Error Sum of  
Squares                    Squares                    Squares

# Partitioning the Variance on Y



$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

↑  
SS Total  
or Total Sum of  
Squares

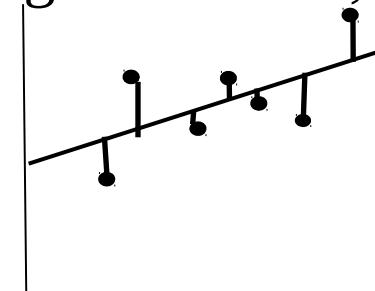
↑  
SS Regression  
or Explained Sum of  
Squares

↑  
SS Residual  
or Error Sum of Squares

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y}_i)^2}{\sum (Y_i - \bar{Y}_i)^2}$$

# Standard Error of the Estimate ( $s_e$ )

Measures *predictive accuracy*: the bigger the standard error, the greater the spread of the observations about the regression line, thus the predictions are less accurate



$S_e^2$  = error mean square, or average squared residual

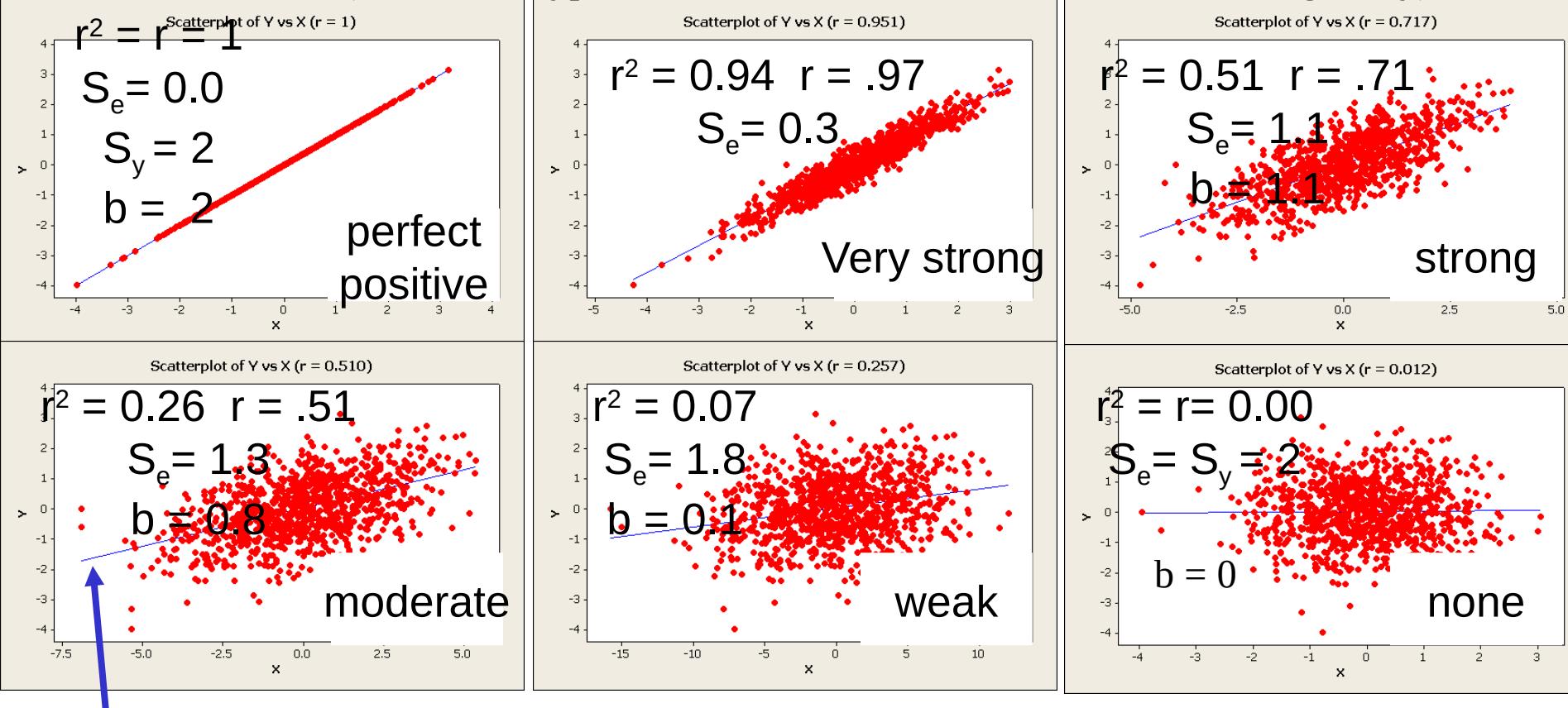
= variance of the estimate, variance about regression

$$S_e = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n - k}}$$

← Sum of squared residuals  
← Number of observations minus *degrees of freedom*  
(for simple regression, degrees of freedom = 2)

# Coefficient of determination ( $r^2$ ), correlation coefficient (r), regression coefficient (b), and standard error ( $S_e$ )

(Values are hypothetical and for illustration of relative change only)



As the coefficient of determination gets smaller, the slope of the regression line (b) gets closer to zero.

As the coefficient of determination gets smaller, the standard error gets larger, and closer to the standard deviation of the dependent variable (Y) ( $S_y = 2$ )

**Regression  
line in  
blue**

## Sample Statistics, Population Parameters and Statistical Significance tests

$$Y_i = a + bX_i + \varepsilon_i \quad a \text{ and } b \text{ are } sample \text{ } statistics$$

which are estimates of

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad population \text{ } parameters \alpha \text{ and } \beta$$

$\beta$  (and  $b$ ) measure the change in  $Y$  for a one unit change in  $X$ . If  $\beta = 0$  then  $X$  has no effect on  $Y$ , therefore

**Null Hypothesis ( $H_0$ ):** in the population  $\beta = 0$

**Alternative Hypothesis ( $H_1$ ):** in the population  $\beta \neq 0$

Thus, we test if our sample regression coefficient,  $b$ , is sufficiently different from zero to reject the Null Hypothesis and conclude that  $X$  has a statistically significant affect on  $Y$

# Test Statistics in Simple Regression

Test statistic for  $b$  is distributed according to the

*Student's t Distribution* (similar to normal):

$$t = \frac{b}{\text{SE}(b)} = \frac{b}{\sqrt{\frac{s_e^2}{\sum_i (X - \bar{X})^2}}}$$

where  $s_e^2$  is the variance of the estimate,  
with degrees of freedom =  $n - 2$

A test can also be conducted on the *coefficient of determination* ( $r^2$ ) to test if it is significantly greater than zero, using the *F* frequency distribution.

$$F = \frac{\text{Regression S.S./d.f.}}{\text{Residual S.S./d.f.}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / 1}{\sum (Y_i - \hat{Y}_i)^2 / n - 2}$$

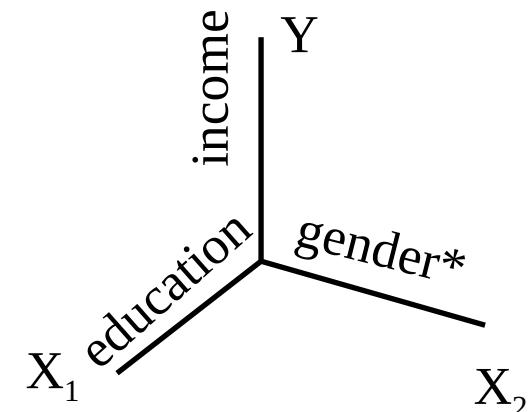
It is mathematically identical to the *t* test.

# Multiple regression

We can rewrite simple regression as:

$$Y = \alpha + \beta X + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$



Multiple regression: Y is predicted from 2 or more independent variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

β<sub>0</sub> is the *intercept* —the value of Y when values of all X<sub>j</sub> = 0

β<sub>1</sub>... β<sub>m</sub> are *partial regression coefficients* which give the change in Y for a one unit change in X<sub>j</sub>, all other X variables held constant

m is the number of independent variables

# Multiple regression: least squares criteria

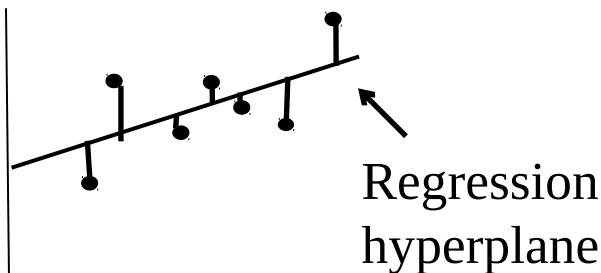
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

or  $Y_i = \sum_{j=0}^m X_{ij} \beta_j + \varepsilon_i$  (actual  $Y_i$ ).

$$\hat{Y}_i = \sum_{j=0}^m X_{ij} b_j$$
 predicted values for Y (regression hyperplane)

$$e_i = Y_i - \sum_{j=0}^m X_{ij} b_j = (Y_i - \hat{Y}_i) = (\text{Actual } Y_i - \text{Predicted } \hat{Y}_i) = \text{residuals}$$

$$\text{Min} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$



As in simple regression, the “least squares” criteria is used. Regression coefficients  $b_j$  are chosen to *minimize the sum of the squared residuals* (the deviations between actual  $Y_i$  and predicted  $\hat{Y}_i$ )

The difference is that  $\hat{Y}_i$  is predicted from 2 or more independent variables, not one.

# Coefficient of Multiple Determination ( $R^2$ )

- Similar to simple regression, the coefficient of multiple determination ( $R^2$ ) measures the proportion of the variance in Y (the dependent variable) which can be predicted or “explained by” all of X variables in combination.

Varies from 0 to 1.

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \leftarrow \begin{array}{l} \text{SS Regression or Explained Sum of Squares} \\ \text{SS Total or Total Sum of Squares} \end{array}$$

As with  
simple  
regression

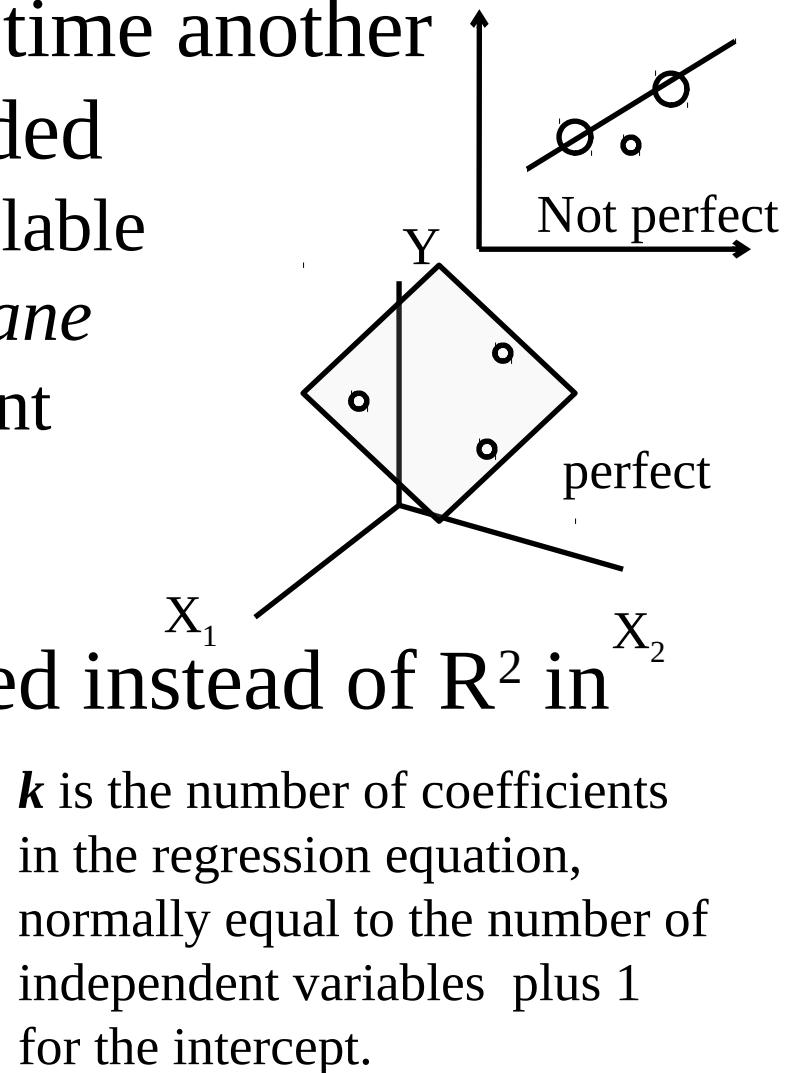
# Reduced or Adjusted

$$\overline{R}^2$$

- R<sup>2</sup> will always increase each time another independent variable is included
  - an additional dimension is available for fitting the regression *hyperplane* (the multiple regression equivalent of the regression line)

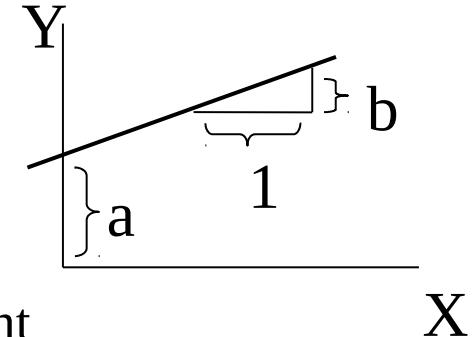
- Adjusted  $\overline{R}^2$  is normally used instead of R<sup>2</sup> in multiple regression

$$\overline{R}^2 = 1 - (1 - R^2) \left( \frac{n - 1}{n - k} \right)$$



# Interpreting *partial regression coefficients*

- The regression coefficients ( $b_j$ ) tell us the change in Y for a 1 unit change in  $X_j$ , all other X variables “held constant”
- Can we compare these  $b_j$  values to tell us the relative importance of the independent variables in affecting the dependent variable?
  - If  $b_1 = 2$  and  $b_2 = 4$ , is the affect of  $X_2$  twice as big as the affect of  $X_1$  ?
- **No, no, no in general!!!!**
- The size of  $b_j$  depends on the measurement scale used for each independent variable
  - if  $X_1$  is income, then a 1 unit change is \$1
  - but if  $X_2$  is rmb or Euro(€) or even cents (₵)  
1 unit is not the same!
  - And if  $X_2$  is % population urban, 1 unit is very different
- Regression coefficients are only directly comparable if the units are all the same: all \$ for example



# **STANDARDIZED PARTIAL REGRESSION COEFFICIENTS**

## ***COMPARING THE IMPORTANCE OF INDEPENDENT VARIABLES***

- How do we compare the relative importance of independent variables?
- We know we cannot use partial regression coefficients to directly compare independent variables unless they are all measured on the same scale
- However, we can use standardized partial regression coefficients (also called *beta weights*, *beta coefficients*, or *path coefficients*).
- They tell us the number of standard deviation (SD) unit changes in Y for a one SD change in X)
- They are the partial regression coefficients if we had measured every variable in *standardized form*

$$\beta_{XY_j} = b_j \left( \frac{s_{X_j}}{s_Y} \right) \longrightarrow z_i = \frac{(x_i - \bar{x})}{s_x}$$

Note the confusing use of  $\beta$  for both standardized partial regression coefficients and for the population parameter they estimate.

# REGRESSION:

## ***TESTING EACH INDEPENDENT VARIABLE***

A test can be conducted for each partial regression coefficient  $b_j$  to test if the associated independent variable influences the dependent variable. It is distributed according to the *Student's t Distribution* (similar to the normal frequency distribution):

Null Hypothesis  $H_0 : b_j = 0$

$$t = \frac{b_j}{\text{SE}(b_j)}$$

with degrees of freedom =  $n - k$ , where  $k$  is the number of coefficients in the regression equation, normally equal to the number of independent variables plus 1 for the intercept ( $m+1$ ).

The formula for calculating the standard error (SE) of  $b_j$  is more complex than for simple regression , so it is not shown here.

# REGRESSION

## *TESTING THE OVERALL MODEL*

- We test the *coefficient of multiple determination* ( $R^2$ ) to see if it is significantly greater than zero, using the  $F$  frequency distribution.
- It is an overall test to see if at least one independent variable, or two or more in combination, affect the dependent variable.
- Does not test if each and every independent variable has an effect

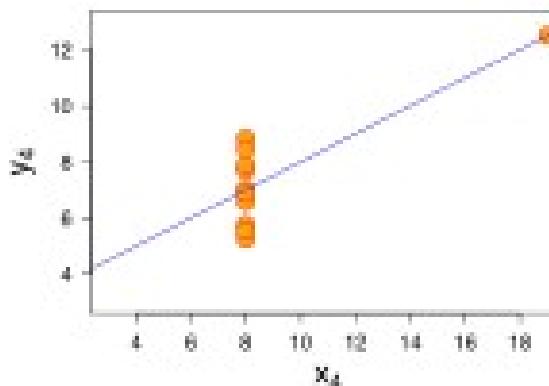
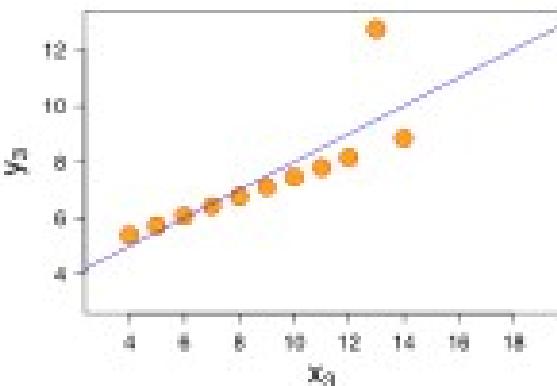
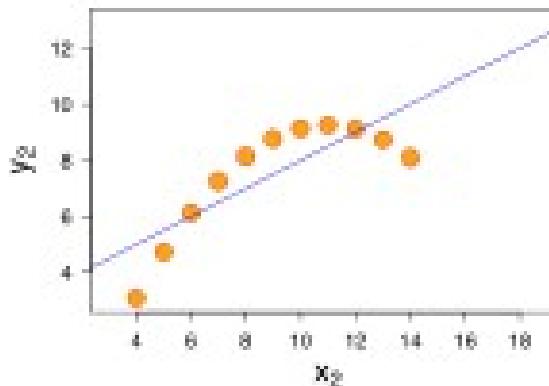
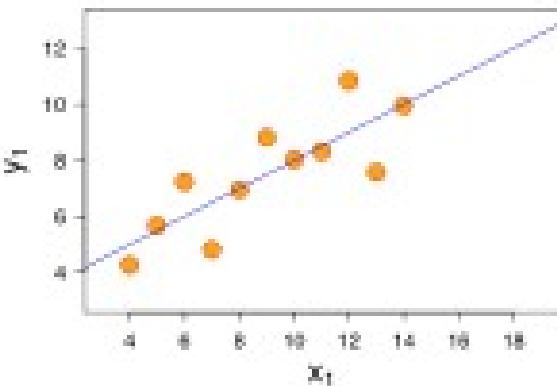
$$F = \frac{\text{Regression S.S./d.f.}}{\text{Residual S.S./d.f.}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / k - l}{\sum (Y_i - \hat{Y}_i)^2 / n - k}$$

Again,  $k$  is the number of coefficients in the regression equation, normally equal to the number of variables ( $m$ ) plus 1.

- Similar to the  $F$  test in simple regression.
  - But unlike simple regression, it is not identical to the  $t$  tests.
- It is possible (but unusual) for the  $F$  test to be significant but all  $t$  tests *not significant*.

# Always look at your data

*Don't just rely on the statistics!*



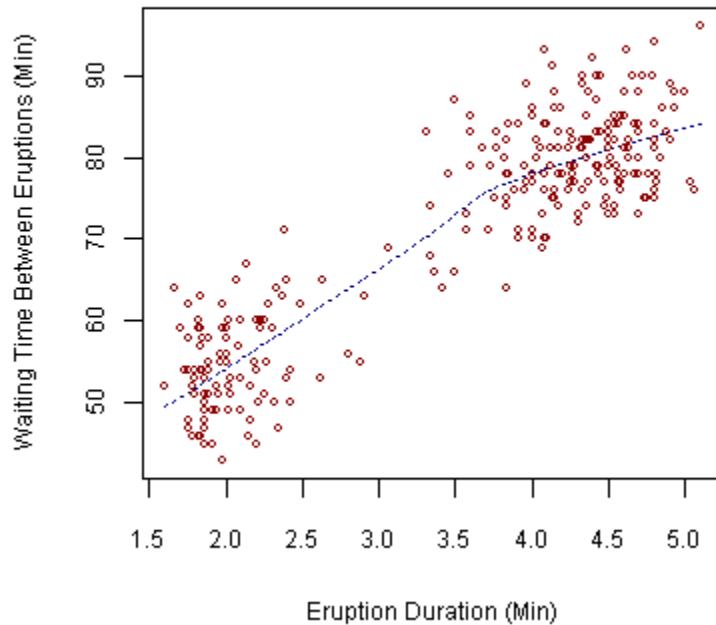
*Anscombe's quartet*

Summary statistics are the same for all four data sets:

mean (7.5),  
standard deviation (4.12),  
correlation (0.816)  
regression line  
 $(y = 3 + 0.5x)$ .

Anscombe, Francis J. (1973). "Graphs in statistical analysis". *The American Statistician* 27: 17–21.

## Old Faithful Eruptions



Real data is almost always more complex than the simple, straight line relationship assumed in regression.

Waiting time between eruptions and the duration of the eruption for the Old Faithful Geyser in Yellowstone National Park, Wyoming, USA. This chart suggests there are generally two "types" of eruptions: short-wait-short-duration, and long-wait-long-duration.

Source: Wikipedia

# Beware Spurious relationships

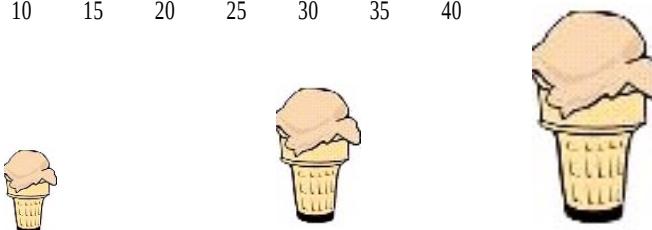
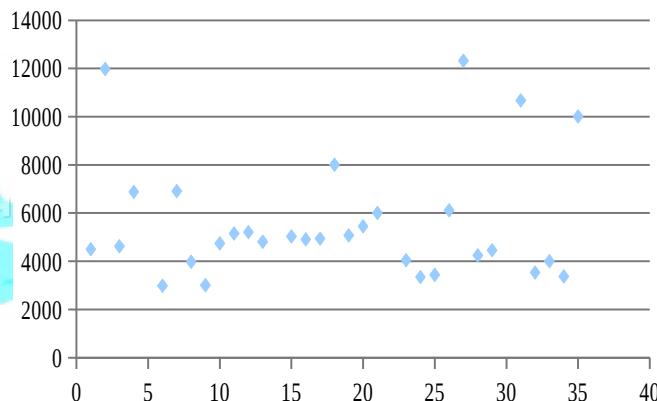
Eating ice cream inhibits swimming ability.

## Ice Cream sales related to Drownings

--eat too much, you cannot swim

*Omitted variable* problem

--both are related to a third variable not included in the analysis

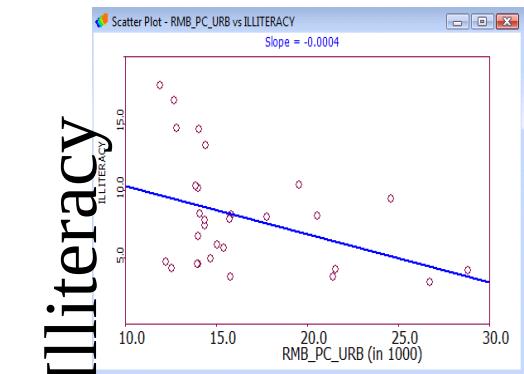


Summer temperatures:  
--more people swim (and some drown)  
--more ice cream is sold

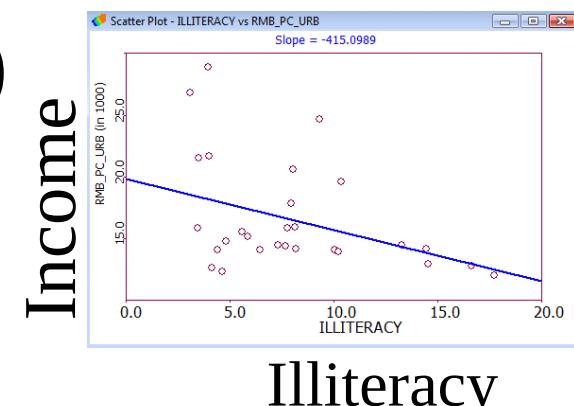
# Regression does not prove direction or cause!

## *Income and Illiteracy*

- Provinces with higher incomes can afford to spend more on education, so illiteracy is lower
  - Higher Income>>>Less Illiteracy
- The higher the level of literacy (and thus the lower the level of illiteracy) the more high income jobs.
  - Less Illiteracy>>>Higher Income
- Regression will not decide!



Illiteracy  
Income



Income  
RMB\_PC\_URB (in 1000)  
ILLITERACY  
Illiteracy

# WHAT IF THERE IS NOT ENOUGH DATA?

- Bootstrap
- Jackknife
- Cross-validation

# **WHAT IS BOOTSTRAP**

***Bootstrapping is a method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample.***

- The bootstrap procedure is a means of estimating the statistical accuracy . . . from the data in a single sample.
- Bootstrapping is used to mimic the process of selecting many samples when the population is too small to do otherwise
- The samples are generated from the data in the original sample by copying it many number of times (Monte Carlo Simulation)
- Samples can then selected at random and descriptive statistics calculated or regressions run for each sample
- The results generated from the bootstrap samples can be treated as if it they were the result of actual sampling from the original population

# *Characteristics of Bootstrapping*

| Sample Size                        |                       |
|------------------------------------|-----------------------|
| Sampling Method                    | Subsample             |
|                                    | Full Sample           |
| Sampling<br>with<br>Replacement    | Jackknife             |
| Sampling<br>Without<br>Replacement | Randomization<br>Test |
| Sampling<br>Without<br>Replacement | Bootstrap             |

# BOOTSTRAPPING EXAMPLE

## Original Data Set

|                |
|----------------|
| Pittsburgh     |
| Michigan       |
| Michigan State |
| Washington     |
| Purdue         |
| USC            |
| BYU            |
| Tennessee      |
| Navy           |
| Syracuse       |
| Stanford       |
| Ohio State     |

Limited number  
of observations

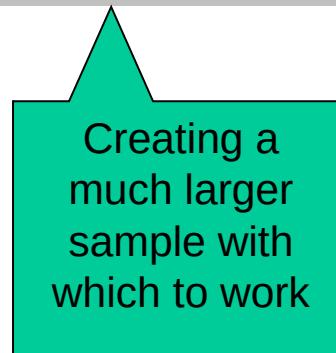
## 1<sup>st</sup> Random Sample

|            |
|------------|
| Navy       |
| Ohio State |
| USC        |
| Washington |
| Ohio State |
| USC        |
| BYU        |
| Stanford   |
| Pittsburgh |
| Ohio State |
| Stanford   |
| Michigan   |

$10^9$  Copies  
of each  
observation

Creating a  
much larger  
sample with  
which to work

Random sampling with replacement can be employed to create multiple independent samples for analysis



# **When it should be used**

***Bootstrapping is especially useful in situations when no analytic formula for the sampling distribution is available.***

- Traditional forecasting methods, like exponential smoothing, work well when demand is constant – patterns easily recognized by software
- In contrast, when demand is irregular, patterns may be difficult to recognize.
- Therefore, when faced with irregular demand, bootstrapping may be used to provide more accurate forecasts, making some important assumptions...

# **ASSUMPTIONS AND METHODOLOGY**

- Bootstrapping makes minimal assumption regarding the population (does need full independence)
- No normality of error terms
- No equal variance
- Allows for accurate forecasts of intermittent demand
- If the sample is a good approximation of the population, the sampling distribution may be estimated by generating a large number of new samples
- For small data sets, taking a small representative sample of the data and replicating it will yield good results. If not representative it will exaggerate bias

# ***Application and Uses***

## ***Setting Parameters***

- Agilent Technologies determined it was time to transfer manufacturing of its 3070 in-circuit test systems from Colorado to Singapore
- Major concern was the change in environmental test conditions (dry vs humid)
- Because Agilent tests to tighter factory limits (“guard banding”), they needed to adjust the guard band for Singapore
- Bootstrap was used to determine the appropriate guard band for Singapore facility

# VARIATIONS

- Bayesian Bootstrap
- Block Bootstrap
- Stratified Block Resampling bootstrap
- **Case resampling**
- **Bias-Corrected Bootstrap**
- **Gaussian process regression bootstrap**
- **Resampling residuals**

# *An Alternative to the bootstrap*

## **Jackknife**

- A statistical method for estimating and removing bias\* and for deriving robust estimates of standard errors and confidence intervals
- Created by systematically dropping out subsets of data one at a time and assessing the resulting variation

*Bias: A statistical sampling or testing error caused by systematically favoring some outcomes over others*



Vision And Security Technology  
University of Colorado Colorado Springs

# A COMPARISON OF THE BOOTSTRAP & JACKKNIFE

## Bootstrap

- Yields slightly different results when repeated on the same data (when estimating the standard error)
- Not bound to theoretical distributions

## Jackknife

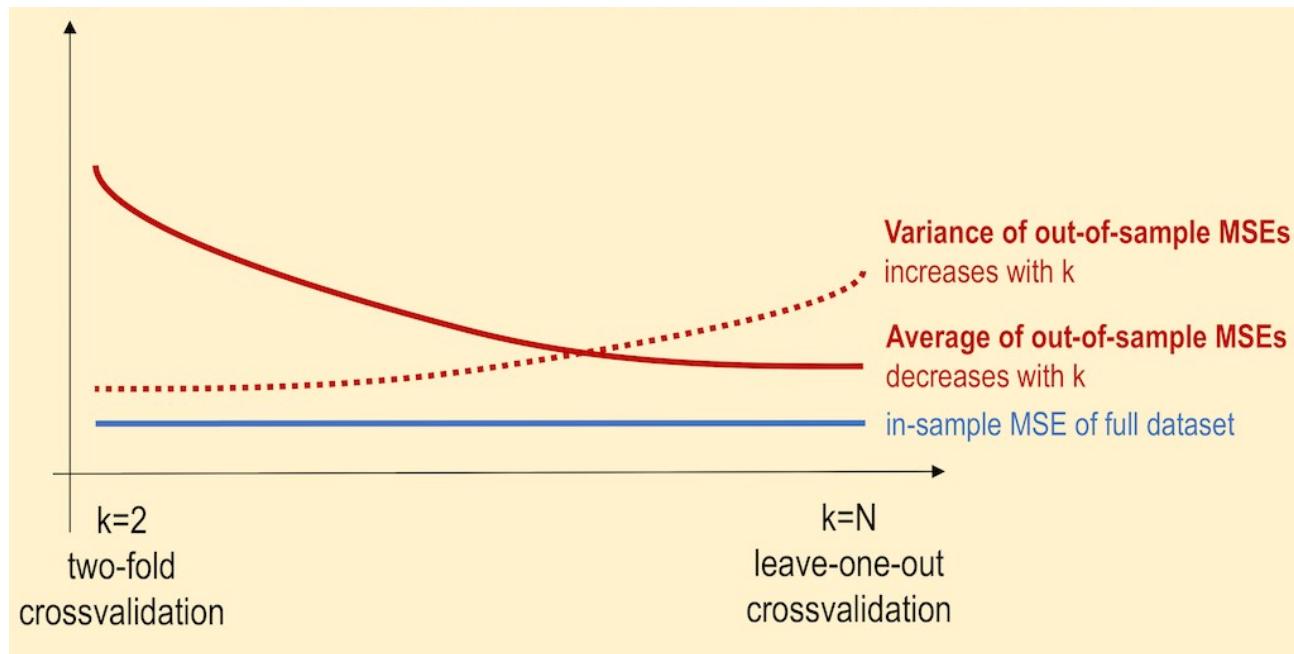
- Less general technique
- Explores sample variation differently
- Yields the same result each time
- Similar data requirements

# ***Another alternative method***

## ***CROSS-VALIDATION***

- The practice of partitioning data into a sample of data into sub-samples such that the initial analysis is conducted on a single sub-sample (training data), while further sub-samples (test or validation data) are retained “blind” in order for subsequent use in confirming and validating the initial analysis

| K-Fold Cross-Validation |              |                 |                 |                 |                 |                 |            |
|-------------------------|--------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------|
|                         | Learning Set | Learning/Test 1 | Learning/Test 2 | Learning/Test 3 | Learning/Test 4 | Learning/Test 5 | Evaluation |
| Learning Set            | Fold 1       | 20%             |                 |                 |                 |                 | 20%        |
| Test Set                | Fold 2       |                 | 20%             |                 |                 |                 | 20%        |
|                         | Fold 3       |                 |                 | 20%             |                 |                 | 20%        |
|                         | Fold 4       |                 |                 |                 | 20%             |                 | 20%        |
|                         | Fold 5       |                 |                 |                 |                 | 20%             | 20%        |
|                         |              |                 |                 |                 |                 |                 | 100%       |



## Cross-Validation

Validation set approach

1-Fold

k-Fold CV

k-Fold

LOOCV

n-Fold

**Bias**

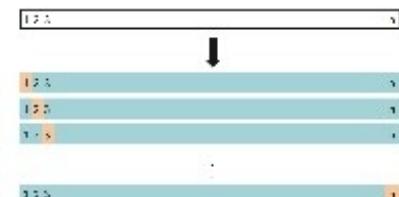
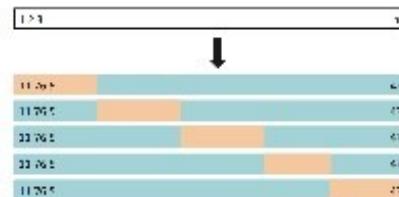
Bias

Bias

Variance

Variance

**Variance**



# **BOOTSTRAP VS. CROSS-VALIDATION**

## Bootstrap

- Requires a small amount of data
- More complex technique – **time consuming**

## Cross-Validation

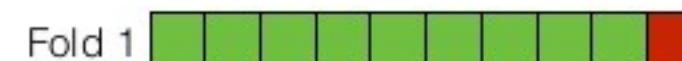
- Not a resampling technique
- Requires large amounts of data
- Extremely useful in data mining and artificial intelligence

# 10-fold cross-validation V.S. out-of-sample bootstrap validation

100 modules, 5% defective ratio



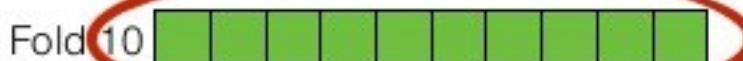
## 10-fold cross-validation



...

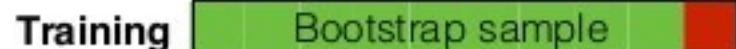


...



*There is a high chance that a testing corpus does not have any defective modules*

## Out-of-sample bootstrap



A sample with replacement with the same size of the original sample



Modules that do not appear in the bootstrap sample

*A bootstrap sample is nearly representative to the original dataset*

# ANOVA: ANALYSIS OF VALUE

## IS YOUR RESEARCH WORTH ANYTHING?

Developed in 1912 by geneticist R.A. Fisher, the Analysis of Value is a powerful statistical tool designed to test the significance of one's work.



am i  
wasting  
my time?

Significance is determined by comparing one's research with the **Dull Hypothesis**:

$$H_0 : \mu_1 = \mu_2 ?$$

where,

$H_0$  : the Dull Hypothesis

$\mu_1$  : significance of your research

$\mu_2$  : significance of a monkey typing randomly on a typewriter in a forest where no one hears it.

WWW.PHDCOMICS.COM

JORGE CHAM © 2007

The test involves computation of the  $F'd$  ratio:

$$F'd = \frac{\text{sum(people who care about your research)}}{\text{world population}}$$

This ratio is compared to the F distribution with  $I-1$ ,  $N_r$  degrees of freedom to determine a  $p(\text{in your pants})$  value. A low  $p(\text{in your pants})$  value means you're on to something good (though statistically improbable).

### Type I/II Errors

The Analysis of Value must be used carefully to avoid the following two types of errors:

Type I: You incorrectly believe your research is not Dull.

Type II: No conclusions can be made. Good luck graduating.

Of course, this test assumes both Independence and Normality on your part, neither of which is likely true, which means *it's not your problem*.

# **“HOW TO DEVELOP A SURVEY PAPER”**

Based in part on examples/slides  
from  
Dr. Manish Kumar Bajpai

# Your Primary paper assignment this term

- Solo: Survey paper at 8-10 pages in length at least 50 references.

Will turn in draft (4 weeks) + final paper (6 weeks)

- Everyone will be giving feedback on at least 3 other paper at some point. You will get feedback from others on each draft.
- Step 1) Get the lay of the land. For next week you are “search” for survey papers in your area and critically read least 3 survey papers, and build a topic map of where there are gaps.

# **WHAT MAKES A GOOD SURVEY?**

- Disucssion in class. Online Stop and think of 3 characteristics of a good survey and put that in your journal

# WRITING SURVEY PAPERS

- Selection of the topic for a good/impactful survey must satisfy the following requirements:
  - 1) The (sub) field is newly emerging.
  - 2) Popularity of the field will grow over time.
  - 3) A critical number of papers with new algorithms/approaches does exist (at least twenty to forty).
  - 4) The author(s) are enthusiastic about the particular topic.
  - 5) ~~A survey paper does not exist.~~
  - 6) Presents a new/useful organization of the space

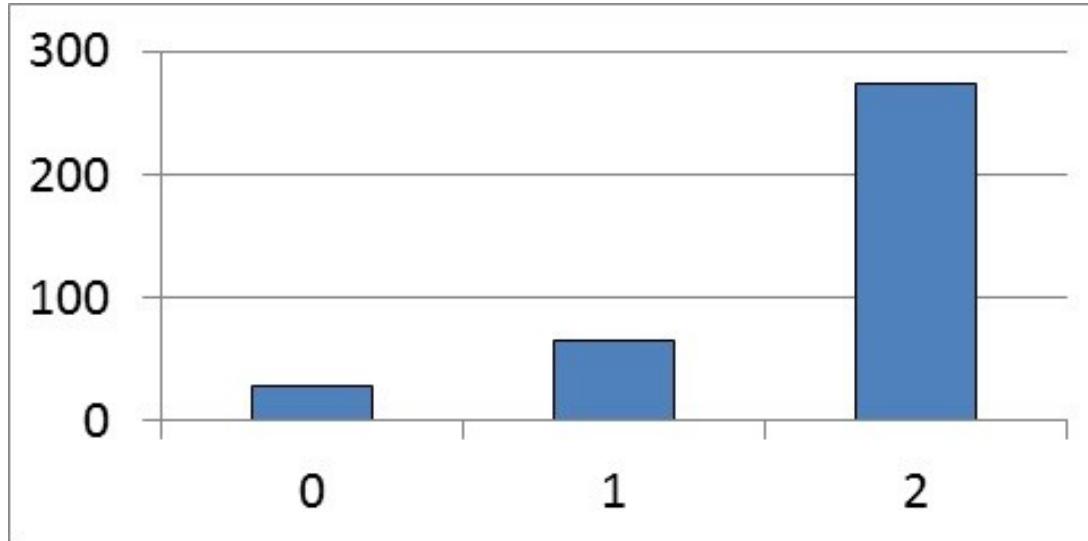


Figure : Impact of the existence of another survey paper.

Explanation: This figure gives a result which was absolutely unexpected.

The expectation was that existence of a survey would decrease citations of our survey, but it happened absolutely the opposite. This means that the quality is more important than the pre-existence of another survey paper on the same subject.

The paper with 2 preceded survey papers was the paper by Protic at al [Protic 1996].

The paper with one preceded survey was the paper by Tomasevic at al [Tomasevic1993].

The paper with no preceded survey was the paper by Jovanovic at al [Jovanovic1999].

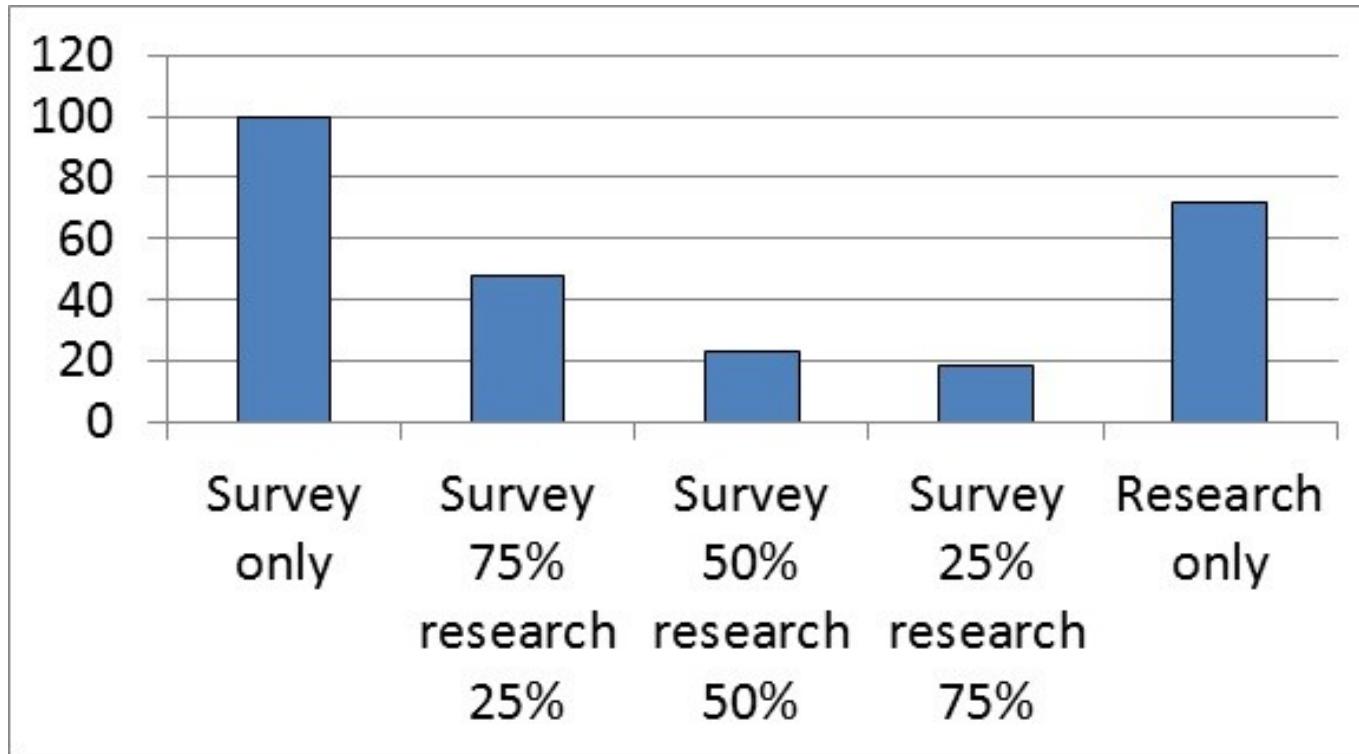


Figure : Survey papers versus research papers, what generates more citations?  
Explanation: Surveys generate more, unless an extraordinary research paper is generated in a popular field.

# SURVEY PAPERS

- With the binary (or n-ary) criteria, one can create either a tree-like classification or a cube-like classification,

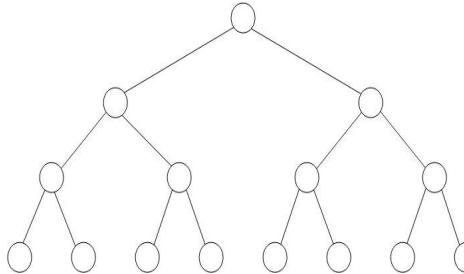
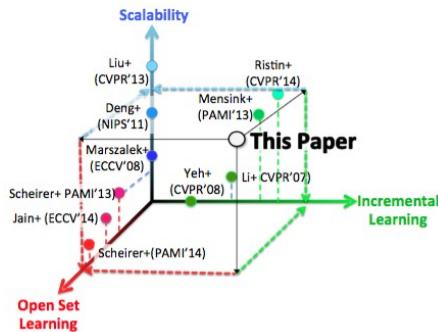


FIGURE. A tree-like classification.  
Classes are only at the leaves of the tree.

- With a **tree-like classification**, one classifies only the approaches that entirely belong to a specific class.
- With a **cube-like classification**, one defines a space in which inner points include, to some extent, characteristics of all existing classes
- What is useful, is to prepare a figure which includes the following:
  - 1) The classification criteria.
  - 2) The classification.
  - 3) Technical mnemonics.
  - 4) Symbolic mnemonics.
  - 5) The number of selected examples per class.
  - 6) The full list of references of selected examples.
  - 7) A vector of relevant characteristics.

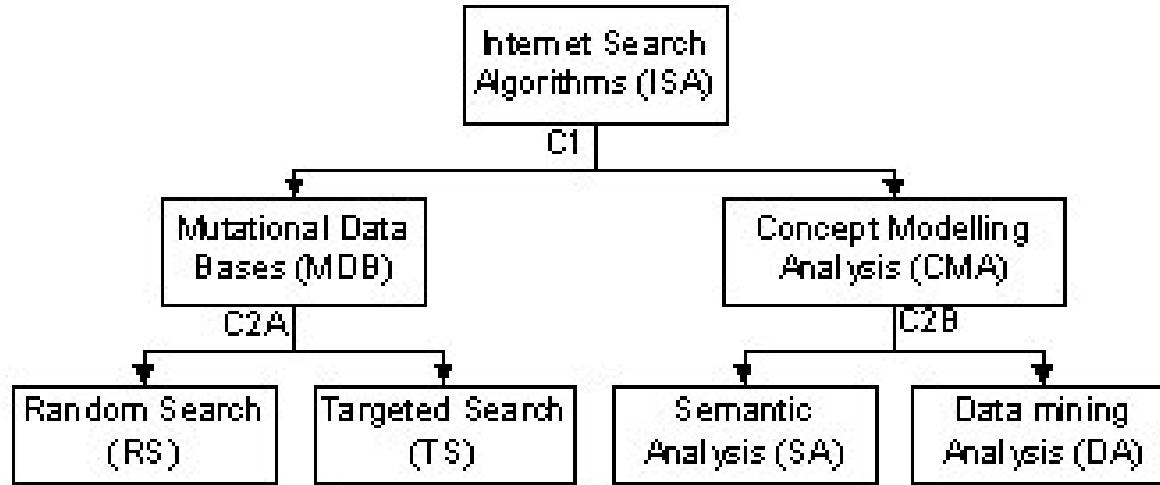


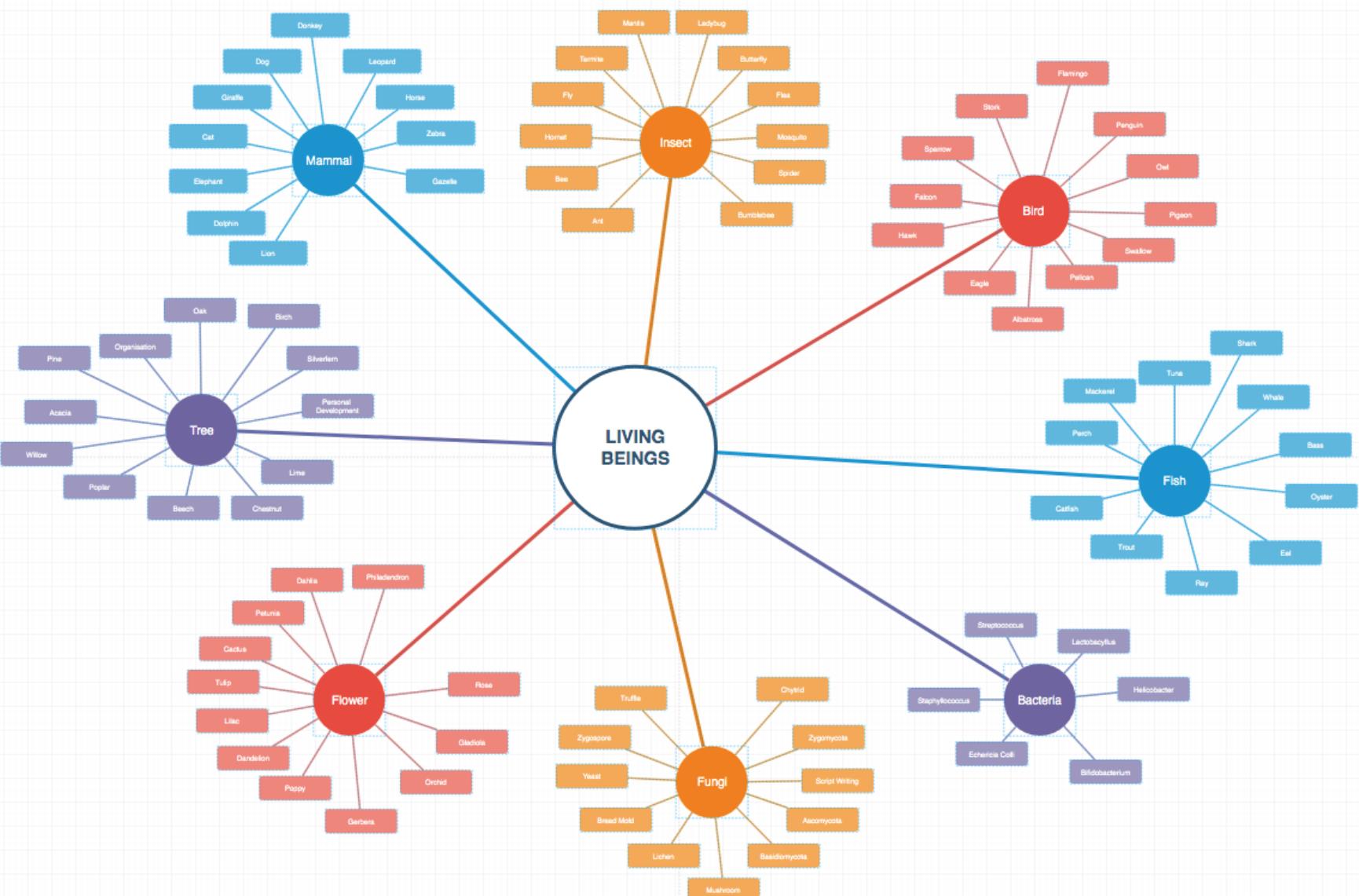
Figure . Classification of Internet Search Algorithms

Legend:

- C1 (criterion #1) = Retrieval-oriented vs Analysis-oriented
- C2A (criterion #2, in the MDB path) = Random Search vs Targeted Search
- C2B (criterion #2, in the CMA path) = Semantics-oriented vs Datamining-oriented

# TRY MIND MAPPING SOFTWARE

- [Coggle](#) (Web) for beginners
- [Mindly](#) (macOS, iOS, Android) for mobile mind mapping
- [Draw.io](#) (Web, macOS, Linux, Windows, Chrome OS) for free mind mapping
- [iMindMap](#) (macOS, Windows, iOS) for performing in-depth analysis on your mind map
- MindMup (Web) for creating public mind maps
- MindMeister (Web, iOS, Android, macOS, Windows) for collaborating on a mind map with a team
- Scapple (macOS, Windows) for fluid, non-traditional mind mapping
- SmartDraw (Web) for linear mind mapping
- Stormboard (Web) for in-person mind mapping sessions
- LucidChart (Web, iOS, Android) for turning your mind map into an organized flowchart



# **SURVEY PAPERS**

- When presenting each particular example, one can start with these 7 items:
  - 1) Seven Ws about the survey example (Who, What, When, Where, Why, for Whom, hoW).
  - 2) Essence (it can extremely difficult to give entire essence in only one sentence).
  - 3) Structure
  - 4) Some relevant details.
  - 5) Example (here one can call a figure that explains an example using a pseudo-code; ideally, the same application case should be used for all surveyed examples).
  - 6) Pros and cons.
  - 7) Author's opinion of this example and its potentials.
- For short surveys, each template element is a sentence.  
For long surveys, each template element is a paragraph.  
For books, each template element can be a page, or more.

# **GOOD SURVEY PAPERS**

- The need to weave that into a story.
- Good surveys are not just a summary.
  - Must to compare and contrast the papers.
  - Must tell a story, with a theme and goal.
  - Should help organize ideas for others
  - Need not be complete but should be thorough within its theme/story.

# **SURVEY PAPER**

- 1) Read about the general subject, to warm up. Pick target journal and locate related surveys
- 2) Determine your proposed classification approach and select 40-80 papers, on various approaches from the open literature based on your classification ideas. Explain why the proposed classification represents a contribution to science.
3. For each example (covered by one or more papers),  
write the main 7 ideas mentioned earlier in this paper.  
Use these place within the chosen template/classification which could enables easy comparison, and therefore represents a contribution to science.

Pick the “key” points of comparison among the papers, may try many more than the final paper..

5. For each key point of comparison, generate two figures (for example, one block scheme of the structure and one pseudo code presentation of the algorithm). Choose the presentation form which indicates the essence of the class that the example belongs to.
6. If the generated classification includes a class without examples (which is highly desirable, since that points to possible new research avenues), define the research strategy of interest for those who decide to take that avenue. Form a section with appropriate discussions.

7. Define the research strategy for those who decide to analyze the hybrid approaches (those consisting of elements of two different classes). Hybrid approaches can be either a symbiosis (the two solutions used interchangeably, as the conditions dictate), or a synergy (the two solutions combined into one). Discuss possible new solutions or both types (symbiosis and synergy). Discuss other possible avenues leading to new inventions

8. Add the preamble and the conclusion and form the final text of the paper. Generate ONE pearl of wisdom that sheds light on the essence of the paper, and increases the probability that the paper be referenced a lot.

9. Ask peers to review your paper, and edit edit edit .

10. Submit the paper to a journal.



# WRITING

CS6000

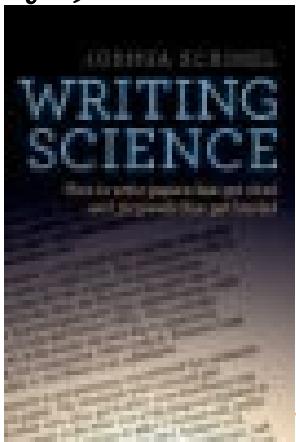
Prof. Boult



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs  
Colorado Springs

# **Suggested reading/books on writing**

*Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded*  
by Joshua Schimel, ISBN 978-0199760244



Writing Science: How to Write Papers That Get Cited and Proposals That  
Get Funded

Joshua Schimel

Watch his video: <https://vimeo.com/17432903>

# The writing process

- How to get started
- A good process
- why process != writing

## **GITOMERS 5.5 THINGS TO IMPROVE YOUR WRITING:**

- 1. Just sit down and write something**
- 2. Capture your though and ideas the second they occur.**
- 3. Write it like you would say it (at least at first).**
- 4. Edit to make sure your thoughts are simple and easy to understand ad complete.**
- 5. Edit early, then often**  
**5.5 Your writing for the reader AND yourself**

# Effective Writing: Process

- All that you need to do is three things:
  - Prewrite (about 25 percent of your time)
  - Write (about 25 percent of your time)
  - Rewrite (about 50 percent of your time)
- Have a good process, a timelines and tools
  - GIT and Latex prefered
  - Word (track changes, dropbox.. )



- The biggest mistake you can make as a writer is waiting until its “ready”, let alone waiting for perfect. The rewriting process begins when you show your first draft to people and start getting feedback.  
Beethoven went through 70 drafts of his symphonies. As William Zinsser put it:
- “Rewriting is where the game is won or lost; rewriting is the essence of writing.”



“Rewriting is the essence of writing well: it’s where the game is won or lost. The idea is hard to accept. We all have an emotional equity in our first draft; we can’t believe that it wasn’t born perfect. But the odds are close to 100% that it wasn’t. Most writers don’t initially say what they want to say, or say it as well as they could. The newly hatched sentence almost always has something wrong with it. It’s not clear. It’s not logical. It’s verbose. It’s klunky. It’s pretentious. It’s boring. It’s full of clutter. It’s full of clichés. It lacks rhythm. It can be read in several different ways. It doesn’t lead out of the previous sentence. It doesn’t... The point is that clear writing is the result of a lot of tinkering.

# SHI..Y FIRST DRAFTS

- or why I don't recommend starting with an outline..
- All first drafts are “bad”.. so might as well get one done so you can start making things better.
- You should do your first draft, getting lots of stuff down on paper. Then you can review it to make an outline.
- (Read  
<http://richardcolby.net/writ2000/wp-content/uploads/2017/09/Bird-by-Bird-Anne-Lamott.pdf>  
chapter on shitty drafts (I'll

## In Class EXERCISE

- Don't let words get in the way..
- Seeing words slows down thinking and often makes one want to edit. You are to close your eyes and write/type, in your native language, about your own research idea. You are free to jump between ideas but must spend at least 30seconds on each idea.
- I'll tell you when to stop.
- (If at home, do it while the video plays)

## A BETTER WAY for many

- Don't write.. talk.. use a voice recorder or use your laptop to record your first draft.
- Get an app that does “dictation” and talk to it.. It gets you even farther as much of the text is there. But DO NOT look at it as it types for you.
- Ideal dictation software saves voices + text, but raw text is often good enough to prompt your memory.

- Now look back and look for the storyline in it. You want a simple sticky story (we'll define that latter)..
- What is both novel and useful/interesting
- Look for possible flows
- What surprises can you find?

***If they cannot remember your work,  
they won't accept/cite your work!***

## **LISTEN TO YOUR INNER CHILD**

- At least for the first draft.. they are more playful, creative and full of ideas. Exactly what you want in the first draft.
- See if you can think in pictures, shapes, characters.

# **FIRST EXERCISE, REDUX**

- Try it again but go with the inner child's voice and tell a story..
- I'll tell you when to stop.

## **SECOND EXERCISE**

- Lose the words.
  - The ideas from what you wrote, and sketch them in very crude figures.. see the figures, connect the concepts as a cartoon or comic book might. You can change the order to make it a more meaningful story.



# WRITING

CS6000

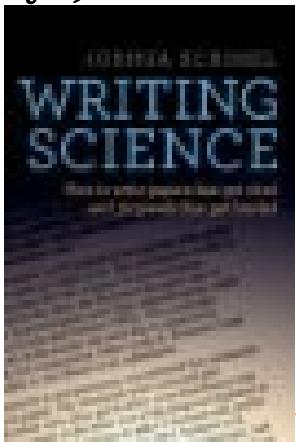
Prof. Boult



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs  
Colorado Springs

# **Suggested reading/books on writing**

*Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded*  
by Joshua Schimel, ISBN 978-0199760244



Writing Science: How to Write Papers That Get Cited and Proposals That  
Get Funded

Joshua Schimel

Watch his video: <https://vimeo.com/17432903>

# The writing process

- How to get started
- A good process
- why process != writing

## **GITOMERS 5.5 THINGS TO IMPROVE YOUR WRITING:**

- 1. Just sit down and write something**
- 2. Capture your though and ideas the second they occur.**
- 3. Write it like you would say it (at least at first).**
- 4. Edit to make sure your thoughts are simple and easy to understand ad complete.**
- 5. Edit early, then often**
  - 5.5 Your writing for the reader AND yourself**

# **Effective Writing: Process**

- **All that you need to do is three things:**
  - Prewrite (about 25 percent of your time)
  - Write (about 25 percent of your time)
  - Rewrite (about 50 percent of your time)
- Have a good process, a timelines and tools
  - GIT and Latex prefered
  - Word (track changes, dropbox.. )



- The biggest mistake you can make as a writer is waiting until its “ready”, let alone waiting for perfect. The rewriting process begins when you show your first draft to people and start getting feedback.  
Beethoven went through 70 drafts of his symphonies. As William Zinsser put it:
- “Rewriting is where the game is won or lost; rewriting is the essence of writing.”



“Rewriting is the essence of writing well: it’s where the game is won or lost. The idea is hard to accept. We all have an emotional equity in our first draft; we can’t believe that it wasn’t born perfect. But the odds are close to 100% that it wasn’t. Most writers don’t initially say what they want to say, or say it as well as they could. The newly hatched sentence almost always has something wrong with it. It’s not clear. It’s not logical. It’s verbose. It’s klunky. It’s pretentious. It’s boring. It’s full of clutter. It’s full of clichés. It lacks rhythm. It can be read in several different ways. It doesn’t lead out of the previous sentence. It doesn’t... The point is that clear writing is the result of a lot of tinkering.

# SHI..Y FIRST DRAFTS

- or why I don't recommend starting with an outline..
- All first drafts are “bad”.. so might as well get one done so you can start making things better.
- You should do your first draft, getting lots of stuff down on paper. Then you can review it to make an outline.
- (Read  
<http://richardcolby.net/writ2000/wp-content/uploads/2017/09/Bird-by-Bird-Anne-Lamott.pdf>  
chapter on shitty drafts (I'll

## In Class EXERCISE

- Don't let words get in the way..
- Seeing words slows down thinking and often makes one want to edit. You are to close your eyes and write/type, in your native language, about your own research idea. You are free to jump between ideas but must spend at least 30seconds on each idea.
- I'll tell you when to stop.
- (If at home, do it while the video plays)

## A BETTER WAY for many

- Don't write.. talk.. use a voice recorder or use your laptop to record your first draft.
- Get an app that does “dictation” and talk to it.. It gets you even farther as much of the text is there. But DO NOT look at it as it types for you.
- Ideal dictation software saves voices + text, but raw text is often good enough to prompt your memory.

- Now look back and look for the storyline in it. You want a simple sticky story (we'll define that latter)..
- What is both novel and useful/interesting
- Look for possible flows
- What surprises can you find?

***If they cannot remember your work,  
they won't accept/cite your work!***

## **LISTEN TO YOUR INNER CHILD**

- At least for the first draft.. they are more playful, creative and full of ideas. Exactly what you want in the first draft.
- See if you can think in pictures, shapes, characters.

# **FIRST EXERCISE, REDUX**

- Try it again but go with the inner child's voice and tell a story..
- I'll tell you when to stop.

## **SECOND EXERCISE**

- Lose the words.
  - The ideas from what you wrote, and sketch them in very crude figures.. see the figures, connect the concepts as a cartoon or comic book might. You can change the order to make it a more meaningful story.



# **WHY ARE YOU WRITING?**

- The art of persuasion and getting your way.
  - If you are not calling the reader to action, you are not persuading them. What are your actions in a paper? Proposal?

What does it mean to persuade?

# Organization is more important than grammar!

- Grammatical errors are not good, but for the reader they're not as bad as these problems:
  - *illogical sequence of ideas* – why am I reading about all these details when I don't yet know the big picture?
  - *no pattern of organization* – is this section moving from most important to least important ideas, or not?
  - *no flow of ideas* – I feel as though I'm starting over again with each new sentence.
  - *Jumpiness* – topic seem to jump around

# OSCAR

Integral elements/order in all good  
(science) stories

- **Opening** (broad hook)
- **Setting the (Science) Stage**
- **Challenge** (the problem)
- **Action** (Alg & Experiments)
- **Resolution** (Must tie back  
to opening)



# **more classical science format is IMRAD Format**

- **I** = **Introduction**, what question (problem) was studied
- **M** = **Methods**, how was the problem studied
- **R** = **Results**, what are the findings
- **A** = and (or Analysis)
- **D** = **Discussion**, what do these findings mean

# **Organization of a scientific paper**

- The most common structure in the rest of science is the IMRAD
- If a number of methods were used to achieve directly related results or for comparison:

M + R = Experimental section

- The results are so complex that they need to be immediately discussed:

R + D = Results and Discussion section

# **OSCAR vs IMRAD**

- **Opening**
- **Setting**
- **Challenge**
- **Action/  
Analysis/  
Approach**
- **Resolution**

I = **Introduction**, what question (problem) was studied

M = **Methods**, how was the problem studied

R = **Results**, what are the findings

A = and (Or analysis)

D = **Discussion**, what do these findings mean

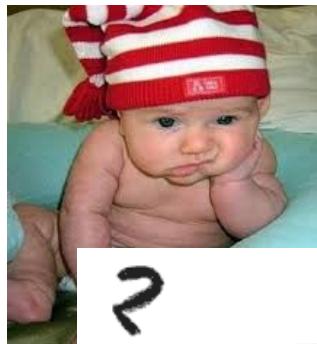
# **Writing as PERSUASIVE story telling**

- Why do you think its PERSUASIVE?
- What makes things PERSUASIVE
- What makes a good story

# FIVE KEY ELEMENTS OF PERSUASIVE PRESENTATION

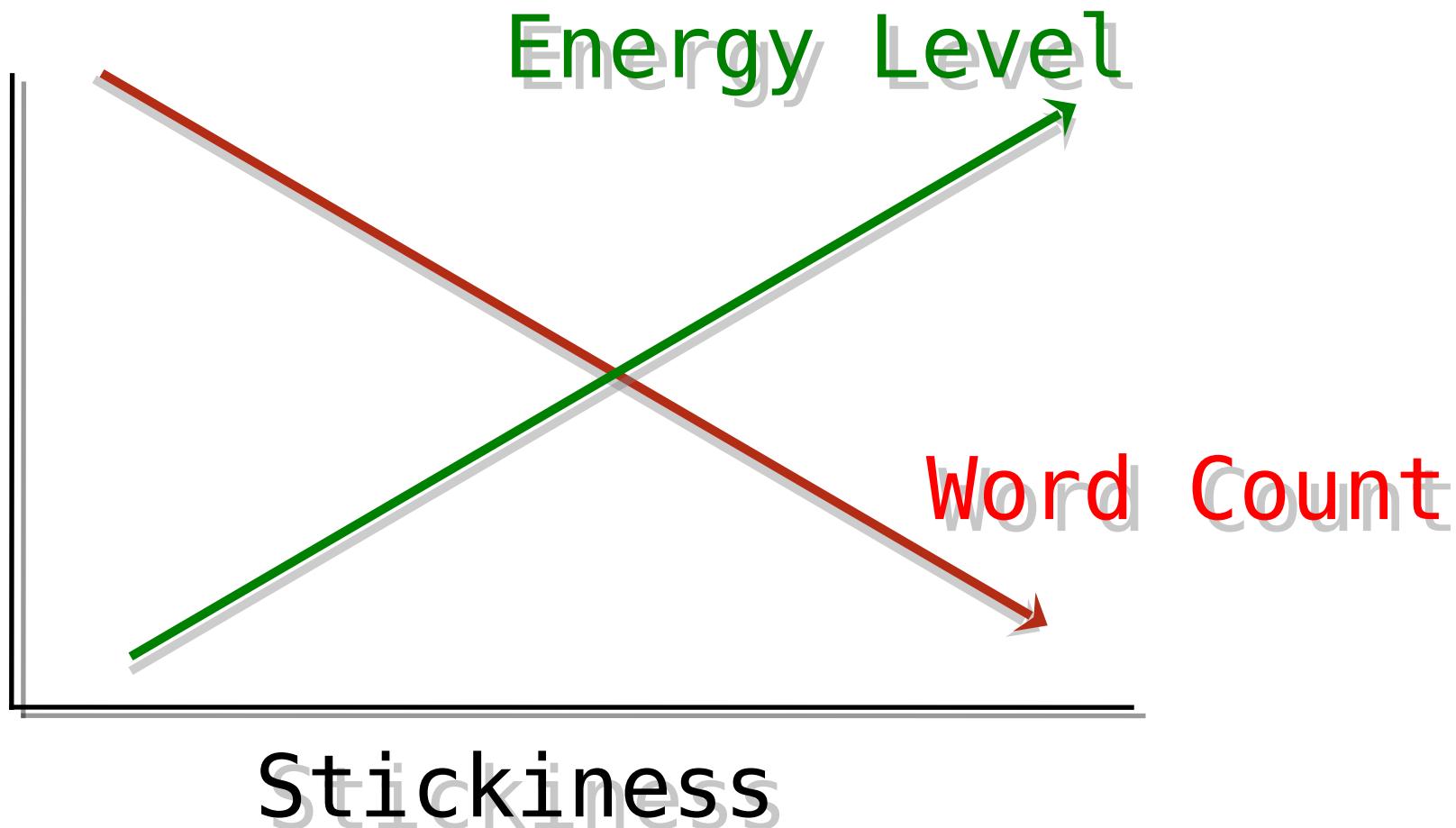
## I 1) Action oriented style with energy!

- Think about everything you say in terms of its impact, not its information.



An investor enjoys yet another set of detailed bullet points

# ANATOMY OF A STICKY STORY



## #2 WHAT'S IN IT FOR THEM



## **#2, WIIFT(WHATS IN IT FOR THEM)**

- #2 Benefits, then features
  - If, as a reader, I don't see a benefit, why would I care about the details. Show the reader you are thinking about them, not you.
  - Think about everything you write in terms of its impact, not its information
  - Features need to be explained to make the benefits creditable.
  - Then work in the benefits again, but as a call to action!

## #3 Structure & Scannable Elements

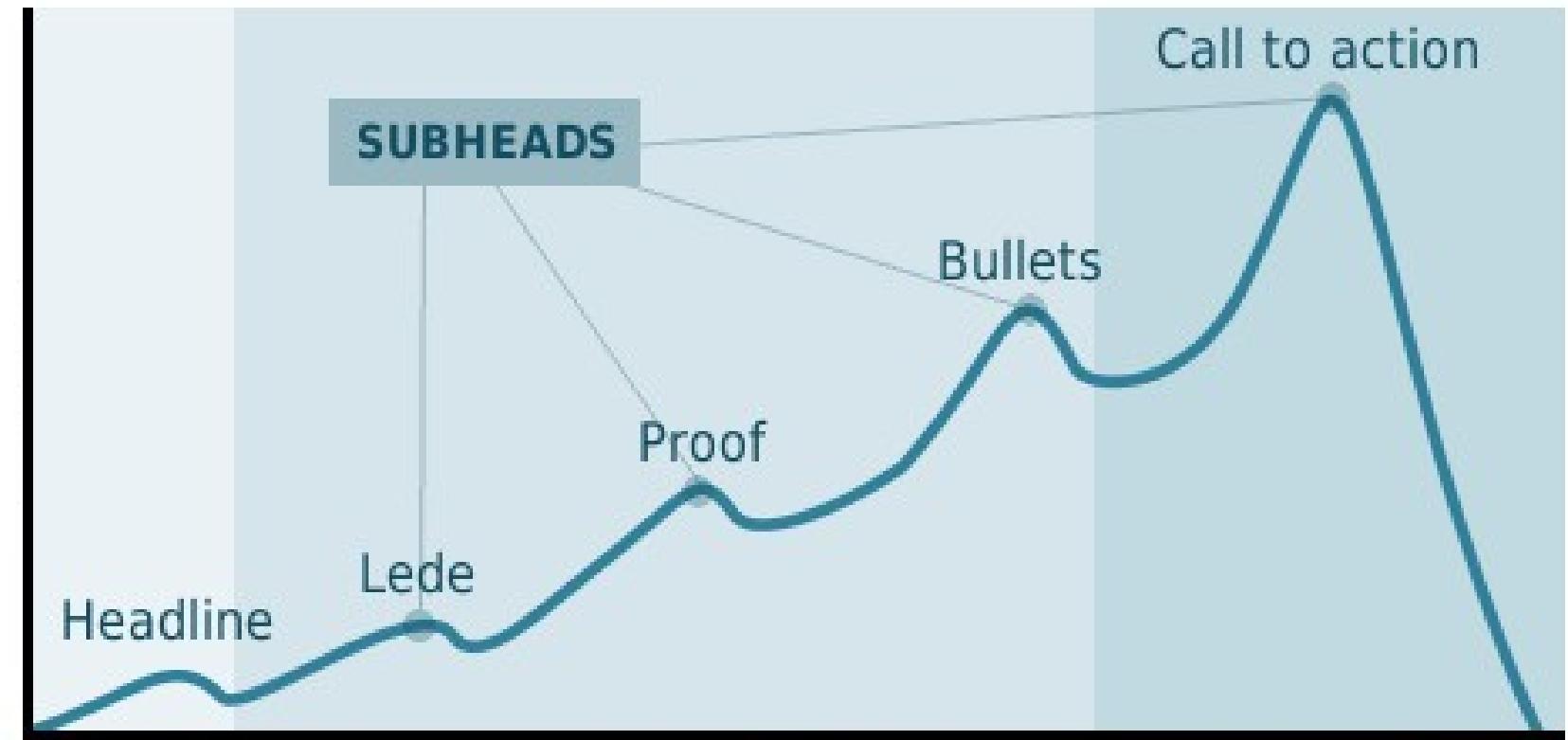
79% of people don't read—they scan.

Must hook them with Heading, Bullets,  
Bold/italics.

Expect things to be in place, section order,  
&key first sentences of paragraphs.

Images that convey value more forcibly than  
text could, such as charts, graphs, or high-  
quality product photos

Captions—these get read by 50% more  
"readers" than body text, and often have a  
nearly 100% recall rate



Opening &  
setup

Body &  
exposition

Climax &  
dénouement

# #4 BUILDING EMOTION AND MEMORY



well, you know what  
they say... a JPEG  
is worth 1,024 words

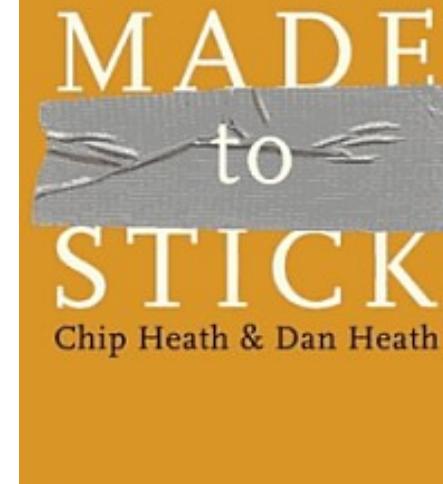


# #5 STICKY STORIES

- Think about everything you say in terms of its impact on your “story” and its stickiness



Why Some Ideas Survive  
and Others Die



# **SUCCESS: STICKY STORIES**

**S**imple

**U**nexpected

**C**oncrete

**C**redible

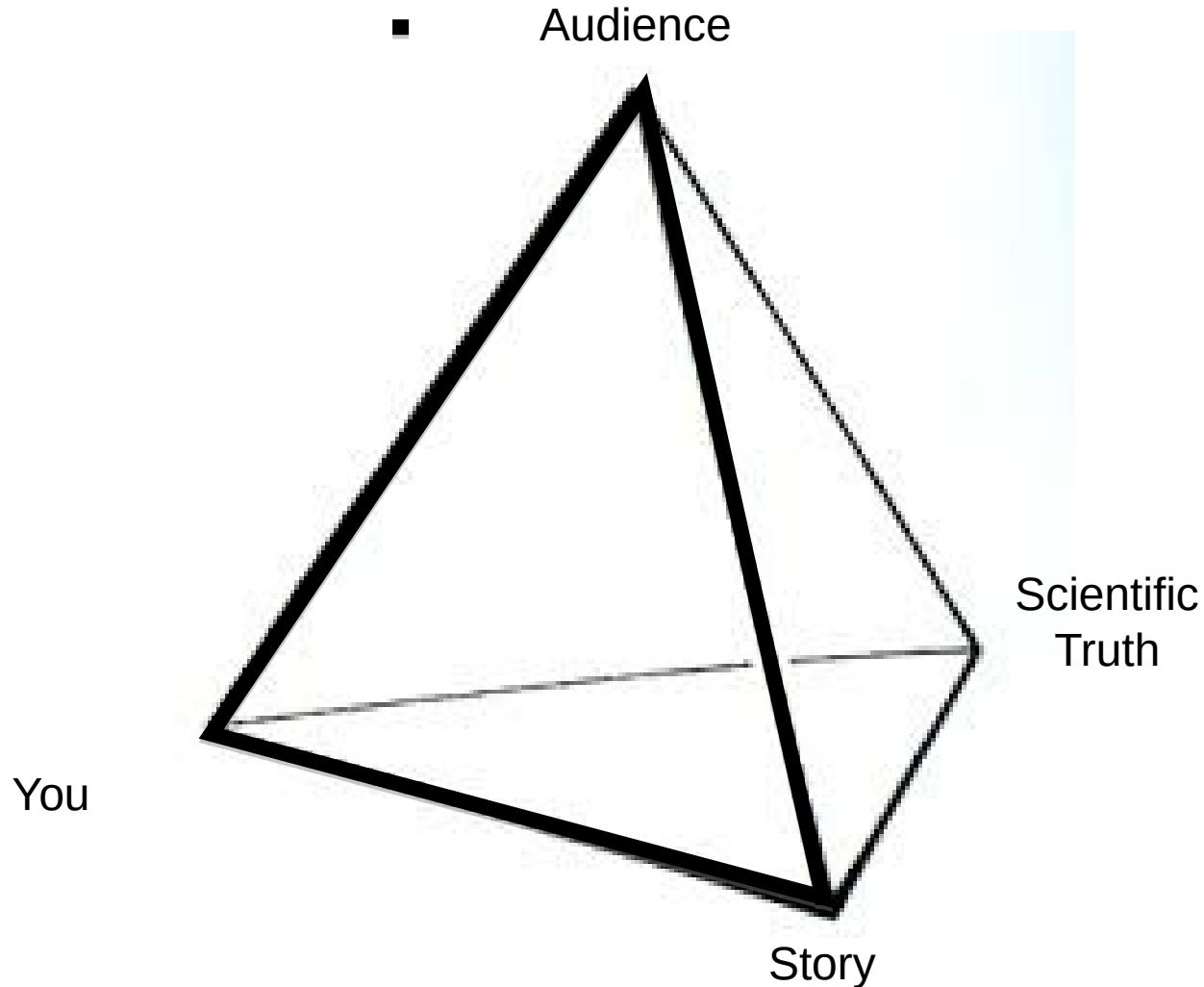
**E**motional

**S**tories

**S**tick!

**PLAN FOR  
SUCCESS**

# Writing as Story Telling



**THERE ARE TWO HALVES TO THE  
WORD**

**“STORYTELLING:”**

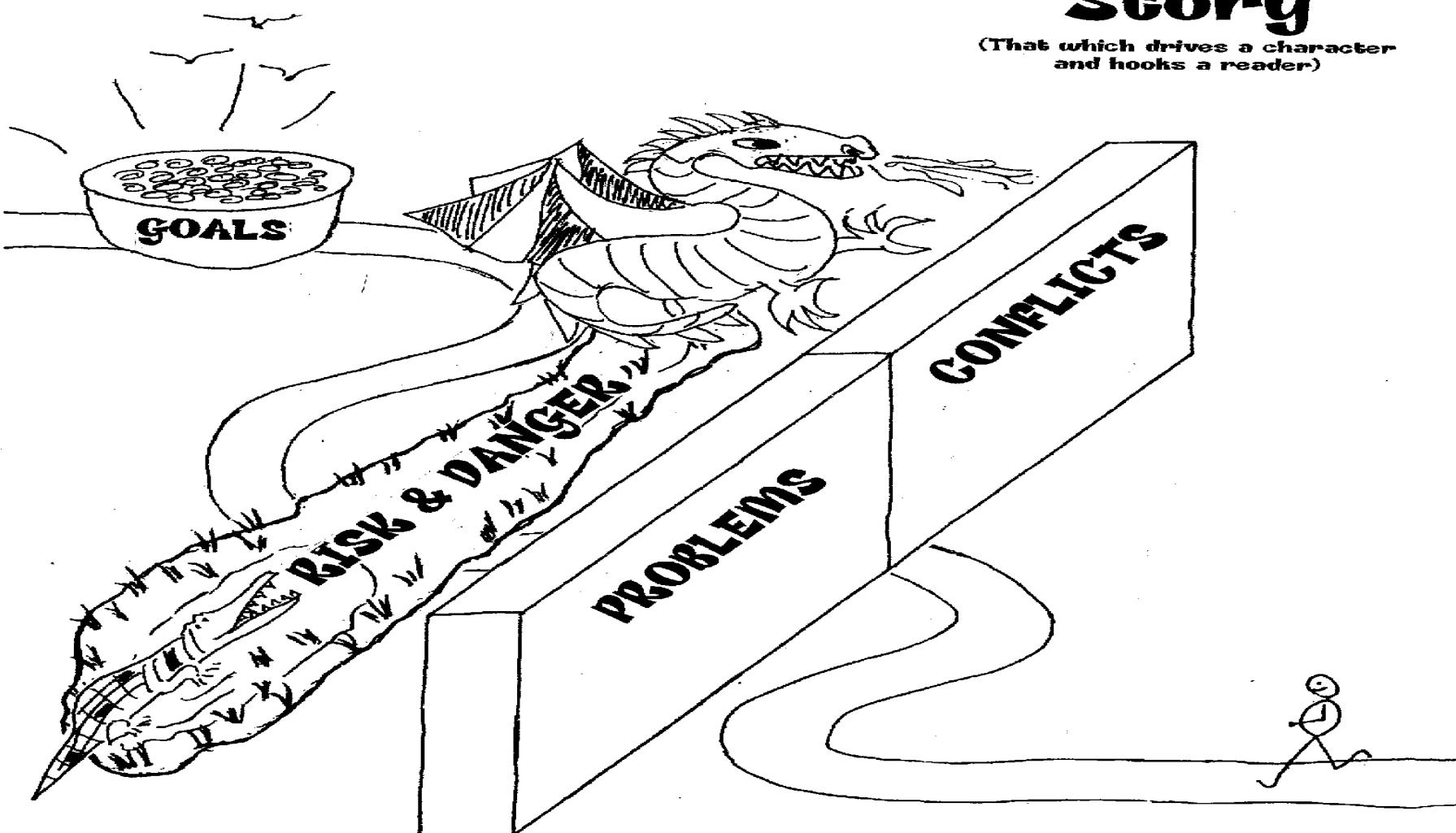
***STORY***

**—AND—**

***TELLING***

# The Core of a Story

(That which drives a character  
and hooks a reader)



# **GETTING WHAT YOU NEED**

You can't always get what you want -

But if you try sometime, yeah,  
You just might find you get what you need!

-Rolling Stones

# **EIGHT ESSENTIAL ELEMENTS**

- 1. CHARACTER**
- 2. CHARACTER TRAITS** (THAT  
MAKE CHARACTERS INTERESTING)
- 3. GOAL**
- 4. MOTIVE**
- 5. CONFLICTS & PROBLEMS**
- 6. RISK & DANGER**
- 7. STRUGGLES**
- 8. DETAILS**

# **THE EIGHT ESSENTIAL ELEMENTS OF EVERY STORY/NARRATIVE**

- 1. WHO IS THE MAIN CHARACTER?**
- 2. WHAT CHARACTER TRAITS MAKE THEM INTERESTING?**
- 3. WHAT DO THE CHARACTER NEED TO DO OR GET (GOAL)?**
- 4. WHY IS THAT GOAL IMPORTANT (MOTIVE)?**
- 5. WHAT CONFLICTS/PROBLEMS BLOCK THE CHARACTER?**
- 6. HOW DO THEY CREATE RISK & DANGER?**
- 7. WHAT DOES THE CHARACTER DO (STRUGGLES) TO REACH GOAL?**
- 8. WHAT SENSORY DETAILS WILL MAKE THE STORY SEEM REAL?**

# Science/Tech story

## 1. WHO IS THE MAIN CHARACTER?

Algorithm/system/theory of the paper

## 2. WHAT CHARACTER TRAITS MAKE THEM INTERESTING?

What helps reader remember WHY it works/is better

## 3. WHAT DO THE CHARACTER NEED TO DO OR GET (GOAL)?

What is measure by which you beat state of the art

## 4. WHY IS THAT GOAL IMPORTANT (MOTIVE)?

Why is that property important..

## 5. WHAT CONFLICTS/PROBLEMS BLOCK THE CHARACTER?

What make it hard to do..

## 6. HOW DO THEY CREATE RISK & DANGER?

What are the risks/danger if not successful.

## 7. WHAT DOES THE CHARACTER DO (STRUGGLES) TO REACH GOAL?

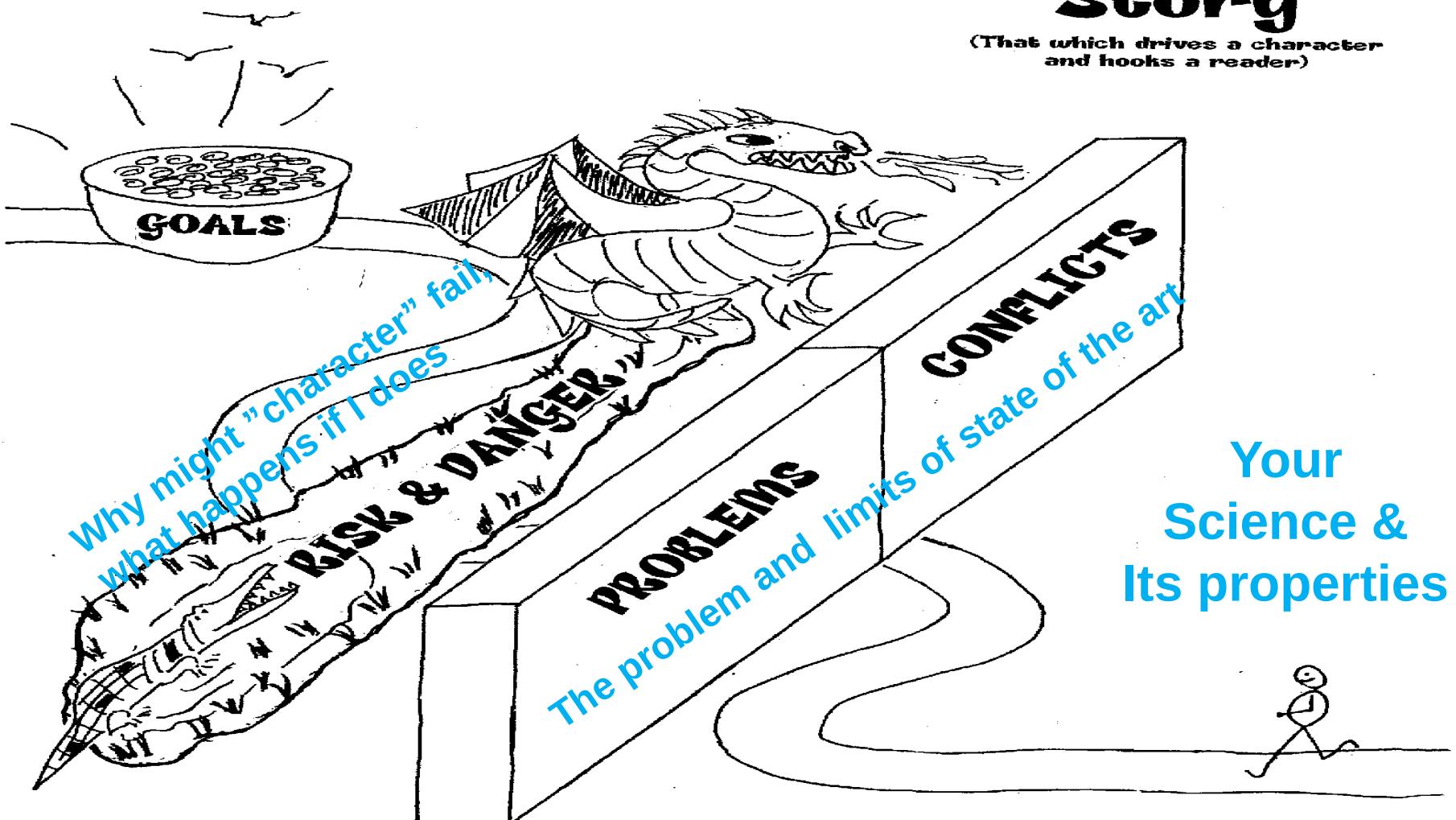
Ablation study or experiments on parts of what make it work (related back to traits)

## 8. WHAT SENSORY DETAILS WILL MAKE THE STORY SEEM REAL?

Show examples; show some failure cases

# The Core of a Story

(That which drives a character  
and hooks a reader)



Your  
Science &  
Its properties

# **THE SCIENCE OF STORY for Science (IN 3 PARTS)**

- 1. THE SCIENCE OF STORY STRUCTURE**
- 2. ANTICIPATING THE MIND OF THE AUDIENCE**
- 3. THE STORY PROCESS**

# **1. THE SCIENCE OF STORY STRUCTURE**

# **PEOPLE LEARNED:**

- TO **READ (most people)...**

**...200 YAG (YEARS AGO)**

# **MOST PEOPLE LEARNED:**

- TO READ

**200 YAG**

- **LOGIC & ARGUMENTATIVE FORMS...**

**...3,200 YAG**

# **PEOPLE LEARNED:**

- TO READ **200 YAG**
- LOGIC & ARGUMENTATIVE FORMS **3,200 YAG**
- TO WRITE (**advanced individuals**)...  
**...7,000 YAG**

# **PEOPLE LEARNED:**

- TO READ **200 YAG**
- LOGIC & ARGUMENTATIVE **3,200**  
**YAG**
- TO WRITE **7,000**  
**YAG**

- TO USE **STORYTELLING...**

**...>100,000 YAG!**

- **100,000 years of storytelling dominance in human interaction**  
**to communicate and to archive learning, wisdom, fact, knowledge, values, beliefs, history, etc.**

Has **evolutionarily rewired** the human brain to think in specific (~~or maybe it's~~ **likely** storytelling evolved to use specific models that are match how humans think/remember).

# **THE HUMAN BRAIN IS PREDISPOSED TO THINK IN **STORY TERMS.****

- **TO UNDERSTAND**
- **TO MAKE SENSE**
- **TO REMEMBER**

**PERSON #1: “WHERE’S JOHN?”**

**PERSON #2: “WELL....I DIDN’T WANT  
TO SAY ANYTHING.**

**BUT...I SAW A GREEN VW PARKED IN  
FRONT OF CAROL’S.”**

- DID ANYONE SAY THAT THERE IS  
**NO CONNECTION?**
- DID ANYONE SAY THAT IT MAKES **NO SENSE?**

**NO!**

HUMAN MINDS AUTOMATICALLY **ASSUME** A CONNECTION  
AND **ASSUME** IT MAKES SENSE

AND USE **STORY STRUCTURES**  
TO FILL IN MISSING INFORMATION.

# OUR STORIED MINDS IN ACTION:

- **Person 1:** “*Hi John.*”
- **Person 2:** “*Shhhh! I’m not here! You never saw me. I’m not here!*”
- **Person 1:** “*It’s okay. Carol’s gone home.*”

## **2. ANTICIPATING THE MIND OF THE AUDIENCE**

# **PRIOR KNOWLEDGE**

**“BANKS OF EXISTING  
KNOWLEDGE AND EXPERIENCE  
ABOUT A RELEVANT TOPIC OR  
NARRATIVE STRUCTURE.”**

- Prior knowledge creates **context** for new information.
- Story structure makes new information **relevant**.

# THE MAGIC INGREDIENTS

- **Context:** “**The surrounding parts that determine meaning; background.**”
- **Relevance:** “**Implying close relationship or importance.**”

**APPLYING PRIOR KNOWLEDGE IS HOW WE REMEMBER.**

**HERE ARE 10 GROUPS WITH 25 LETTERS.  
IN 10 SECONDS, TRY TO MEMORIZE THE  
GROUPS  
IN ORDER:**

**AT TIB MAP PLEM ICR OS  
OF TT WIT TER**

**HOW MANY CAN YOU RECALL IN ORDER?  
(THE AVERAGE IS JUST UNDER 10  
LETTERS.)**

**DIFFERENT SPACING WILL EVOKE  
PRIOR KNOWLEDGE**

**PRIOR KNOWLEDGE CREATES CONTEXT  
AND RELEVANCE.**

**ATT IBM APPLE MICROSOFT TWITTER**

**IT'S THE SAME LETTERS.**

**AT TIB MAP PLEM ICR OS OF TT  
WIT TER**

**NOW MEMORY LAUGHABLY EASY.**

**THAT IS THE POWER OF PRIOR KNOWLEDGE AND  
PROPER GROUPING.**

# **PRIOR KNOWLEDGE**

**CONSIDER THE FOLLOWING SEVEN  
SENTENCES:**

**JOHN WALKED ON THE ROOF.  
BILL PICKED UP THE EGGS.  
PETE HID THE AX.**

**JIM FLEW THE KITE.**

**FRANK BUILT THE BOAT.**

**HARVEY FLIPPED THE ELECTRIC SWITCH.  
TED WROTE THE PLAY.**

**HOW MANY CAN YOU REMEMBER?**

**WHO BUILD THE BOAT? WHO FLEW THE  
KITE? ETC.?**

**YOU UNDERSTOOD THE SENTENCES,  
BUT HAVE NO CONTEXT OR RELEVANCE  
FOR THEM,  
SO WEREN'T ABLE TO REMEMBER.**

**CHANGE THE NAMES TO INVOKE PRIOR  
KNOWLEDGE.**

**SANTA CLAUSE WALKED ON THE ROOF.  
THE EASTER BUNNY PICKED UP THE EGGS.**

**GEORGE WASHINGTON HID THE AX.**

**BENJAMIN FRANKLIN FLEW THE KITE.**

**NOAH BUILT THE BOAT.**

**THOMAS EDISON FLIPPED THE ELECTRIC  
SWITCH.**

**WILLIAM SHAKESPEARE WROTE THE PLAY.**

**PRIOR KNOWLEDGE  
AND  
FAMILY STORIES**

**WHEN YOU TELL STORIES TO THE  
FAMILY THERE IS MUCH YOU  
OMIT BECAUSE THEY ALREADY  
KNOW (THEY HAVE BANKS OF  
PRIOR KNOWLEDGE).**

## **THE CURSE OF KNOWLEDGE:**

**“ONCE YOU ***KNOW***, IT IS ***IMPOSSIBLE***  
TO REMEMBER WHAT IT WAS LIKE  
TO ***NOT KNOW***.”**

**ONCE YOU ***KNOW***, YOU TEND TO WRITE  
(AND TELL OTHERS) AS IF ***EVERY***  
***READER/LISTENER ALSO KNEW***.**

**WRITE FROM THE AUDIENCES KNOWN  
BANKS OF PRIOR KNOWLEDGE**

**NOT FROM YOURS; NOT FROM THE  
BEST SCIENTISTS; AT MOST WRITE  
FROM THE AVERAGE AUDIENCE  
MEMBER.**

**BETTER WRITE TO BELOW AVERAGE BUT  
MAKE THEM FEEL ABOVE AVERAGE**

**CAN HAVE 2-3 PARAGRAPHS FOR THE  
BEST SCIENTISTS**

**FACT IS MADE RELEVANT BY TURNING IT INTO STORY.**

**TURNER (1996).** "SCIENCE WRITINGS IMPLY (BURY) MOST OF THE KEY STORY ELEMENTS. MAKE THOSE EXPLICIT AND INFORMATION TURNS INTO STORY! "

**EXAMPLE:** **MOTHER POURS MILK INTO A GLASS.** (AN EVENT)  
**UNSTATED:** GOAL/MOTIVE/RESOLUTION/OBSTACLES

**MAKE THOSE EXPLICIT:** MOTHER'S LEFT SIDE HAS BEEN PARALYZED BY A CRIPPLING STROKE. SHE FIGHTS TO REGAIN USE OF LEFT HAND AND ARM. FOR THE FIRST TIME SHE TRIES A COMPLEX ACTION WITH HER LEFT HAND. SHE IS DETERMINED TO POUR THE MILK TO PROVE THAT SHE CAN....

**NOW ITS A STICKY STORY!**

# **THE STORY *PROCESS***

**THE PROCESS:**

**MERGING YOUR INFORMATION  
CONTENT**

**WITH**

**EFFECTIVE STORY STRUCTURE**

# The Story Process:

1. Create **theme** and “take away” message.
2. Search for a core **metaphor**.
3. Define/develop story **characters**.
4. Build story **elements** around character and metaphor.
5. **Adjust** for audience and media.

# Theme Versus Goal

## **GOAL:**

Is for **characters**

Should be physical,  
tangible, identifiable

Defines the story's ending

Creates purpose and  
meaning for the story

## **THEME:**

Is for **readers**

Relates to the human  
condition

What writer wants readers  
to ponder

Let's readers interpret  
characters & events

Characters need not be  
aware of theme

**Theme** is the core of your message.

**Story** is the vehicle that effectively communicates your message and

can help the message ***stick!***

# THE **BARE BONES** VERSION OF THE **EIGHT ESSENTIAL ELEMENTS**

\_\_\_\_\_ (CHARACTER)

**NEEDED** \_\_\_\_\_

**BECAUSE** \_\_\_\_\_

**BUT** \_\_\_\_\_

**SO,** \_\_\_\_\_

**FINALLY** \_\_\_\_\_

# OSCAR And Story Elements



- Opening
- Setting
- Challenge
  
- Action
- Resolution

(CHARACTER) \_\_\_\_\_  
NEEDED \_\_\_\_\_  
BECAUSE \_\_\_\_\_  
BUT \_\_\_\_\_

SO, \_\_\_\_\_

FINALLY \_\_\_\_\_

# **Remeber OSCAR is applied recursively**

- Paper level
- Section Level
- Paragraph level (S and C become smaller)
- Sentence Level often but not always the order often some parts missing. More variety here. e.g. can also be OAR or RA depending on context.

**Learn to frame what's important**



# **Openings**

- Are critical to framing the paper. The first sentence is very audience specific.
- Generally needs be a hook, intellectually or emotionally for Audience
- Worth spend lots of time on this.. for abstract and for intro and for conclusion.

# Setting the Science Scene

- Intro should have a bit of setting the science context and the “problem”, maybe even danger. What is the most relevant existing science. Why should we care
- Setup non-standard important terms.
- This is NOT a lit review (Many papers will also have a “related work” section..
- Its just framing the problem and showing reader that others are working in related areas too.
- Be Positive, not critical.

# **The Challenge is defines your stories PROBLEMS & CONFLICTS AND RISK & DANGER**

- ARE **ALWAYS THERE**
- DRIVE **ALL STORIES**
- DOES **NOT MEAN STORIES**

**ARE ABOUT FAILURE**

**• THEY ARE CHALLENGES  
ALONG THE ROAD TO **SUCCESS.****

**DOES “DANGER” AUTOMATICALLY MEAN “PHYSICAL  
DANGER?”**

**ABSOLUTELY *NOT!***

**ANY EMOTIONAL, MENTAL, SOCIAL, PROFESSIONAL, OR  
PHYSICAL DANGER WILL DO.**

**(EMBARRASSMENT, REJECTION, LACK OF KNOWLEDGE &  
UNDERSTANDING, LACK OF INFORMATION, FAILURE,  
RIDICULE, BELITTLEMENT, BEING LAUGHED AT, BEING  
ABANDONED, BEING SCORNED, BEING UNWANTED, FEARS,  
ETC., ETC.)**

**IN SCIENCE/ENG ACCURACY, SPEED, COST ARE ALSO  
DANGERS.**

# Resolution

- Conclusion vs Discussion.
- Good model is OSCAR in reverse.. state the resolution, review the action, science setting (now science clearly stating contribution) and resolve the opening.
- What about future work?
  - Best is to put it in the body of the method or experiments (no need to summarize at end)
  - “Bad new sandwich”
  - Do it as But YES.. not YES But.. because But yes Ends strong

# **General comments on writing**

- Start early.
- Work on the story before writing too much and before doing any/all experiments
- Edit often
- Review how the story works with experimental results
- Edit some more
- Get feedback
- Edit even more

# **Common Avoidable Sins in papers**

- Grandmotherhood opening/ending
- A table of contents in a paragraph form
- Undercutting yourself with textbook level material
- Conclusions that don't conclude  
(summary != Conclusion)
- Opening/Conclusion mismatch

# Highlight information by repeating it.

- Repeat *major ideas* in several places in the document.  
**BUT ALWAYS DIFFERENT views of the same idea**
- Repeat same *keywords*:
  - “deep network” vs “CNN”
  - Don’t go for “variety” to keep it from being repetitive. You may read it 100 times, but the reader does it 1-2 times. So its not nearly as repetitive as you might think. If you used it < 20 times, go ahead reuse it as its not really “repetitive” yet.
  - DO avoid instant repetition/tautologies (e.g. basic fundamental, final ending, true facts mutual cooperation

# Writing Startup

- 1) Manage/Plan your time. This is a key element of being effective.
- 2) Don't think about writing. This is about communication.
- 3) Generation of ideas
  - 1) Using outlines
  - 2) Questioning, Question/listing, Answering..
  - 3) Brainstorm outlining ..
    - (for a twist.. try brainstorming with only action verbs, words that rhyme, words that start with a particular letter, etc. )
  - 4) Post-it Notes or cards..
  - 5) Free writing
  - 6) Free talk / Dictations..

# Writing a Research Paper

Based in part my own experiences plus slides from E. Kraemer and S. Peyton Jones. Note in parallel to "story telling" this is more about structure/process



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

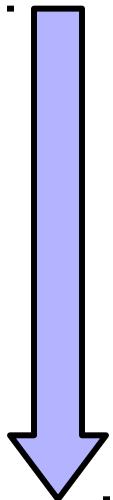
# Related work != Literature Survey

- ◆ “Related work” section is different than a survey paper.  
Don’t need a “categorization”
- ◆ Want 1-2 papers that set the stage of the problem 3-5  
papers that use related tools, 2-3 papers that are state  
of the art and to which you will compare.
- ◆ If your approach is a combination of 2-3 ideas may need  
more.
- ◆ Be positive and talk about the good ideas/results in the  
papers, but if they missed some dimension you will  
explore, say so.

# Writing papers is a skill

- ◆ Many papers are badly written
- ◆ Good writing is a skill you can learn
- ◆ It's a skill that is worth learning:
  - You will get more papers accepted, and brownie points (e.g. best paper awards)
  - Your ideas will have more impact
  - You will have better ideas

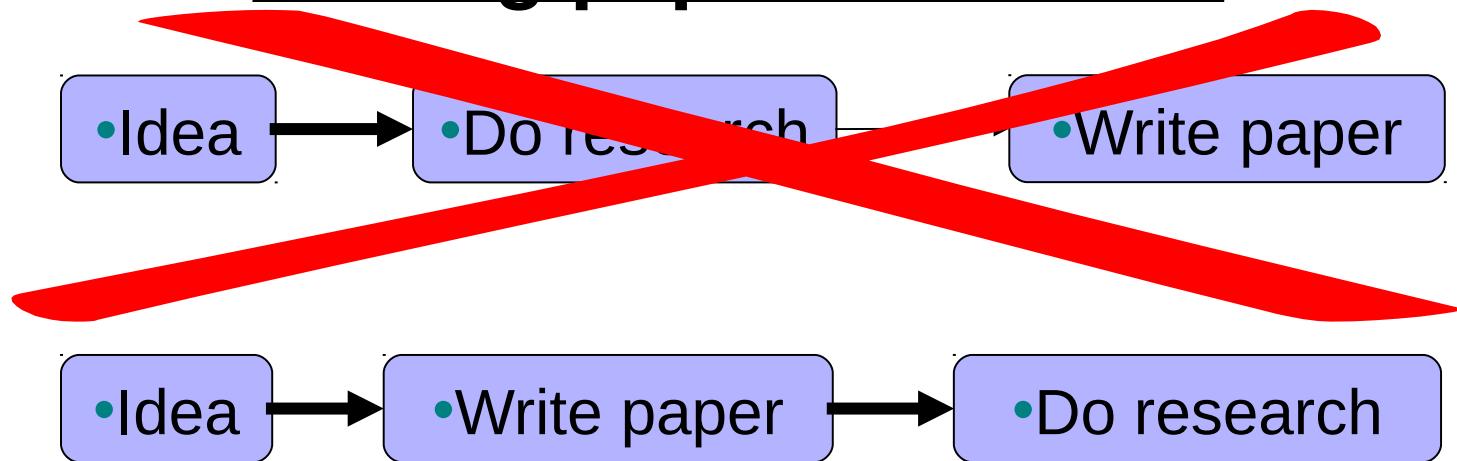
Increasing importance



# Writing papers: model 1

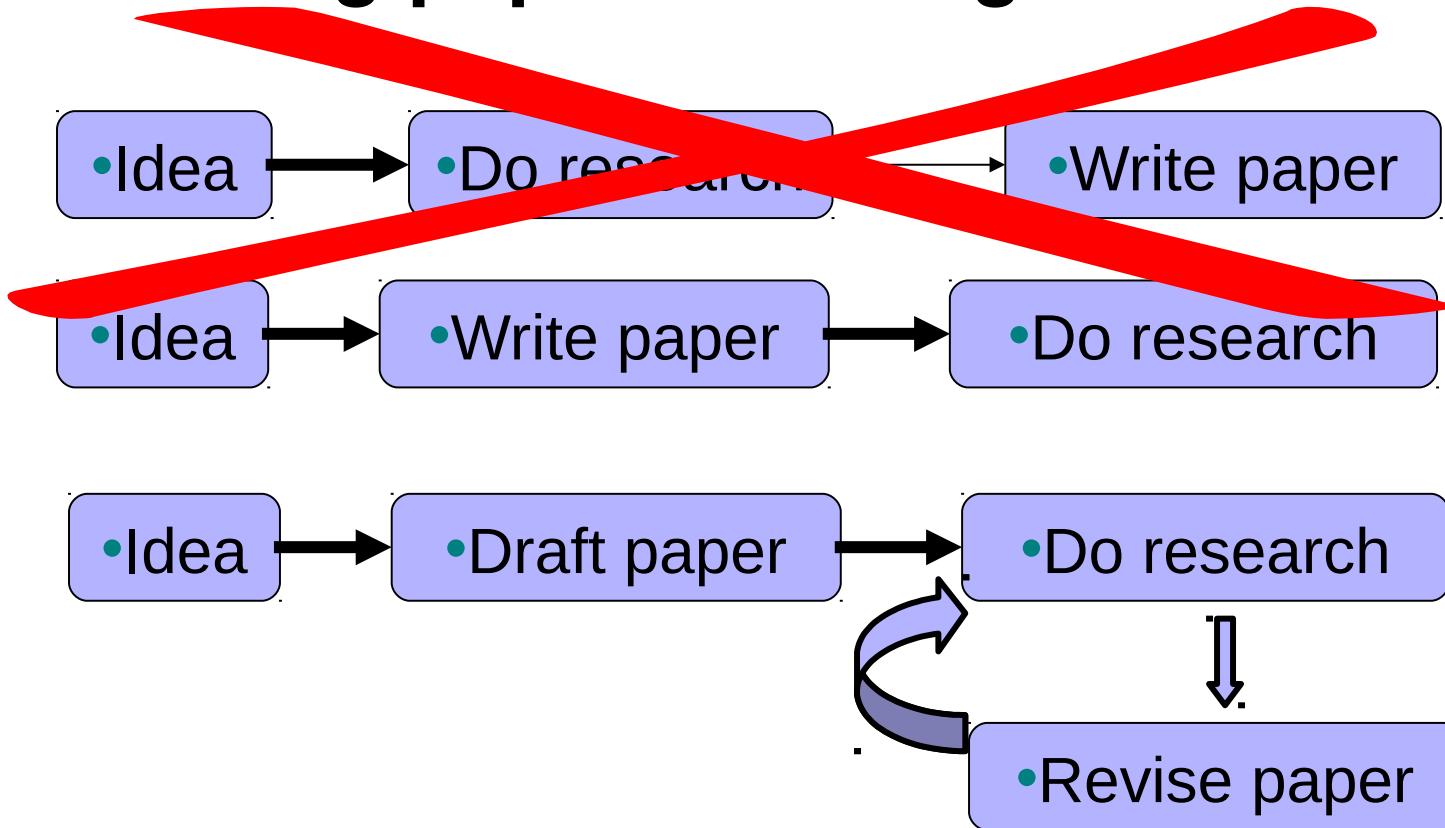


## Writing papers: model 2



- ◆ Forces us to be clear, focused
- ◆ Crystallizes what we don't understand
- ◆ Opens the way to dialogue with others:  
reality check, critique, and collaboration

# Writing papers: The right model



- ◆ Adjust to what is learned in research experiments.
- ◆ Allows you to find the best “story” in what you have

## **Do not be intimidated**

### **Fallacy**

You need to have a fantastic idea before you can write a paper.  
(Everyone else seems to.)

- Write a paper,

and give a talk, about

## **any novel idea,**

no matter how weedy and insignificant it may seem to you

## **Do not be intimidated**

- Write a paper, and give a talk, about any novel idea, no matter how insignificant it may seem to you

- **Writing the paper is how you develop the idea in the first place**
- It usually turns out to be more interesting and challenging than it seemed at first

# The purpose of your paper

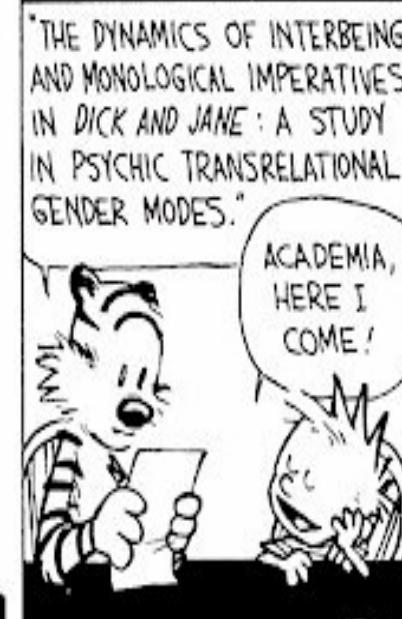
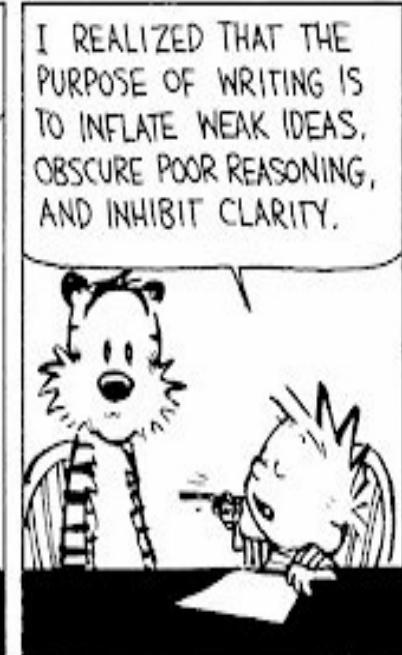


University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

# Why bother?

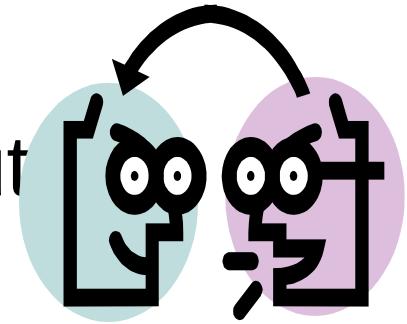
## Fallacy

we write papers and give talks mainly to impress others, gain recognition, and get promoted



# Papers communicate ideas

- ◆ Your goal: to infect the mind of your reader with **your idea**, like a virus
- ◆ Papers ideas are far more durable than programs
- ◆ Ideas may be a dime a dozen, but Good papers are worth a lot.



- The greatest ideas are (literally) worthless if you keep them to yourself

# The Idea

Idea

A re-usable insight,  
useful to the reader

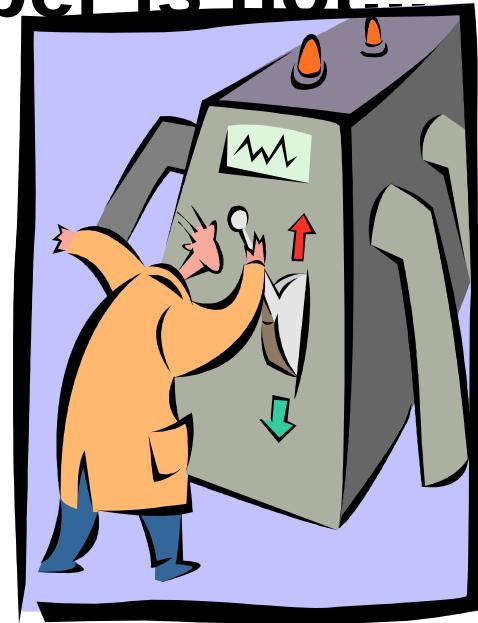
- ◆ Figure out what your idea(s) is/are. These are the MAIN CHARACTERS in your story. Have at least 1 main and at most 2 other supporting ideas in the paper
- ◆ Make certain that the reader is in no doubt what the idea is. Be 100% explicit, usually at end of intro:
  - “The main idea of this paper is....”
  - The main contributions of this paper are: 1) 2)...
- ◆ Many papers contain good ideas, but do not distil what they are.

## **One “ping” - one sticky idea**

- ◆ Your paper should have just one “ping”: one clear, sharp idea – the main character. Any other contributions are supporting players in the story.
- ◆ Read your intro and whole paper again: can you hear the “ping”? Is there one think will stick with the reader
- ◆ You may not know exactly what the ping is when you start writing; but you must know when you finish
- ◆ If you have lots of ideas, write lots of papers. One main idea and 1-2 support is enough for a 8-20 page paper.

# The purpose of your paper is not ...

## To describe the WizWoz system



- Your reader does not have a WizWoz. If they did then the paper is not novel
- She is primarily interested in re-usable brain-stuff, not executable artefacts

## Your narrative flow

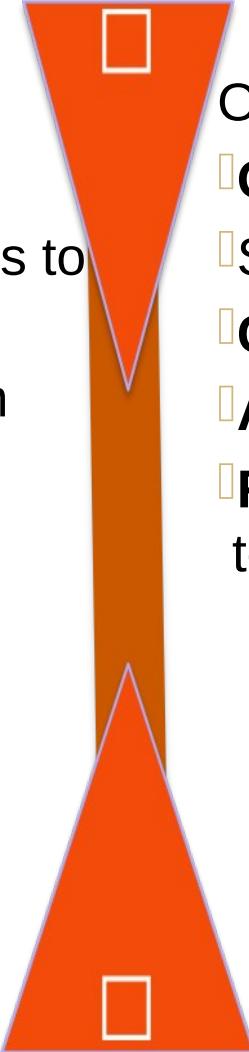
- ◆ Here is a problem
- ◆ It's an interesting problem
- ◆ Here's how my idea compares to other people's approaches
- ◆ See it is an unsolved problem
- ◆ **Here is my idea**
- ◆ My idea works (details, data)

I wish I knew how to solve that!

I see how that works. Ingenious!



# Your narrative flow

- ◆ Here is a problem
  - ◆ It's an interesting problem
  - ◆ Here's how my idea compares to other people's approaches
  - ◆ See it is an unsolved problem
  - ◆ **Here is my idea**
  - ◆ My idea works (details, data)
- 
- OSCAR
- | Opening (brotsad hook)
  - | Setting the (Science) Stage
  - | Challenge (the problem)
  - | Action (Alg & Experiments)
  - | Resolution (Must tie back to opening)

## **Structure (conference paper)**

Title (1000 readers)

Abstract (4 sentences, 100-1000 readers)

Introduction (1 page, 100 readers)

(a) Related work (1-2 pages, 10 readers)

The problem (1 page, 10-50 readers)

Approach (Key idea\_ (2 pages, 10-50 readers)

The details (5 pages, 5 readers)

(b) Related work (1-2 pages, 10 readers)

Conclusions and further work (0.5 pages)

## The abstract

- ◆ I usually write the abstract first and also last
- ◆ Used by program committee members to decide which papers to read
- ◆ Four sentences [Kent Beck] (=OSCAR)
  1. State the problem
  2. Say why it's an interesting problem
  3. Say what your solution achieves
  4. Say what follows from your solution

## Example

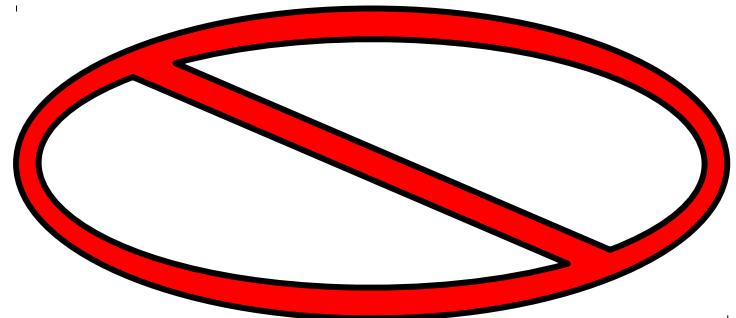
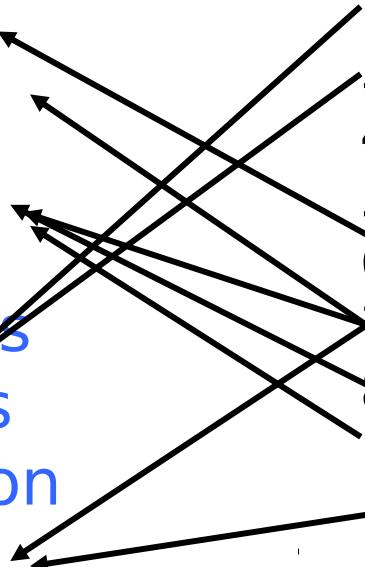
1. Many papers are badly written and hard to understand
2. This is a pity, because their good ideas may go unappreciated
3. Following simple guidelines can dramatically improve the quality of your papers
4. Your work will be used more, and the feedback you get from others will in turn improve your research

## EIGHT ESSENTIAL ELEMENTS

Four sentences [Kent Beck]

1. State the problem
2. Say why it's an interesting problem
3. Say what your solution achieves
4. Say what follows from your solution

1. CHARACTER
2. CHARACTER TRAITS (THAT MAKE CHARACTERS INTERESTING)
3. GOAL
4. MOTIVE
5. CONFLICTS & PROBLEMS
6. RISK & DANGER
7. STRUGGLES
8. DETAILS



## Abstract No-nos

- ◆ Avoid putting "citations" in abstract. Why?
  - ◆ But do refer to the state of the art. How?
  - ◆ Refer to any datasets? Why? How?
- 
- ◆ Avoid putting equations in abstract why?
  - ◆ Keep it short, it is not the introduction.

# Structure

Abstract (4 sentences)

**Introduction** (1 page)

The problem (1 page)

My idea (2 pages)

The details (5 pages)

Related work (1-2 pages)

Conclusions and further work (0.5 pages)

# The introduction (1 page)

1. **Describe the problem**
2. **1-2 paragraph(s) on w is not solved by SOTA**
3. **State your contributions**

...and that is all

One Figure + ONE PAGE of text max.  
(Unless its intro+related work then 2 max)

# Describe the problem

## 1 Introduction

There are two basic ways to implement function application in a higher-order language, when the function is unknown: the *push/enter* model or the *eval/apply* model [11]. To illustrate the difference, consider the higher-order function `zipWith`, which zips together two lists, using a function `k` to combine corresponding list elements:

```
zipWith :: (a->b->c) -> [a] -> [b] -> [c]
zipWith k []      []      = []
zipWith k (x:xs) (y:ys) = k x y : zipWith xs ys
```

Use an example to introduce the problem

Here `k` is an *unknown function*, passed as an argument; global flow analysis aside, the compiler does not know what function `k` is bound to. How should the compiler deal with the call `k x y` in the body of `zipWith`? It can't blithely apply `k` to two arguments, because `k` might in reality take just one argument and compute for a while before returning a function that consumes the next argument; or `k`

# Use a "teaser figure"

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.??, NO. ??, MARCH 2017

1

## The Extreme Value Machine

Ethan M. Rudd, Lalit P. Jain,  
Walter J. Scheirer, *Senior Member, IEEE*,  
and Terrance E. Boult, *IEEE Fellow*

*Abstract*—It is often desirable to be able to recognize when inputs to a recognition function learned in a supervised manner correspond to classes unseen at training time. With this ability, new class labels could be assigned to these inputs by a human operator, allowing them to be incorporated into the recognition function — ideally under an efficient incremental update mechanism. While good algorithms that assume inputs from a fixed set of classes exist, e.g., artificial neural networks and kernel machines, it is not immediately obvious how to extend them to perform incremental learning in the presence of unknown query classes. Existing algorithms take little to no distributional information into account when learning recognition functions and lack a strong theoretical foundation. We address this gap by formulating a novel, theoretically sound classifier — the Extreme Value Machine (EVM). The EVM has a well-grounded interpretation derived from statistical Extreme Value Theory (EVT), and is the first classifier to be able to perform nonlinear kernel-free variable bandwidth incremental learning. Compared to other classifiers in the same deep network derived feature space, the EVM is accurate and efficient on an established benchmark partition of the ImageNet dataset.

*Index Terms*—Machine Learning, Supervised Classification, Open Set Recognition, Open World Recognition, Statistical Extreme Value Theory

## I. INTRODUCTION

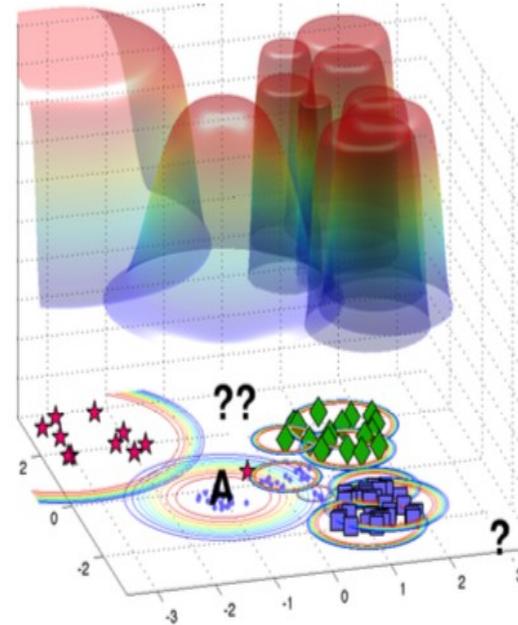


Fig. 1. A solution from the proposed EVM algorithm trained on four classes: dots, diamonds, squares, and stars. The colors in the isocontour rings show a  $\Psi$ -model (probability of sample inclusion) for each *extreme vector* (EV) chosen by the algorithm, with red near 1 and blue near .005. Via kernel-free non-linear modeling, the EVM supports open set recognition and can reject the three “?” inputs that lie beyond the support of the training set as “unknown.” Each  $\Psi$ -model has its own independent shape and scale parameters learnt from the data, and supports a soft-margin. For example, the  $\Psi$ -model for the blue dots corresponding to extreme vector A has a more gradual fall off, due to the effect of the outlier star during training.

# Ideal "Teaser" = graphic of key idea + results

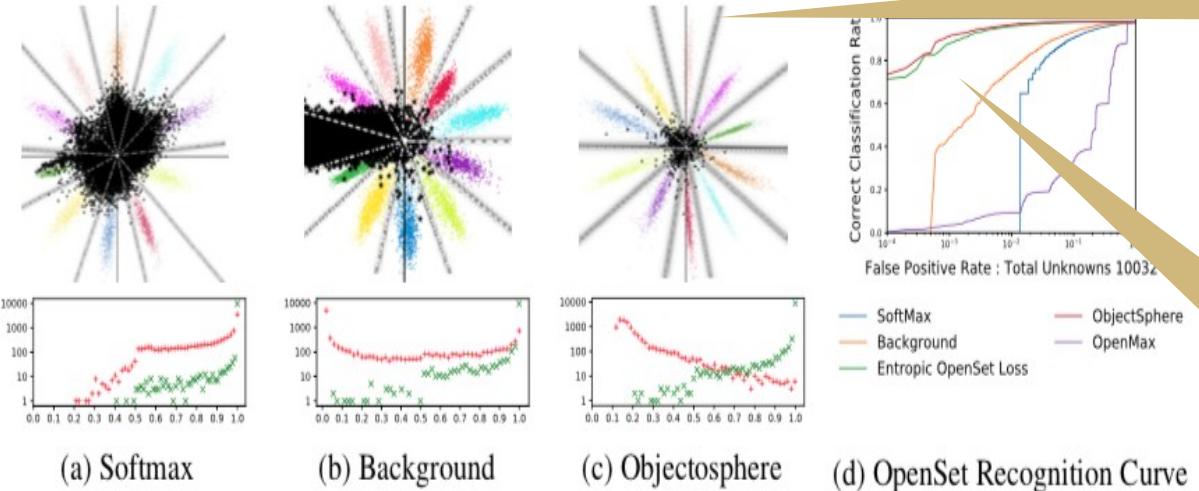


Figure 1: LENET++ RESPONSES TO KNOWNS AND UNKNOWNS. The network in (a) was only trained to classify the 10 MNIST classes while the networks in (b) and (c) added MNIST digits [17] as background examples (known unknowns). In the feature representation plots on top, colored dots represent  $\mathcal{D}_c$ , i.e., test samples from the ten MNIST classes, while black dots represent samples from the Devanagari [29] dataset ( $\mathcal{D}_a$ ). The dashed gray-white lines indicate class borders, i.e., where softmax scores for neighboring classes are equal. This paper addresses how to improve recognition by reducing the overlap of network features from known samples  $\mathcal{D}_c$  with features from background or unknown samples  $\mathcal{D}_u$ . The figures in the bottom are histograms of softmax probability values for samples of  $\mathcal{D}_c$  and  $\mathcal{D}_a$  with a logarithmic vertical axis. The OSCR curve in (d) depicts the probability of the correct class for samples of  $\mathcal{D}_c$  and the maximum probability of any known class for samples of  $\mathcal{D}_a$ . In an application, a score threshold  $\theta$  should be chosen to optimally separate unknown from known samples. Unfortunately, such a threshold is difficult to find for either (a) or (b), a better separation is achievable with the Objectosphere loss (c). The proposed Open-Set Classification Rate (OSCR) curve in (d) depicts the high accuracy of our approach even at a low false positive rate.

Graphics that tell the story of the idea

Results "teaser" (refer back later)

Caption that explains it all

# State your contributions

- ◆ Write the list of contributions first
- ◆ **The list of contributions drives the entire paper:** the paper substantiates the claims you have made
- ◆ Aim to have the Reader think “gosh, if they can really deliver this, that’s be exciting; I’d better read on”
- ◆ Last part of introduction, with itemize, Italic/bold to catch the eye

# State your contributions

performance of classification module of detection networks by better handling false positives from the region proposal network.

**Our Contributions:** In this paper, we make four major contributions: *a*) we derive a novel loss function, the Entropic Open-Set loss, which increases the entropy of the softmax scores for background training samples and improve the handling of background and unknown inputs, *b*) we extend that loss into the Objectosphere loss, which further increases softmax entropy and performance by minimizing the Euclidean length of deep representations of unknown samples, *c*) we propose a new evaluation metric for comparing the performance of different approaches in an unknown space, and *d*) we show that the new loss functions advance the state of the art for open-set image classification.

Do not leave the reader to guess what  
your contributions are!

If you have space itemize/enumerate list  
is better But with Bold/italics can do inline  
list

Enumerated  
list of  
contribution  
s

# be refused

Can get section information in contributions

| NO!   | YES!  |
|---|---|
| We describe the WizWoz system. It is really cool. | We give the syntax and semantics of a language that supports concurrent processes (Section 3). Its innovative features are...                 |
| We study its properties                           | We prove that the type system is sound, and that type checking is decidable (Section 4)   |
| We have used WizWoz in practice                   | We have built a GUI toolkit in WizWoz, and used it to implement a text editor (Section 5). The result is half the length of the Java version. |

## No “rest of this paper is...”

- ◆ Never:

“The rest of this paper is structured as follows. Section 2 introduces the problem. Section 3 ... Finally, Section 8 concludes”.

- ◆ Instead, **use forward references from the narrative in the introduction.**

The introduction (including the contributions) should touch on all elements of the whole paper, and therefore can forward reference every important part. (But generally not needed)

# Ideal Structure (but carries some risk)

Abstract (4 sentences)

Introduction (1 page)

## Related work

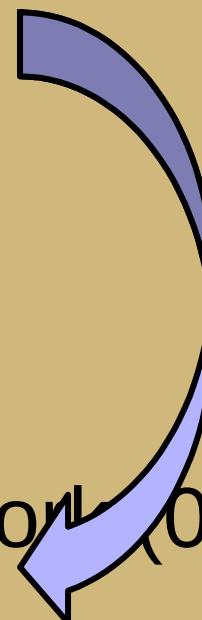
The problem (1 page)

My idea (2 pages)

The details (5 pages)

Related work (1-2 pages)

Conclusions and further work (0.5 pages)

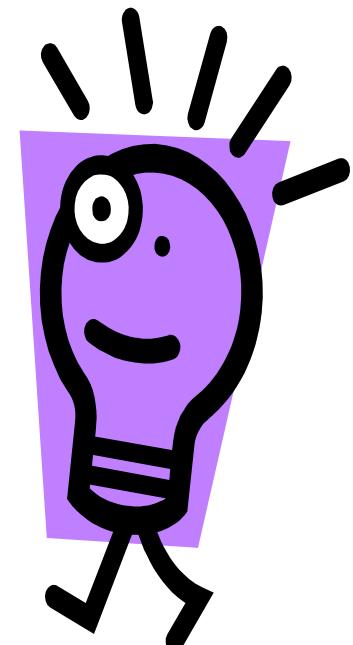
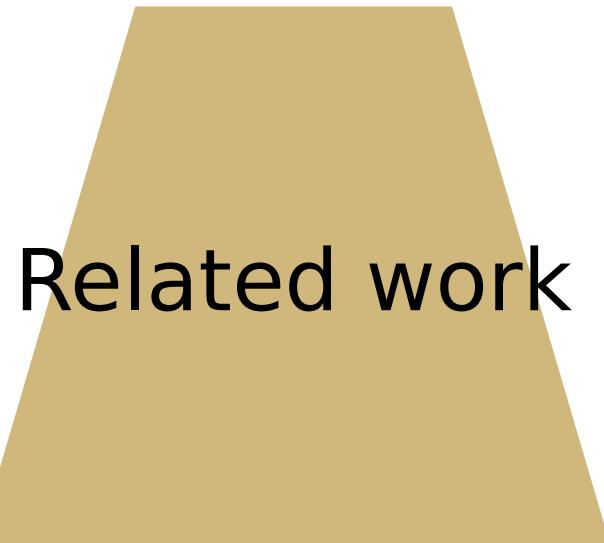


# General papers -- No related work

## yet!



Your reader

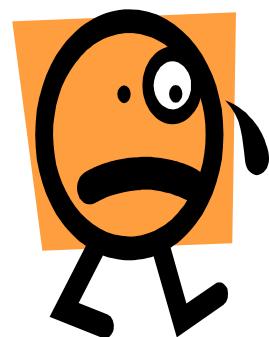


Your idea

We adopt the notion of transaction from Brown [1], as modified for distributed systems by White [2], using the four-phase interpolation algorithm of Green [3]. Our work differs from White in our advanced revocation protocol, which deals with the case of priority inversion as described by Yellow [4].

## No related work yet

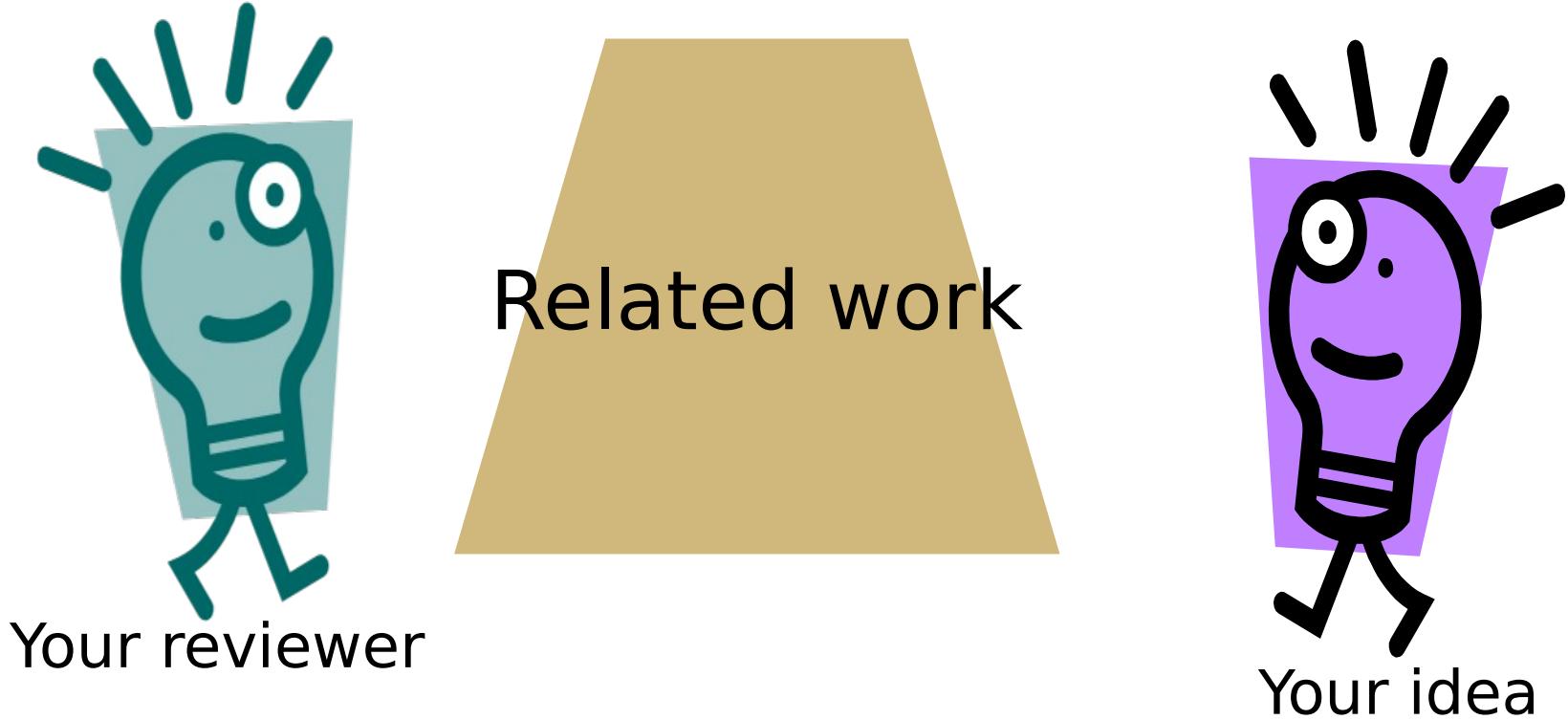
- ◆ **Problem 1:** the reader knows nothing about the problem yet; so your (carefully trimmed) description of various technical tradeoffs is absolutely incomprehensible
- ◆ **Problem 2:** describing alternative approaches gets between the reader and your idea
- ◆ **Problem/Solution 3:** Readers skip it in first read, and come back later.



I feel  
stupid

I feel  
tired

## With "aggressive" reviewers



We adopt the notion of transaction from Brown [1], as modified for distributed systems by White [2], using the four-phase interpolation algorithm of Green [3]. Our work differs from White in our advanced revocation protocol, which deals with the case of priority inversion as described by Yellow [4].

# **Structure**

Abstract (4 sentences)

Introduction (1 page)

The problem (1 page)

Approach (My idea) (2 pages)

The details & Experiments (5 pages)

Related work (1-2 pages)

Conclusions and further work (0.5 pages)

## **Presenting the idea**

### **3. The idea**

Consider a bifircuated semi-lattice D, over a hyper-modulated signature S. Suppose  $p_i$  is an element of D. Then we know for every such  $p_i$  there is an epi-modulus j, such that  $p_j < p_i$ .

- Sounds impressive...but
- Sends readers to sleep
- In a paper you MUST provide the details,  
but FIRST convey the idea

## Presenting the idea

- ◆ Explain it as if you were speaking to someone using a whiteboard. Someone not in your field
- ◆ **Conveying the intuition is primary**, not secondary
- ◆ Once your reader has the intuition, she can follow the details (but almost never vice versa)
- ◆ Even if she skips the details, she still takes away something valuable

## Putting the reader first

- ◆ **Do not** recapitulate your personal journey of discovery. This route may be soaked with your blood, but that is not interesting to the reader.
- ◆ Instead, choose the most direct route to the idea.

# The payload of your paper

Introduce the problem,  
and your idea, using

**EXAMPLES**

and only then present the  
general case

## 2 Background

To set the scene for this paper, we begin with a brief overview of the *Scrap your boilerplate* approach to generic programming. Suppose that we want to write a function that computes the size of an arbitrary data structure. The basic algorithm is “for each node, add the sizes of the children, and add 1 for the node itself”. Here is the entire code for `gsize`:

```
gsize :: Data a => a -> Int  
gsize t = 1 + sum (gmapQ gsize t)
```

The type for `gsize` says that it works over any type `a`, provided `a` is a *data* type — that is, that it is an instance of the class `Data`<sup>1</sup>. The definition of `gsize` refers to the operation `gmapQ`, which is a method of the `Data` class:

```
class Typeable a => Data a where  
  ...other methods of class Data...
```

Example  
right  
away

## The details: evidence

- ◆ Your introduction makes claims
- ◆ The body of the paper provides **evidence to support each claim**
- ◆ Check each claim in the introduction, identify the evidence, and forward-reference it from the claim
- ◆ Evidence can be: analysis and comparison, theorems, measurements, case studies

# Adding Mathematical rigor

- ◆ Try to formalize the approach and add rigor.
- ◆ Especially important if the idea seems “Simple”
- ◆ Great place to engage your advisor or bring in collaborators.
- ◆ Often this stage can change the idea/algorithm as it evolves between ad-hoc heuristics and a formal mathematical model that explains the data and/or works well.



## How to build a model

- ◆ What models are used in related papers?
- ◆ How do you adjust those models to your problem/approach?
- ◆ What mathematical models related to your fundamental assumptions?

# Structure

Abstract (4 sentences)

Introduction (1 page)

The problem (1 page)

My idea (2 pages)

The details (5 pages)

**Related work** (1-2 pages)

Conclusions and further work (0.5 pages)

## Related work

### Fallacy

To make my work look good,  
I have to make other  
people's work look bad

## The truth: credit is not like money

Giving credit to others does not diminish  
the credit you get from your paper

- Warmly acknowledge people who have helped you
- Be generous to the competition. “In his inspiring paper [Foo98] Foogle shows.... We develop his foundation in the following ways....”
- Acknowledge weaknesses in your approach

# Acadmeic Credit is not like Finacial Credit

If you imply that an idea is yours, and the referee knows it is not, then either

- You don't know that it's an old idea (bad)
- You do know, but are pretending it's yours (very bad)

Failing to give credit to others can kill  
your paper

# Structure

Abstract (4 sentences)

Introduction (1 page)

Background and related work (1-2 pages)

My idea (2 pages)

The details (5 pages)

Related work (1-2 pages)

Conclusions and further work (0.5 pages)

## **Conclusions and further work**

- ◆ Be brief.
- ◆ Be a conclusion, not a summary
- ◆ Don't have to explicitly talk about future work.
  - Some risk if you do the reviewers might say.. That is what the paper should be, come back when that is done.

# The process of writing



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

## The process

- ◆ Start early. Very early.
  - Hastily-written papers get rejected.
  - Papers are like wine: they need time to mature
  - Start on the story long before you do experiments
- ◆ Collaborate; Use GIT to support collaboration
- ◆ Discuss “story” with collaborators; Don’t expect your advisor to be your grammar checker/editor, but do address when you get that Feedback

## Getting help

Get your paper read by as many friendly guinea pigs as possible

- ◆ Experts are good
- ◆ Non-experts are also very good
- ◆ Each reader can only read your paper for the first time once! So use them carefully
- ◆ Explain carefully what you want (“I got lost here” is much more important than “Jarva is mis-spelt”)

## Getting expert help

- ◆ A good plan: when you think you are done, send the draft to the competition saying “could you help me ensure that I describe your work fairly?”.
- ◆ Often they will respond with helpful critique (they are interested in the area)
- ◆ They are likely to be your referees anyway, so getting their comments or criticism up front is Jolly Good.

# Listening to your reviewers

**Treat every review like gold dust**

Be (truly) grateful for criticism  
as well as praise

This is really, really, really hard  
But it's  
really, really, really, really, really, really,  
really, really, really, really  
important

## Listening to your reviewers

- ◆ Read every criticism as a positive suggestion for something you could explain more clearly
- ◆ DO NOT respond “you stupid person, I meant X”. Fix the paper so that X is apparent even to the stupidest reader.
- ◆ Thank them warmly. They have given up their time for you.

# Language and style



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

## Basic stuff

- ◆ Submit by the deadline, upload a nearly finished draft > 5 hrs before deadline, then can approach it closely.
- ◆ Keep to the length restrictions
  - Do not narrow the margins
  - Do not use 6pt font
  - On occasion, supply supporting evidence (e.g. experimental data, or a written-out proof) in an appendix
- ◆ Always use a spell/grammer checker

# Visual structure

- Give strong visual structure to your paper using
  - sections and sub-sections
  - bullets
  - italics
  - laid-out code
- Find out how to draw pictures, and use them

# Virtual structure

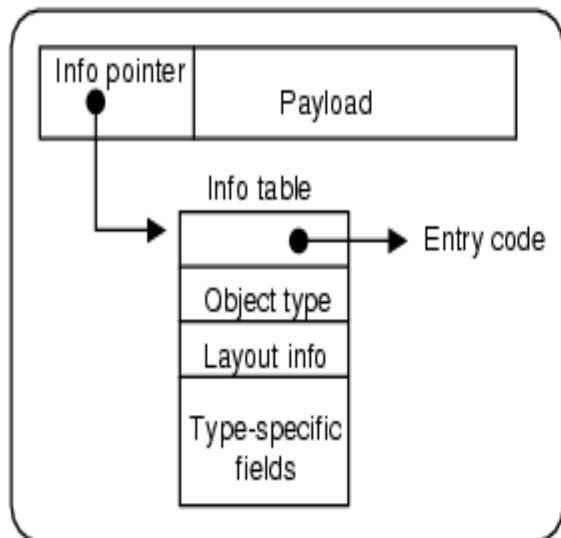


Figure 3. A heap object

The three cases above do not exhaust the possible forms of  $f$ . It might also be a *THUNK*, but we have already dealt with that case (rule THUNK). It might be a *CON*, in which case there cannot be any pending arguments on the stack, and rules UPDATE or RET apply.

## 4.3 The eval/apply model

The last block of Figure 2 shows how the eval/apply model deals with function application. The first three rules all deal with the case of a *FUN* applied to some arguments:

- If there are exactly the right number of arguments, we behave exactly like rule *KNOWNCALL*, by tail-calling the function. Rule *EXACT* is still necessary — and indeed has a direct counterpart in the implementation — because the function might

remainder of the object is called the *payload*, and may consist of a mixture of pointers and non-pointers. For example, the object  $CON(C\ a_1\dots a_n)$  would be represented by an object whose info pointer represented the constructor  $C$  and whose payload is the arguments  $a_1\dots a_n$ .

The info table contains:

- Executable code for the object. For example, a *FUN* object has code for the function body.
- An object-type field, which distinguishes the various kinds of objects (*FUN*, *PAP*, *CON* etc) from each other.
- Layout information for garbage collection purposes, which describes the size and layout of the payload. By “layout” we mean which fields contain pointers and which contain non-pointers, information that is essential for accurate garbage collection.
- Type-specific information, which varies depending on the object type. For example, a *FUN* object contains its arity; a *CON* object contains its constructor tag, a small integer that distinguishes the different constructors of a data type; and so on.

In the case of a *PAP*, the size of the object is not fixed by its info table; instead, its size is stored in the object itself. The layout of its fields (e.g. which are pointers) is described by the (initial segment of) an argument-descriptor field in the info table of the *FUN* object which is always the first field of a *PAP*. The other kinds of heap object all have a size that is statically fixed by their info table.

# Use the active voice

The passive voice is “respectable” but it DEADENS your paper. Avoid it almost always.

**NO**

It can be seen that...

34 tests were run

These properties were thought desirable

It might be thought that this would be a type error

**YES**

We can see that...

We ran 34 tests

We wanted to retain these properties

One might think this would be a type error

“We” =  
you and  
the  
reader

“We” =  
the  
authors

One or  
“You” =  
the reader

# language

**NO**

**YES**

The object under study was  
displaced horizontally

The ball moved sideways

On an annual basis

Yearly

Endeavour to ascertain

Find out

It could be considered that the  
speed of storage reclamation  
left something to be desired

The garbage collector was slow,  
taking 20% of run time

Abstract (4 sentences)

Introduction (1 page)

The problem (1 page)

My idea (2 pages)

The details (5 pages)

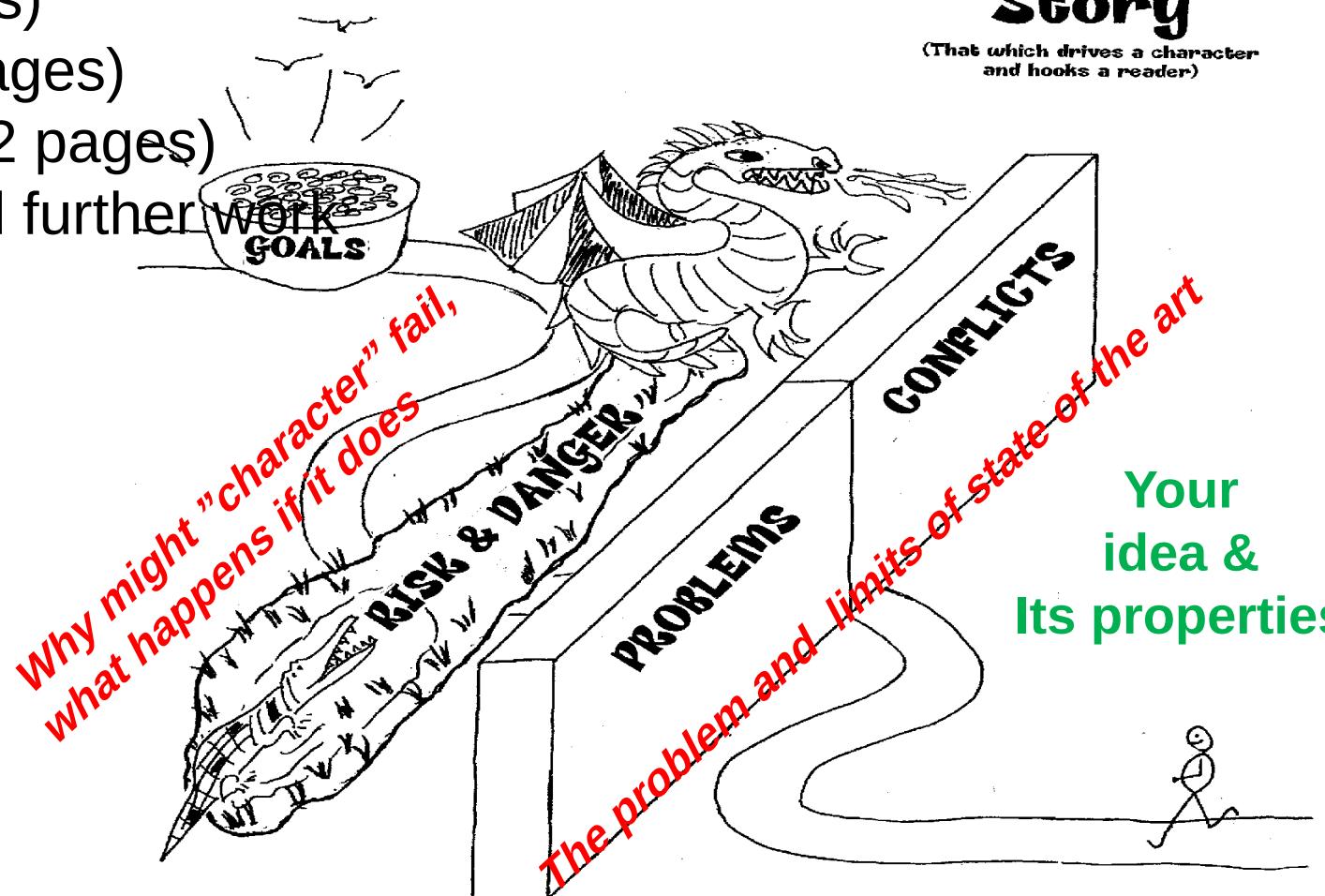
**Related work** (1-2 pages)

Conclusions and further work

(0.5 pages)

## The Core of a Story

(That which drives a character  
and hooks a reader)



# Summary

If you remember nothing else:

- ◆ **Tell a story**
- ◆ **Identify your key idea**
- ◆ **Make your contributions explicit**
- ◆ **Use examples**
- ◆ **Follow the rules**

# Structure of a Survey Paper??



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

## A survey paper is...

"a paper that summarizes and organizes recent research results in a novel way that integrates and adds understanding to work in the field. A survey article assumes a general knowledge of the area; it emphasizes the classification of the existing literature, developing a perspective on the area, and evaluating trends."

## Goals of a Survey

- ◆ Provide reader with a view of existing work that is well organized and comprehensive
  - Not all details must be included, which one's should/shouldn't?
  - Make sure to cover all relevant material completely
  - Logical structure of organization
  - State-of-the-art view

# Your survey paper should ...

- ◆ Summarize the research in 20-80 papers on a particular topic
  - ◆ Include your own commentary on the significance of the approach and the solutions presented in each paper
  - ◆ Provide a critical assessment of the work that has been done
  - ◆ Include a discussion on future research directions
- 
- ◆ **REMEMBER**
    - Everything you write in this survey paper has to be in your own words
    - All ideas, paraphrases of other people's words must be correctly attributed in the body of the paper and in the references
    - Any evidence of it in the survey paper will result in a fail grade

## Article Structure

- ◆ It should not be just a concatenation of paper reviews
- ◆ A typical structure of a paper includes:
  - Title
  - Abstract
  - Introduction
  - Body of paper
  - Conclusion/Future Work
  - References

## Article Structure

- ◆ It should not be just a concatenation of paper reviews
- ◆ A typical structure of a paper includes:
  - Title
  - Abstract
  - Introduction
  - Organizational Approach
  - Selection Criterion
  - Body
  - Conclusion/Future Work
  - References

# Article Structure

## ◆ Introduction

- Importance and significance of the topic
- Discuss the background and target audience
- Summarize the surveyed research area and explain why the surveyed area has been studied
- Summarize the classification scheme you used to do the survey
- Summarize the surveyed techniques with the above classification scheme

# Article Structure

- ◆ Survey details/Body of paper
  - Present the surveyed techniques using the classification scheme in detail
  - Identify the trends in the surveyed area. Give evidences for your decision
  - Identify some leading research/products/companies/web-sites
  - Identify the unresolved problems/difficulties, and future research issues

# Article Structure

- ◆ Conclusions/Future work
  - Summarize the conclusions of your survey
- ◆ References
  - List all the citations referenced in your paper

## Figures

- ◆ Can be taken from papers as long as appropriate credit is given
  - “Figure taken from [28]”.
- ◆ Draw your own figures to show classification or structure of the survey
- ◆ Use tables to organize comparisons between applications/systems/etc

# How to Cite a Reference

- ◆ Cite the full info about the paper
  - Author names
  - Paper title
  - Publication details
  - Page numbers
  - Year, etc

In the text, use [1] to refer  
[1] Adomavicius G, Tuzhilin A., "Toward the Next Generation of  
Recommender Systems: A Survey of the State-of-the-Art and Possible  
Extensions", IEEE Transactions on Knowledge and Data Engineering,  
Vol. 17, No. 6. (June 2005), pp. 734-749.

There are many bibliography formats. Select one and **stick to it**.  
[http://standards.ieee.org/guides/style/2009\\_Style\\_Manual.pdf](http://standards.ieee.org/guides/style/2009_Style_Manual.pdf) (Chap 19)  
<http://sgs.umkc.edu/pdfs/ACM-STYLE-EXAMPLES.pdf>

# General Rules for Bibliography

- ◆ Avoid use of et al. in a bibliography unless list is very long (five or more authors).
- ◆ Internet drafts must be marked ``work in progress".
- ◆ Book citations include publication years, but no ISBN number.
- ◆ It is now acceptable to include URLs to material, but it is probably bad form to include a URL pointing to the author's web page for papers published in IEEE and ACM publications, given the copyright situation. Use it for software and other non-library material. Avoid long URLs; it may be sufficient to point to the general page and let the reader find the material. General URLs are also less likely to change.
- ◆ Leave a space between first names and last name, i.e., "J. P. Doe", not "J.P.Doe".

# What not to do....

## 8. References:

- Embedded systems security

[http://www.ece.cmu.edu/~koopman/security/koopman04\\_embedded\\_security.pdf](http://www.ece.cmu.edu/~koopman/security/koopman04_embedded_security.pdf)

Philip Koopman, Carnegie Mellon University

- Security in embedded systems [ACM, TECS]:

<http://portal.acm.org/citation.cfm?id=1015049> Srivaths Ravi ,Anand Raghunathan NEC Laboratories America

- C. Mulliner, G. Vigna, D. Dagon, and W. Lee, UCSB, "Using Labeling to Prevent Cross-Service Attacks Against Smart Phones [DIMVA]."

[http://www.cs.ucsb.edu/~vigna/publications/2006\\_mulliner\\_vigna\\_dagon\\_lee\\_DIMVA.pdf](http://www.cs.ucsb.edu/~vigna/publications/2006_mulliner_vigna_dagon_lee_DIMVA.pdf)

- Symbian OS Project <http://www.symbian.com/symbianos/index.asp>

- Mobile code security for MMS, SMS type data.

[http://www.cs.ucsb.edu/~vigna/publications/2005\\_felmetsger\\_vigna\\_ICECCS05.pdf](http://www.cs.ucsb.edu/~vigna/publications/2005_felmetsger_vigna_ICECCS05.pdf)

# What not to do....

## V. References

1. Two Fast Handover Solutions for the IMS Handover in the Presence of Mobile IPv6 by using Context Transfer Procedures  
Reza Farahbakhsh, Naser Movahhedinia  
, University of Isfahan  
, Computer Engineering Department  
, Isfahan, Iran
2. QoS-Conditionalized Handoff for Mobile IPv6 (2002)
3. QoS in Mobile Ipv6, Zhigang KAN, Dongmei ZHANG, Runtong ZHANG, Jian MA
4. Quality of Service in the IP Multimedia Subsystem, A. Hernández, M. Álvarez-Campana,  
E. Vázquez Departamento de Ingeniería de Sistemas Telemáticos, Universidad  
Politécnica de Madrid, E.T.S.I. de Telecomunicación, Av. Complutense, s/n, E-28040  
Madrid, Spain.{albertoh, mac, enrique}@dit.upm.es
5. Comparative Analysis of Network Layer and Application Layer IP Mobility Protocols for  
IPv6 Networks A. Dutta, S. Das, Telcordia Technologies, NJ, T. Chiba, H. Yokota, A. Idoue,  
KDDI Labs, Japan,K. D Wong, Malaysia University of Science and Technology,H.  
Schulzrinne, Columbia University, NY
6. <http://www.newport-networks.com/whitepapers/IMS-2.html>
7. [http://en.wikipedia.org/wiki/IP\\_Multimedia\\_Subsystem](http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem)

## A few Latex/Bibtex issues

- ◆ Don't start a sentence with \cite{}.
- ◆ Bibtex loose Capitalization except for first word of sentence. Put items in {} to keep case.
- ◆ Will not detect duplicated ref if the "tag" is different. If using "in order" you won't see it either so do a check by using alphabetical ordering.
- ◆ Google scholar not always correct, read it to check.
- ◆ Lots of standard ways to shorten when needed. Edit the bibtex .bib file, not the bbl.

- Machine Learning*, 2017, pp. 1224–1232.
- [12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, “Stackgan++: Realistic image synthesis with stacked generative adversarial networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1947–1962, 2018.
  - [13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- ▼ **BOOKTITLE= {{IEEETEXML.XCOMI.}} ON  
{Computer}{V}ision {{ICCV}}.}**

# **ORAL/COMPREHENSIVE EXAM AND THESIS PROPOSAL**



**Vision And Security Technology**  
University of Colorado **Colorado Springs**

- *The oral qualifying examination* is an oral presentation with a written report that surveys the literature in the planned research area that a student may pursue. A few example papers, recommended by the PhD program committee will be posted and available to students for reference. The examining committee will be organized by the advisor if identified by the student, or by the program director if the advisor has not been chosen by the student. The examining committee consists of three faculty members from the host department.

# ORAL EXAM

- The topic of the exam will be determined by the student's advisor in consultation with the examining committee. Committee is 3 people.
- Students having any sponsorship via the advisor's research funding or departmental teaching should pass the oral qualifying examination by the end of the first year after admission to the program. Other PhD students should pass the oral qualifying examination by the end of the second year after admission to the program. A student has one chance to take the oral exam.

# WAIVERS

- Waiver to the oral qualifying exam will be given if a student, as the first author, has one paper published or accepted for publication at a peer-reviewed international journal or technical conference with PUBLISHED selection rate <=40%. The request is initiated by the advisor.

- For CS the written examination for the Computer Science focus area consists of four required topics: Computer Architecture, Operating Systems, Automata, and Algorithms.
- For Security written examination consists of four distinct topics, Computer Communication, Fundamentals of Network & Computer Security, Applied Cryptography, and Homeland Security.
- The examination takes place twice a year for all students in May and December.

- A student can take the examination up to **two** times: Students only need to retake the failed topics in the subsequent attempt. Students having any sponsorship via the advisor's research funding or departmental teaching should pass all topics by the end of the second year after admission to the program. Other PhD students should pass all topics by the end of the third year after admission to the program

# PH.D IN CS COURSE WAIVER

- Waiver to the written qualifying examination will be given if the student has passed the required courses (CS 5200, CS 5500, CS 5700 and CS 5720) of the four qualifying examination topics at UCCS with a cumulative GPA of 3.75 or above. The program committee will decide if the grade of a transferred class can be used. Courses cannot be retaken to increase the GPA in order to qualify for the waiver.

# PHD IN SECURITY COURSE WAIVER

- Waiver to the written qualifying examination will be given if the student has passed the required courses (CS 5220, CS 5910, CS 5920 and CS 5950) of the four qualifying examination topics at UCCS with a cumulative GPA of 3.75 or better. The program committee will decide if the grade of a transferred class can be used. Courses may not be retaken in order to increase the GPA to qualify for the waiver.

- ***Thesis Proposal /Comprehensive Examination*** The purpose of the comprehensive examination is to ensure that the student possesses the following:
  - Sufficient grasp of the fundamentals of the chosen dissertation area to begin research, normally achieved through a thorough study of the current literature on the topic
  - Ability to conduct innovative research
  - Ability to exchange ideas and information with members of the Advisory Committee

# TIMELINE

- Expected proposal done by end of year 3.
- Max time for Ph.d. (without exceptions) 6 years. After that “course” start expiring and need grad school dean permission to continue.

- PhD students cannot enroll in more than 24 credit hours of dissertation prior to passing the comprehensive exam. Comprehension of existing literature and course material pertinent to the dissertation research, as well as the reasonableness of the unknown or undeveloped concepts that the student proposes, will be assessed by the Advisory Committee. The responsibility of the Advisory Committee is to review the research proposal and the qualifications of the student to complete the research successfully. If the research and the approach are found to be significant and appropriate and the student is judged capable of completing the research, the Advisory Committee will approve the research direction.

- Grad School: Must complete 15 credits after proposal is accepted.
- If the Advisory Committee does not find the student ready to begin dissertation research, it must suggest further preparation by the student and plan on a subsequent taking of the comprehensive examination.
- A passing grade in the examination is given if at least four of the five members of the public Committee including the student's advisor vote affirmatively.

# WHAT IS IN THE PROPOSAL

## Parts of a Proposal

- Title
- Abstract
- Introduction
- Background
- Overall Research Theme
- Research Question1
- Research Question2
- Research Question3
- ...
- Bibliography

## For Each Research Question:

- Review of Literature
- Significance/Implications
- Preliminary work & Plan of Work
- Possible Methodology
- Plan for Result Validation
- Anticipated Timeline

# A TYPICAL PROCESS

- Decide a general area of interest with your advisor
- Initial literature review to narrow down and sharpen the problem definitions
- For a Ph.D. thesis, usually multiple related problems under one broader theme, so find multiple problem
- Write the problem statements
- Justify potential innovation
- Design methods

# A TYPICAL PROCESS (CONTINUED)

- Experimental Design (or Theoretical proof?)
- Plan for validation/evaluation
- Estimate a timeline
  
- Having published papers adds significant credibility to your claims about ability to do research and that your problems are solveable.

# WHAT PROBLEM TO EXPLORE/PROPOSE



# THESIS TIMELINE

- Work with your advisor on the timeline.
- Check important dates for submitting and defending dissertations.
- Take various factors into consideration: IRB approval, travel; design, testing, and length of experiments; negotiation of entry into the study site; purchase of necessary equipment; drafting; redrafting).
- Timeline is not always respected but writing it helps you to avoid potential pitfalls
- Do a Gantt chart.

# MY “PERSONAL” ADVICE/APPROACH

- Year 1: If you can, Go big try for Tier 1 level problems/papers. But either way try to publish something.
- Year 2: Reassess where you are, go for Tier 2/3 to get 1-2 papers out (reworking paper from year1. Maybe new idea for Tier1).
- Year 2 or 3, do thesis proposal & 1-2 more papers
- Year 3 or 4, graduate. & 1-2 more papers

# Introduction To GIT

Based in part on slides from tboult Ling and others



University of Colorado  
Bachelor of Innovation™  
University of Colorado Colorado Springs

# Key Git Files/Directories

- `~/.gitconfig`
- `.git`
  - In top level of repository
  - Contains all objects, commits, configuration, for project
  - `.git/config` has project specific configurations
- `.gitignore`
  - Stored in directory for ignoring

# Git Advantages

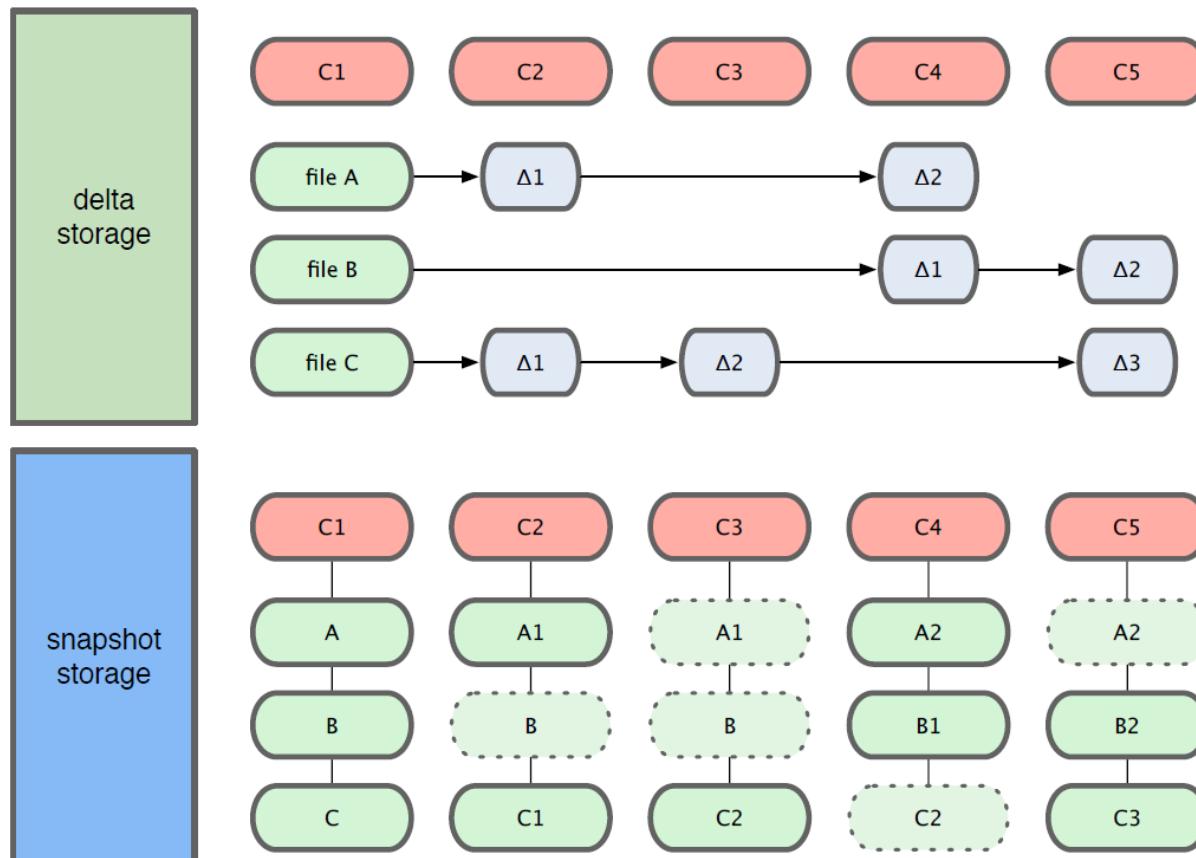
- Resilience
  - No one repository has more data than any other
- Speed
  - Very fast operations compared to other VCS (source-safe, CVS and Subversion)
- Space
  - Compression can be done across repository not just per file
  - Minimizes local size as well as push/pull data transfers
- Simplicity
  - Object model is very simple
- Large userbase with robust tools

# Some GIT Disadvantages

- Definite learning curve, especially for those used to centralized systems
  - Can sometimes seem overwhelming to learn
    - Conceptual difference
    - Huge amount of commands

# Getting Started

- Git uses snapshot storage (logical view – it does use deltas in packs)



# Working With Git

- `echo "hello" >> hello.txt`
- `git diff`
  - Shows changes we have made
- `git status`
  - Shows list of modified files
- `git add hello.txt`
- `git diff`
  - No changes shown as diff compares to the index
- `git diff HEAD`
  - Now can see the changes in working version
- `git status`
- `git commit -m 'Second commit'`

# Getting Started

## Three trees of Git

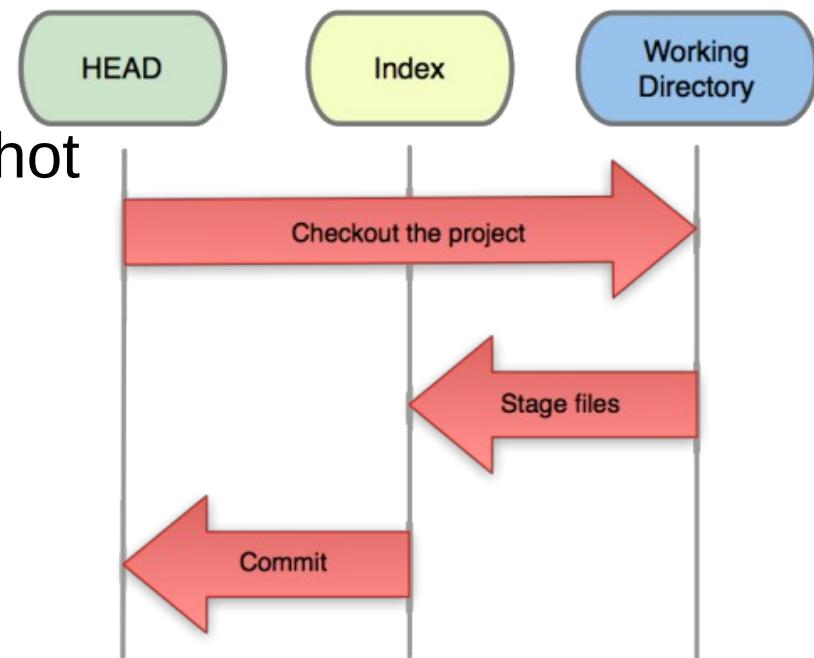
- The HEAD
  - last commit snapshot, next parent

- Index

- Proposed next commit snapshot

- Working directory

- Sandbox



# Getting Started

## ■ A basic workflow

- (Possible init or clone) Init a repo
- Edit files
- Stage the changes
- Review your changes
- Commit the changes

# Getting Started

## Init a repository

```
tboult:gittest-> git init
Initialized empty Git repository in /home/tboult/gitdemo/.git/
```

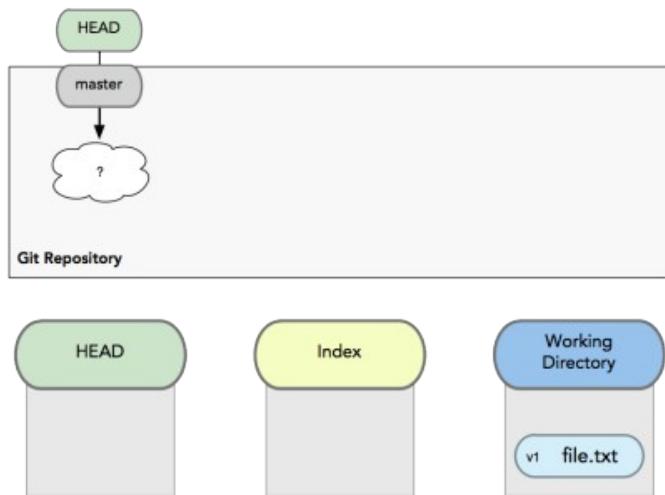
```
tboult:gittest-> ls -l .git/
total 32
drwxr-xr-x 2 tboult tboult 4096 2011-08-28 14:51 branches
-rw-r--r-- 1 tboult tboult    92 2011-08-28 14:51 config
-rw-r--r-- 1 tboult tboult   73 2011-08-28 14:51 description
-rw-r--r-- 1 tboult tboult   23 2011-08-28 14:51 HEAD
drwxr-xr-x 2 tboult tboult 4096 2011-08-28 14:51 hooks
drwxr-xr-x 2 tboult tboult 4096 2011-08-28 14:51 info
drwxr-xr-x 4 tboult tboult 4096 2011-08-28 14:51 objects
drwxr-xr-x 4 tboult tboult 4096 2011-08-28 14:51 refs
```

## Git init

# Getting Started

## ■ A basic workflow

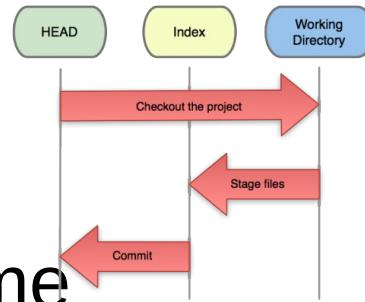
- Edit files
- Stage the changes
- Review your changes
- Commit the changes



## ■ Use your favorite editor

### ■ E.g:

- Visual Studio
- Eclipse
- Vscode
- Atom
- Emacs
- VIM
- Sublime
- NotePad++
- <https://help.github.com/articles/associating-text-editors-with-git/#platform-windows>

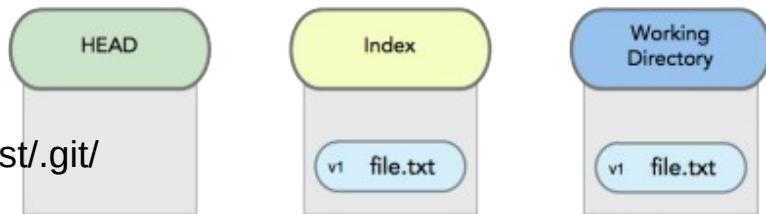
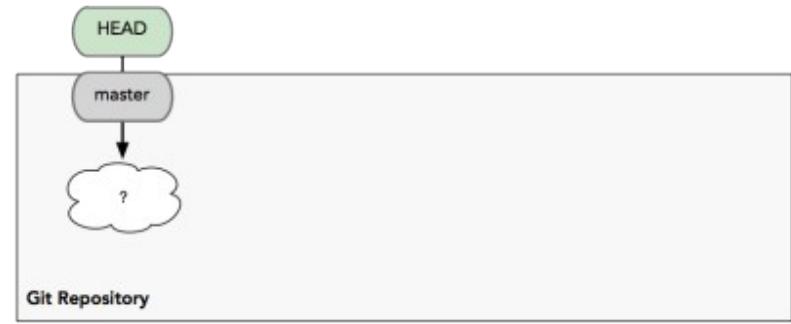


# Getting Started

## A basic workflow

- Edit files
- Stage the changes
- Review your changes
- Commit the changes

## Git add filename



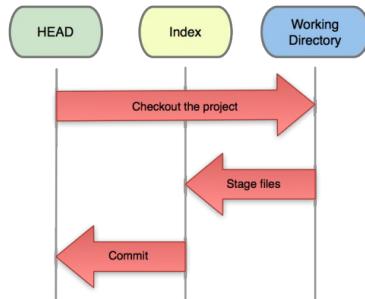
```

tboult:gittest->git init
Initialized empty Git repository in /Users/tboult/CS6000/gittest/.git/
tboult:gittest->echo "hello" > hello.txt
tboult:gittest->git status
On branch master
Initial commit
Untracked files:
  (use "git add <file>..." to include in what will be committed)

      hello.txt

```

nothing added to commit but untracked files present (use "git add" to track)  
tboult:gittest->



# Getting Started

## A basic workflow

- Edit files
- Stage the changes
- Review your changes
- Commit the changes

```
tboult:gittest->git add hello.txt
```

```
tboult:gittest->git status
```

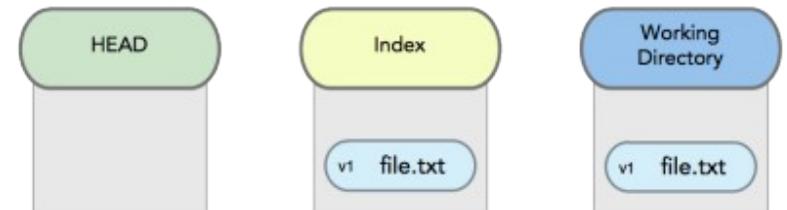
On branch master

Initial commit

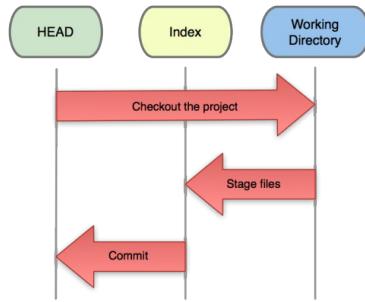
Changes to be committed:  
(use "git rm --cached <file>..." to unstage)

new file: hello.txt

## Git status



```
tboult:gittest->
```

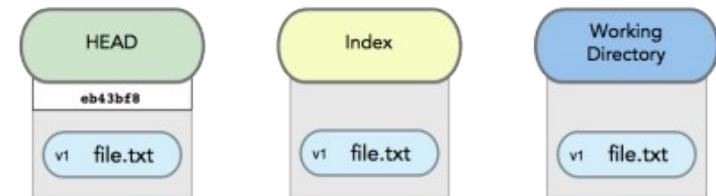


# Getting Started

## A basic workflow

- Edit files
- Stage the changes
- Review your changes
- Commit the changes

## Git commit

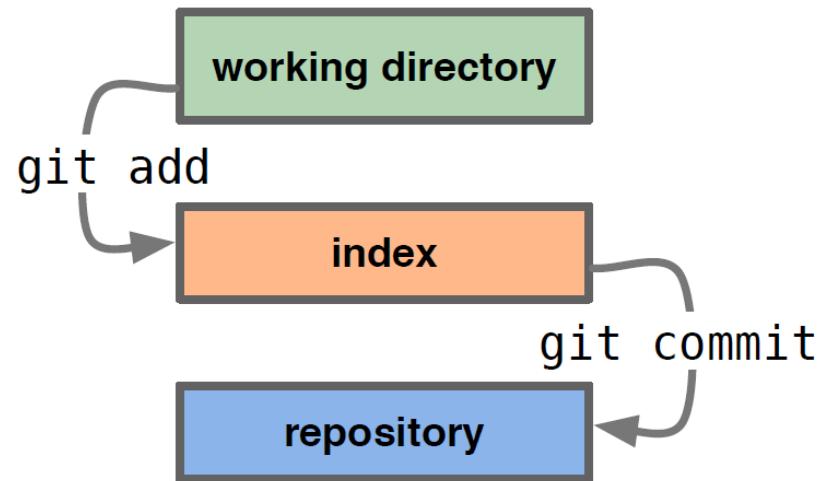


**git commit**

```
tboult:gittest->git commit -am "demo message"
[master (root-commit) ff0f799] demo message
 1 file changed, 1 insertion(+)
 create mode 100644 hello.txt
tboult:gittest->
```

# Getting Started

- A basic workflow
  - Edit files
  - Stage the changes
  - Review your changes
  - Commit the changes



# Getting Started

- View changes
- Git diff
  - Show the difference between working directory and staged
- Git diff --cached
  - Show the difference between staged and the HEAD

- View history

- Git log

```
toul:git>echo "hello2" > hello.txt  
toul:git>git commit -am "adding second line"  
[master 03e952e] adding second line  
1 file changed, 1 insertion(+), 1 deletion(-)
```

```
toul:git>git log  
WARNING: terminal is not fully functional  
- (press RETURN)  
commit  
03e952e99190d4a4d2cad12504a5474921419daf  
Author: Terry Boult <toul@vast.uccs.edu>  
Date: Sat Aug 18 12:53:33 2018 -0600  
  
        adding second line
```

```
commit ff0f799d1115ab1306dad78893fa1be3bee91af5  
Author: Terry Boult <toul@vast.uccs.edu>  
Date: Sat Aug 18 12:52:22 2018 -0600
```

demo message  
toul:git>toul:git> 623

# Revert changes (Get back to a previous version)

## • and branches

```
tboult:gittest->git checkout ff0f -b newbranch
Switched to a new branch 'newbranch'
tboult:gittest->
tboult:gittest->cat hello.txt
hello
tboult:gittest->echo "Branc lin" >>hello.txt
tboult:gittest->git commit -am "commit on branch"
[newbranch 5dfa020] commit on branch
 1 file changed, 1 insertion(+)
tboult:gittest->cat hello.txt
hello
Branc lin
tboult:gittest->git checkout ff0f
Note: checking out 'ff0f'.
```

You are in 'detached HEAD' state. You can look around, make experimental changes and commit them, and you can discard any commits you make in this state without impacting any branches by performing another checkout.

If you want to create a new branch to retain commits you create, you may do so (now or later) by using -b with the checkout command again. Example:

```
git checkout -b <new-branch-name>
```

```
HEAD is now at ff0f799... demo message
tboult:gittest->cat hello.txt
hello
tboult:gittest->git checkout 03e9
Previous HEAD position was ff0f799... demo message
HEAD is now at 03e952e... adding second line
tboult:gittest->cat hello.txt
hello2
tboult:gittest->
```

# Undoing What is Done

- **git checkout**
  - Used to checkout a specific version/branch of the tree
- **git reset**
  - Moves the tree back to a certain specified version
  - Use the --force to ignore working changes
- **git revert**
  - Reverts a commit
  - Does not delete the commit object, just applies a patch
  - Reverts can themselves be reverted!
- **Git never deletes a commit object**
  - It is very hard to shoot yourself in the foot!

# Branching

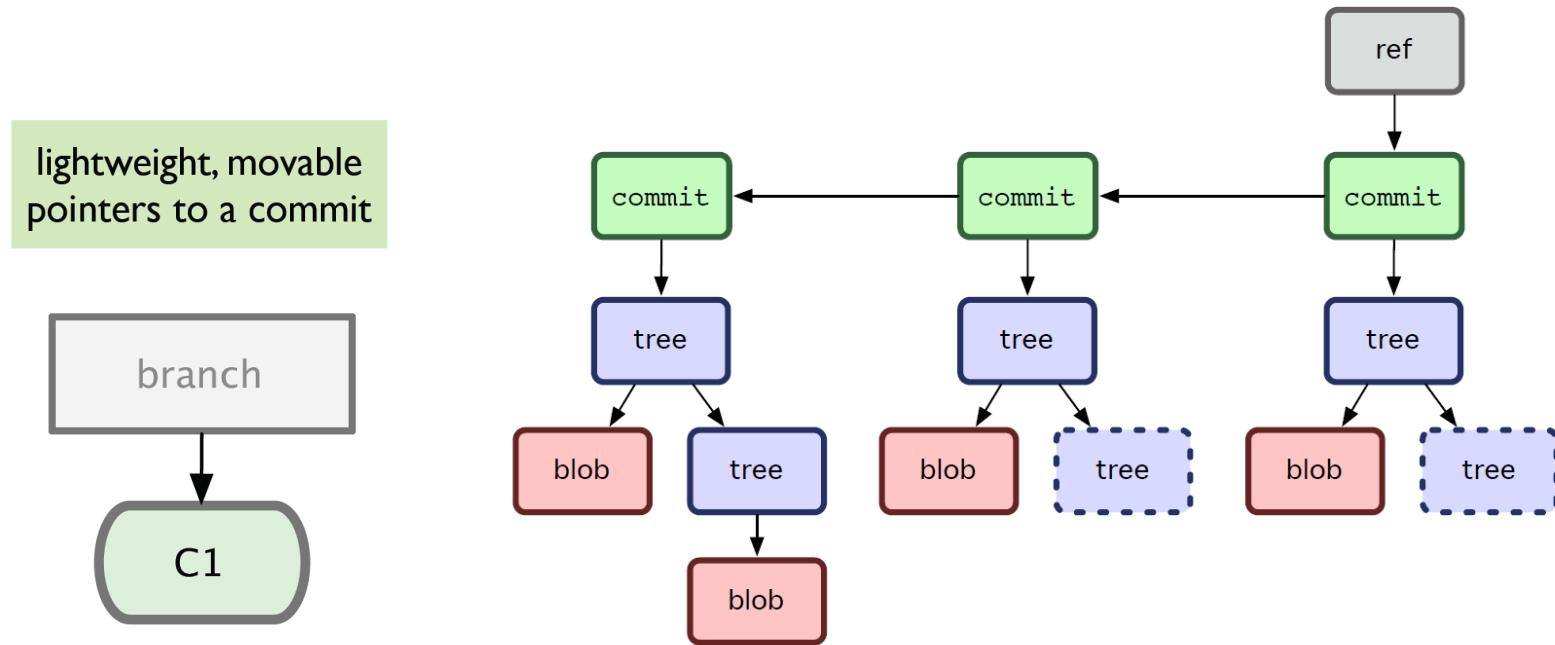
- Git branching is lightweight
  - No massive copying a la CVS/Subversion
  - Tools for helping merge branches and changes easily
- You are ALWAYS on a branch
- Branches can be local or remote
- Key commands
  - git branch
  - git merge
  - git cherry-pick
    - Allows you to choose specific commits to apply
    - You can edit the commits while cherry picking

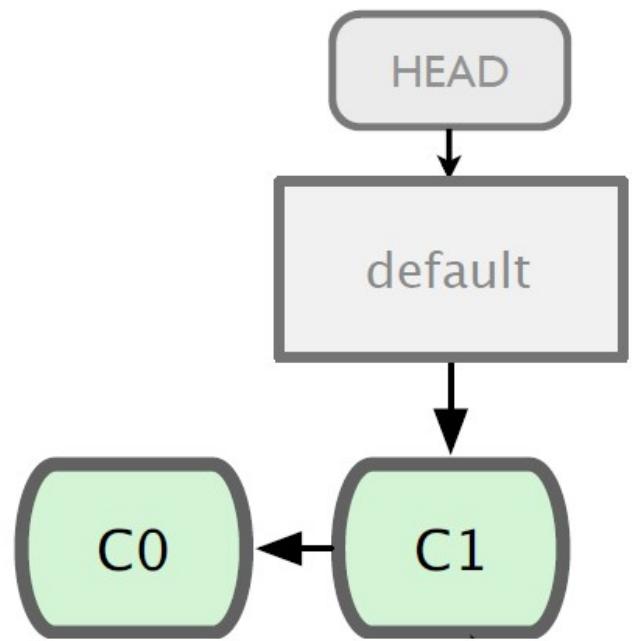
# Using Branches

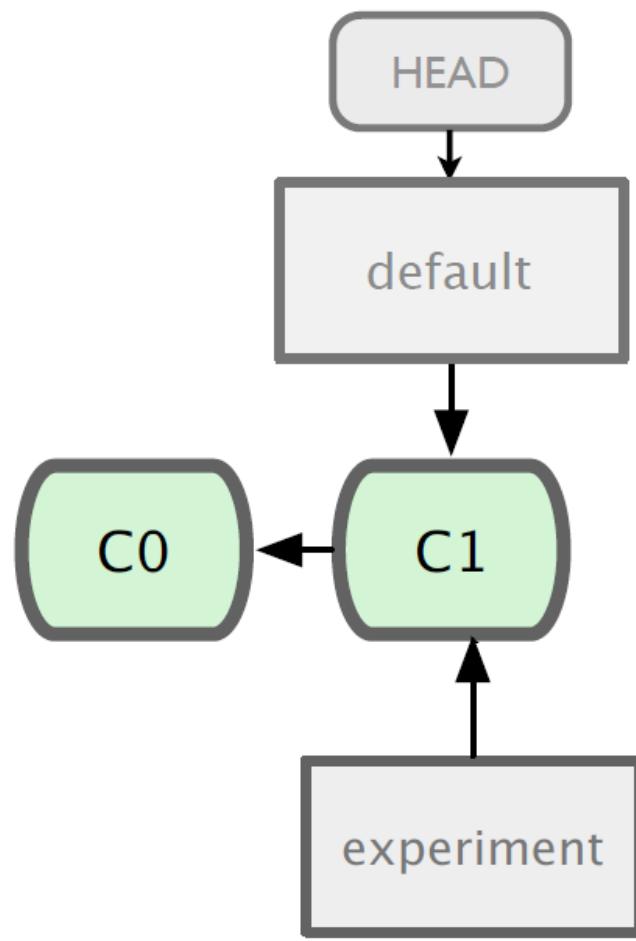
- `git checkout -b branch`
- `git checkout -b devel/branch`
- `git branch`
  - Lists all local branches available
- We can now make changes in one branch and propagate change using
  - `git merge`
  - `git cherry-pick`

# Branching

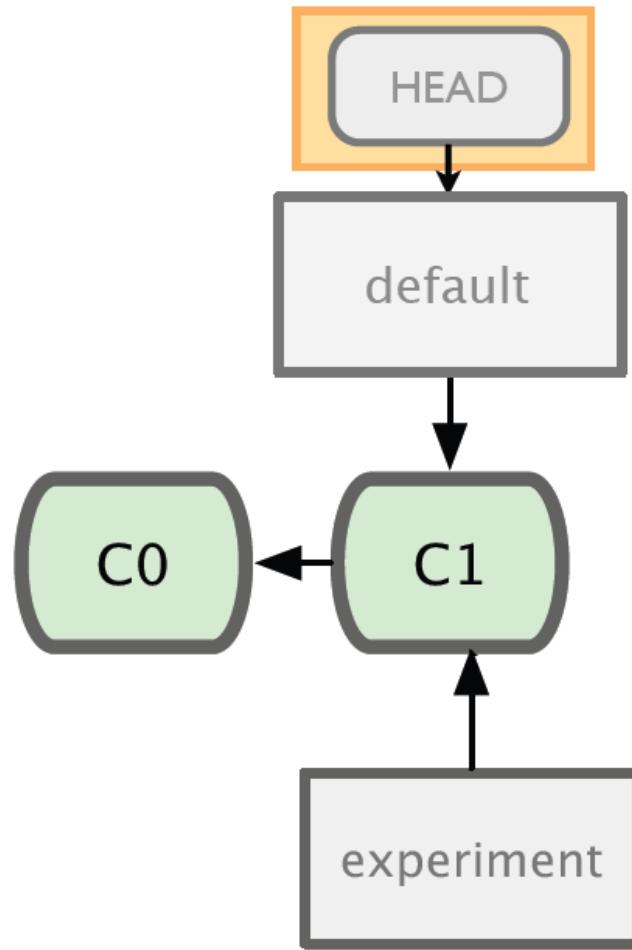
- Git sees commit this way...
- Branch annotates which commit we are working on



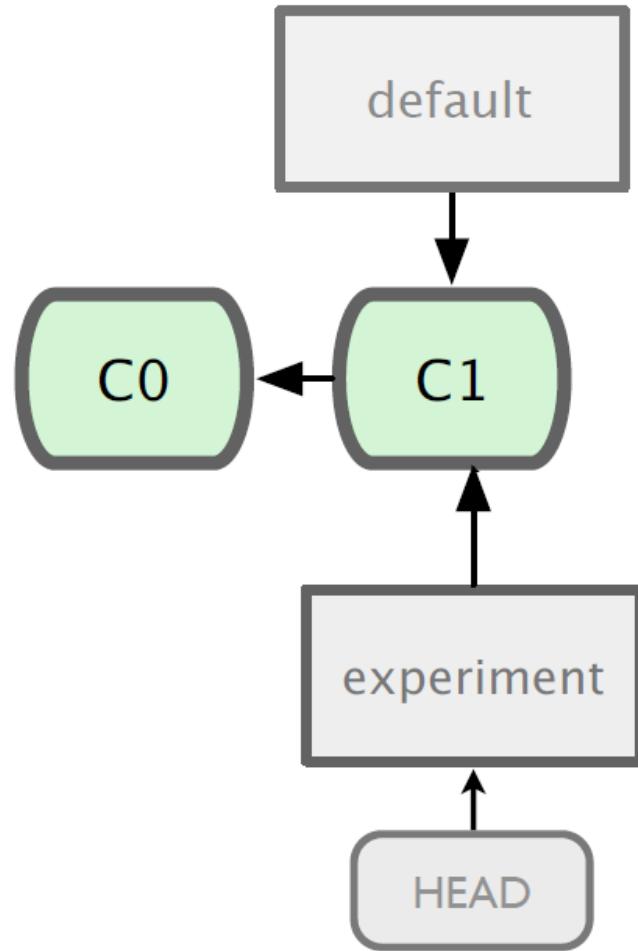




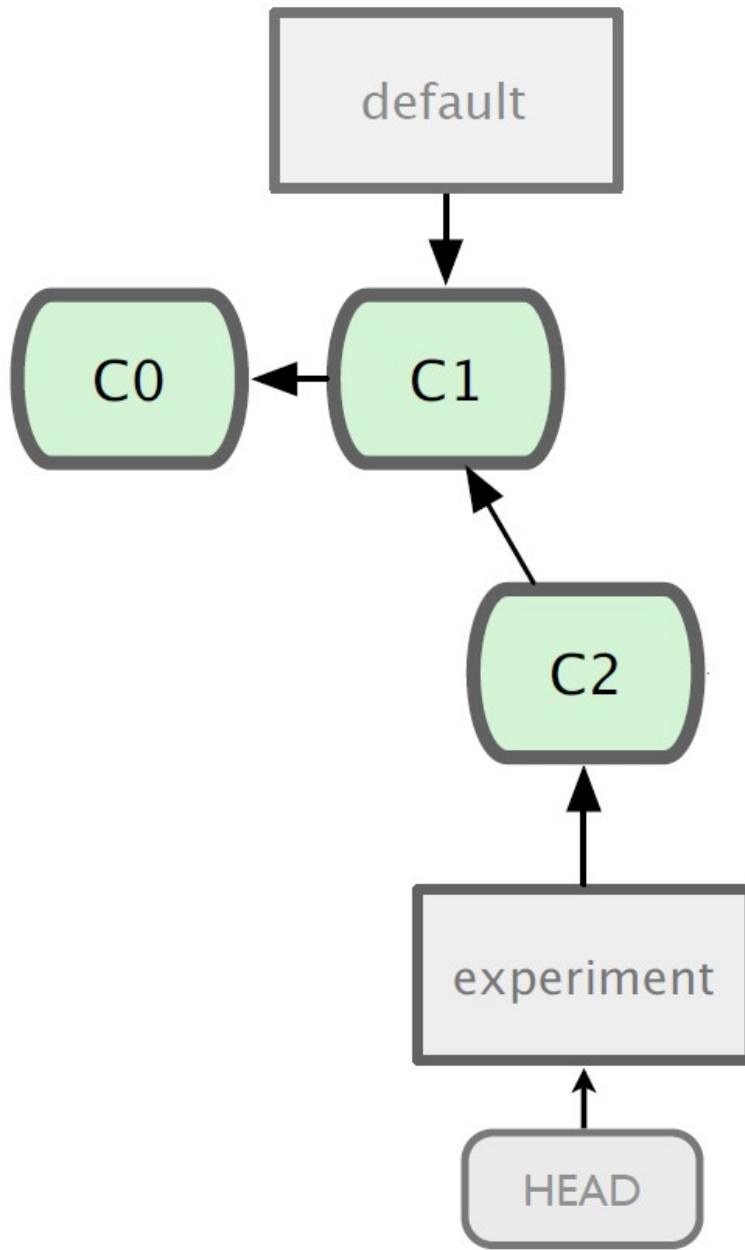
**git branch experiment**



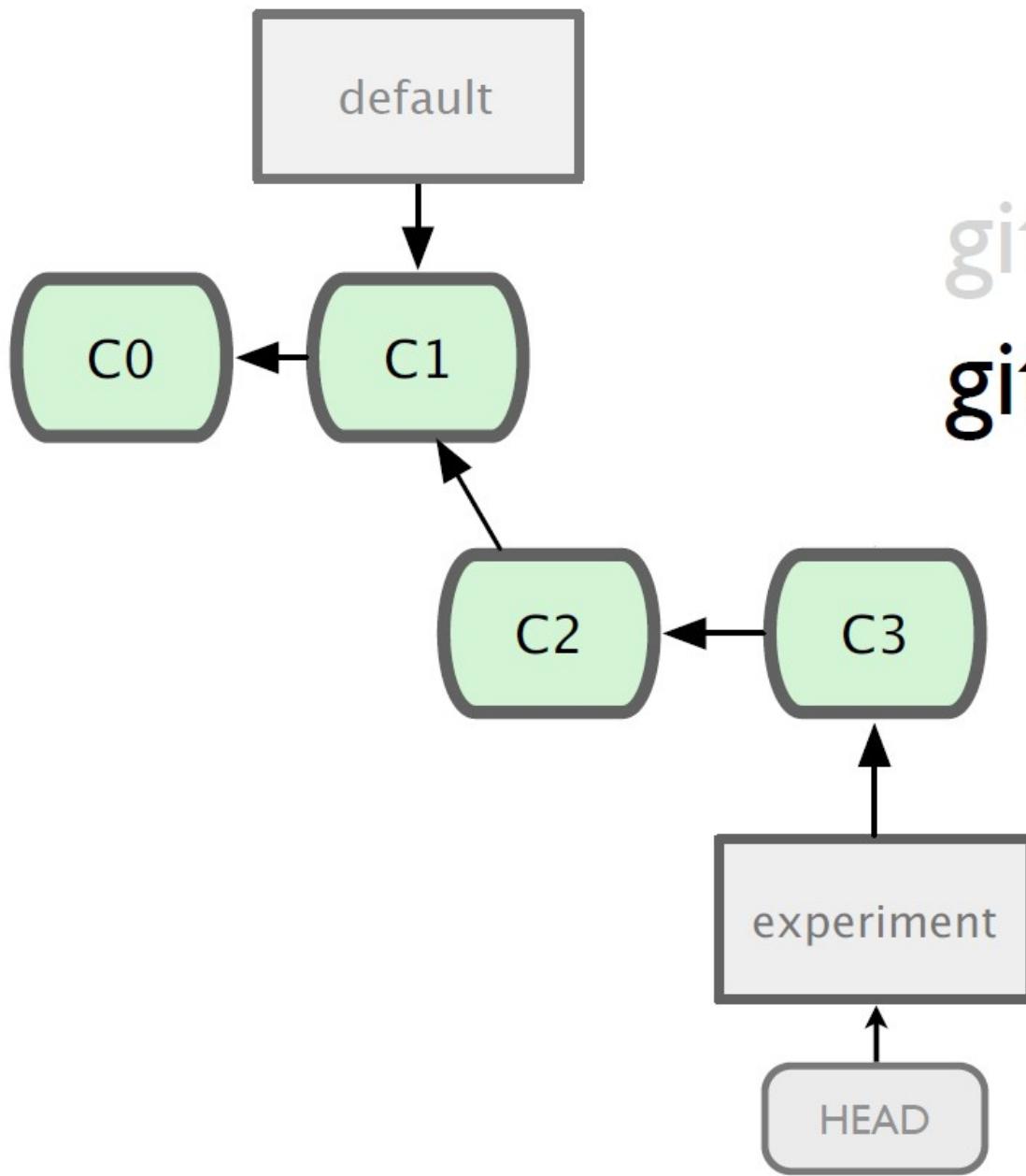
```
$ git branch  
* default  
  experiment
```



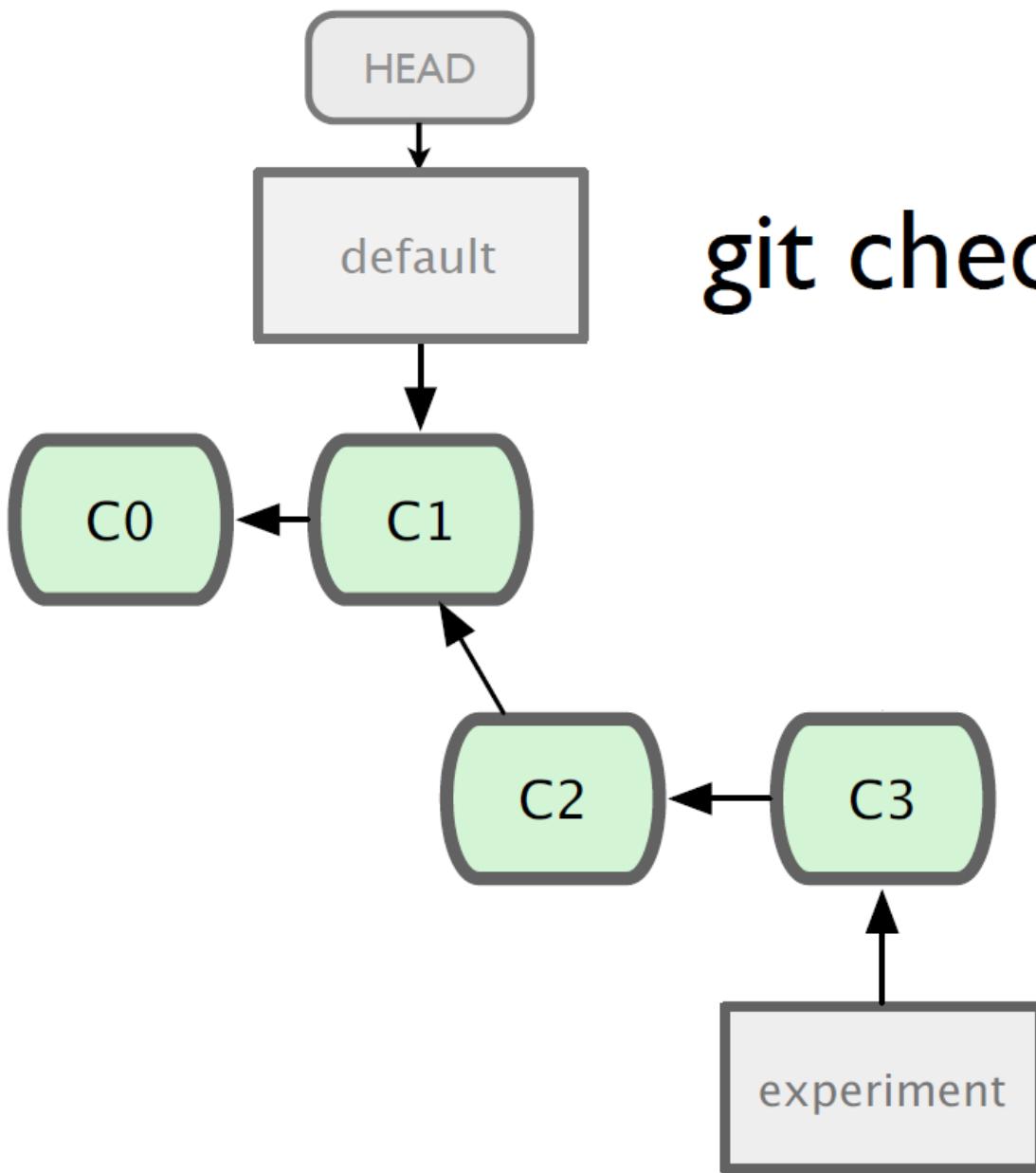
**git checkout experiment**



git commit

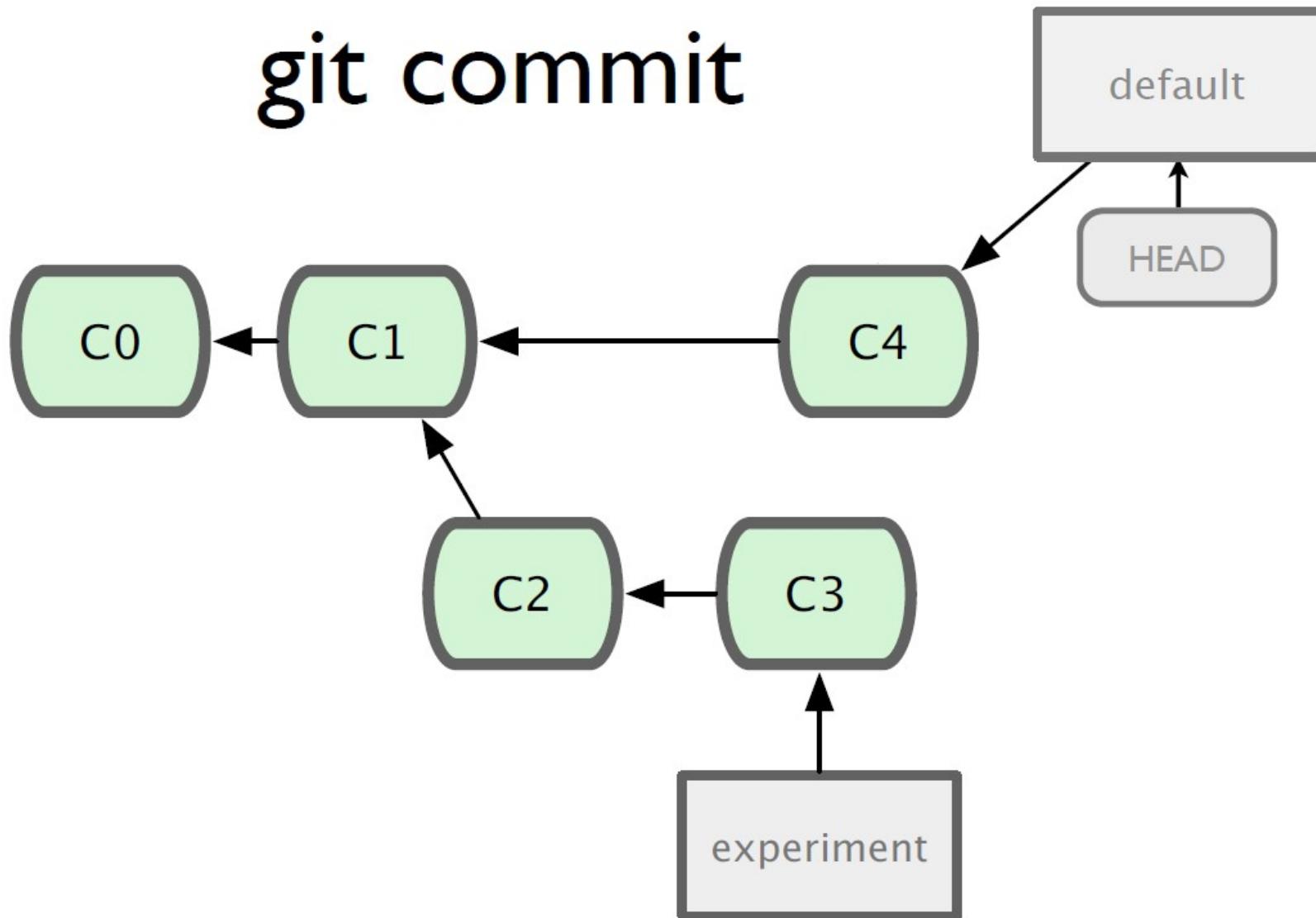


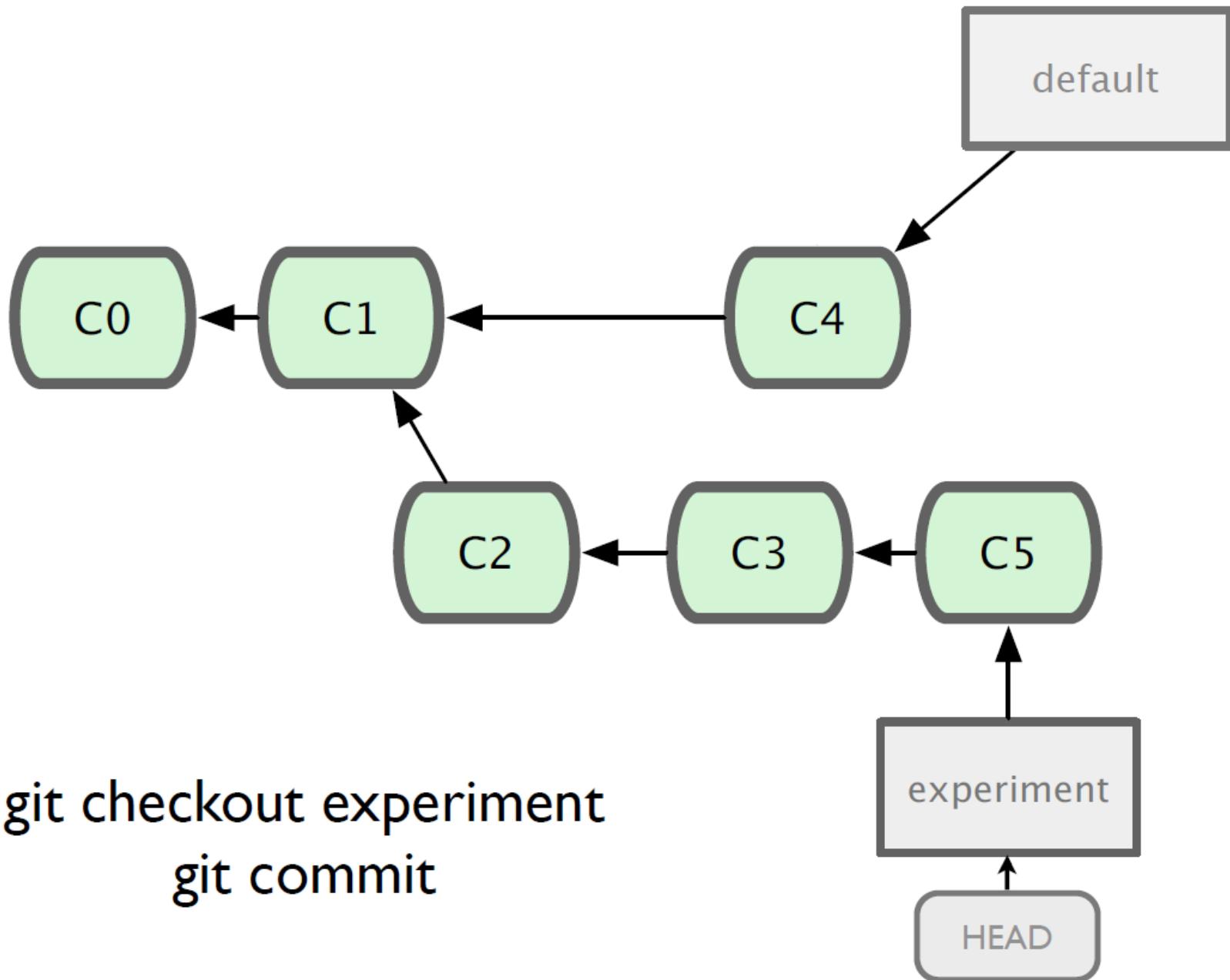
git commit  
git commit



# git checkout default

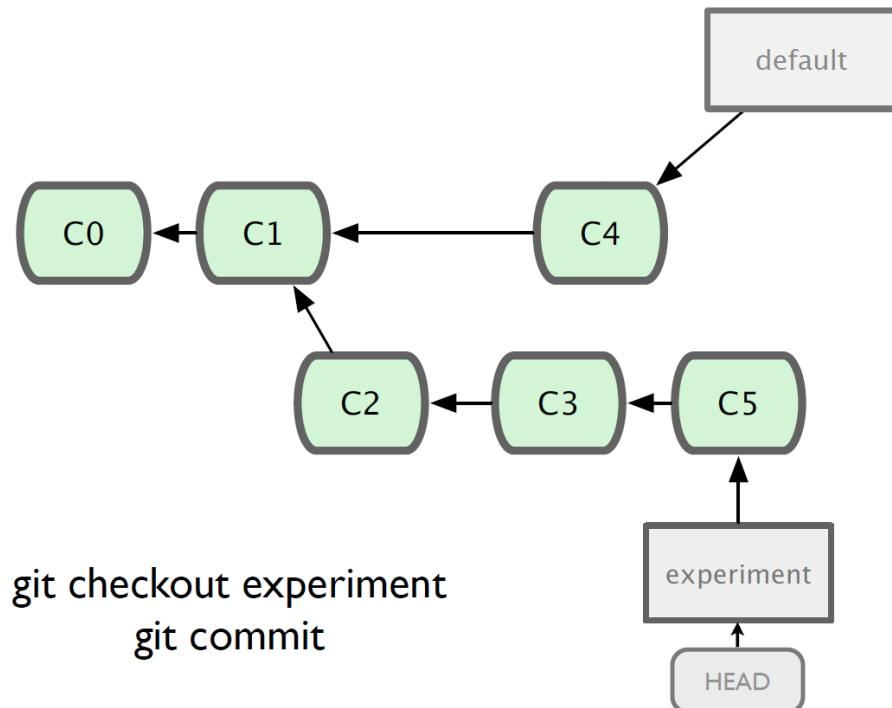
# git commit





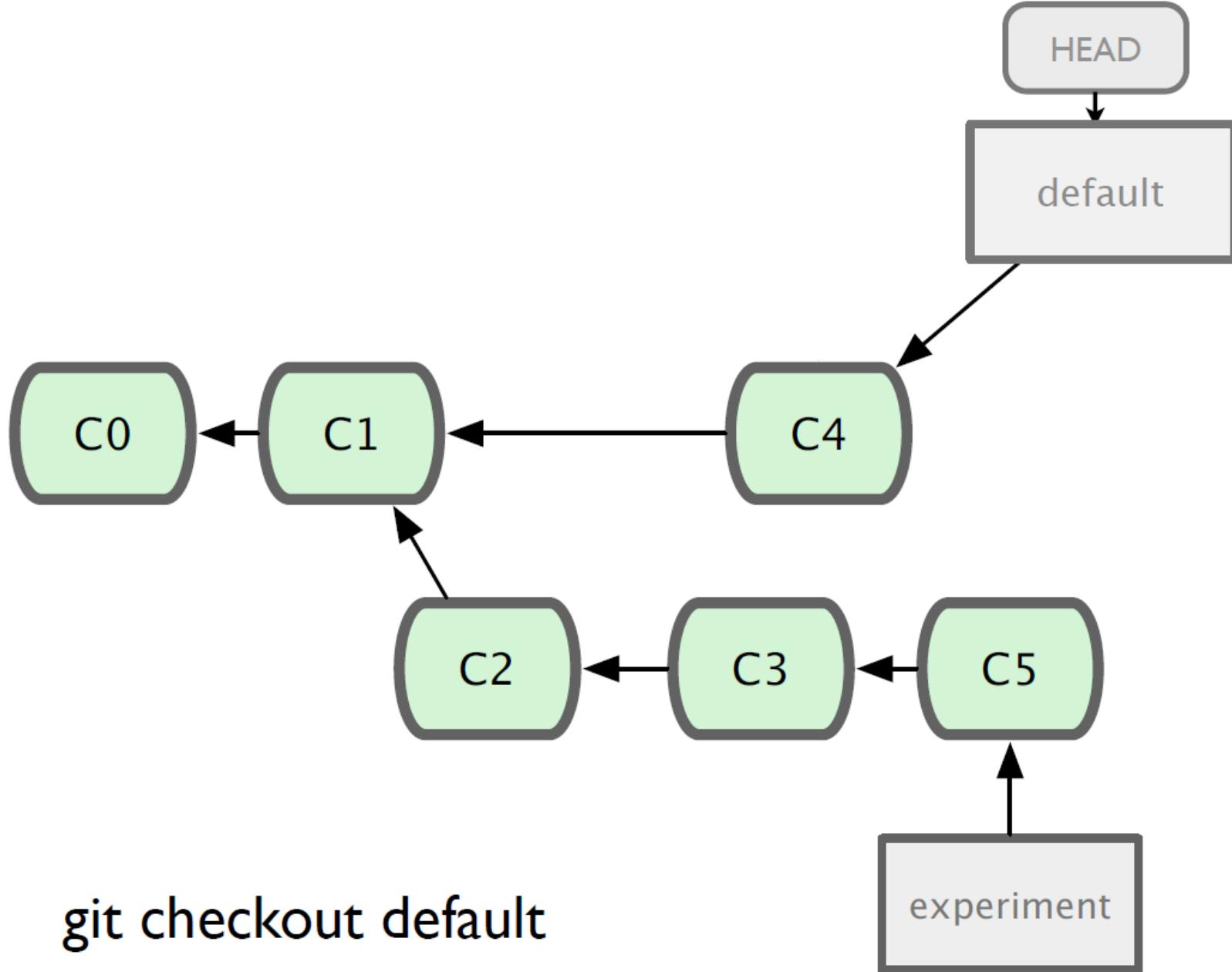
# Merging

- What do we do with this mess?
  - Merge them

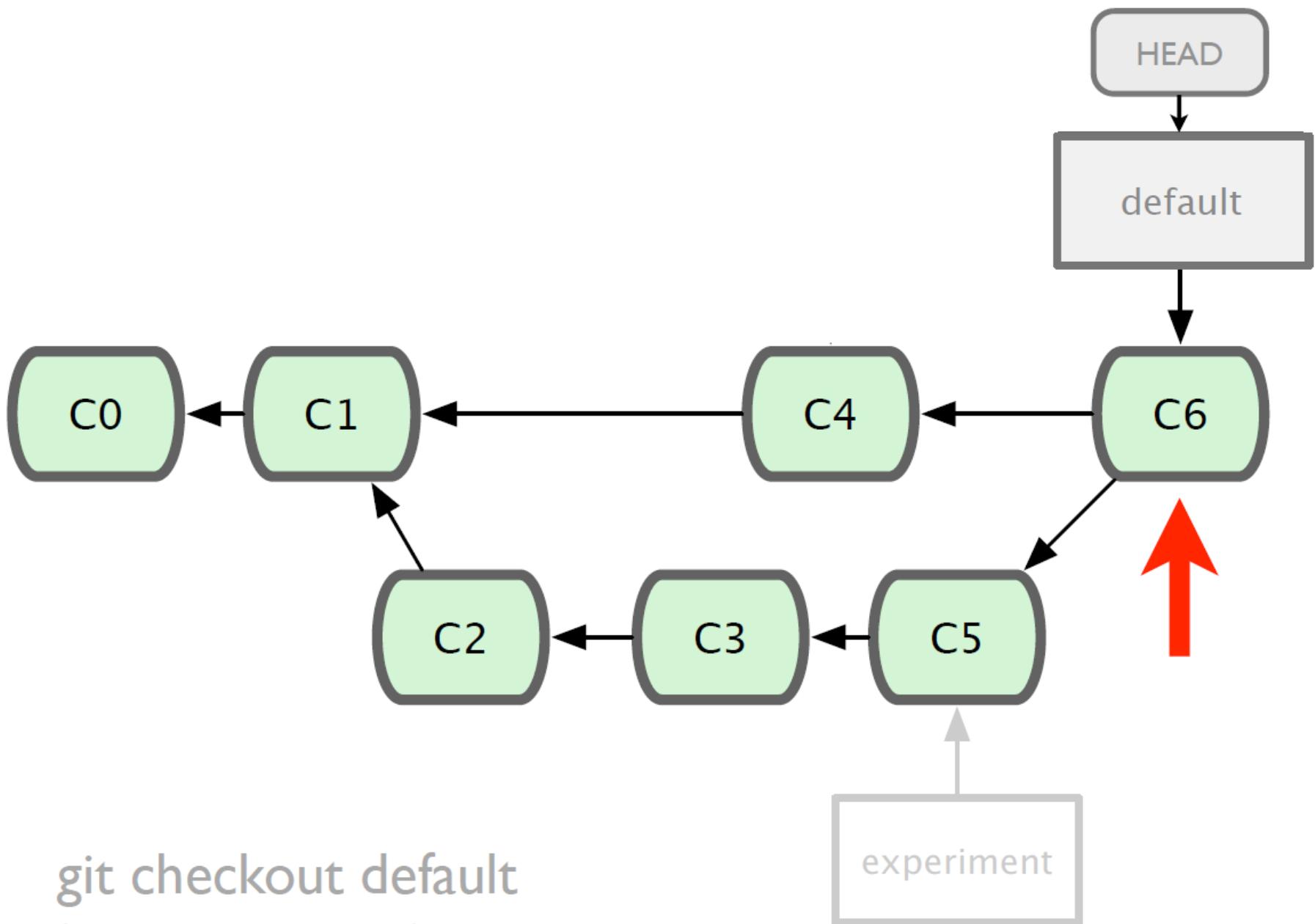


# Merging

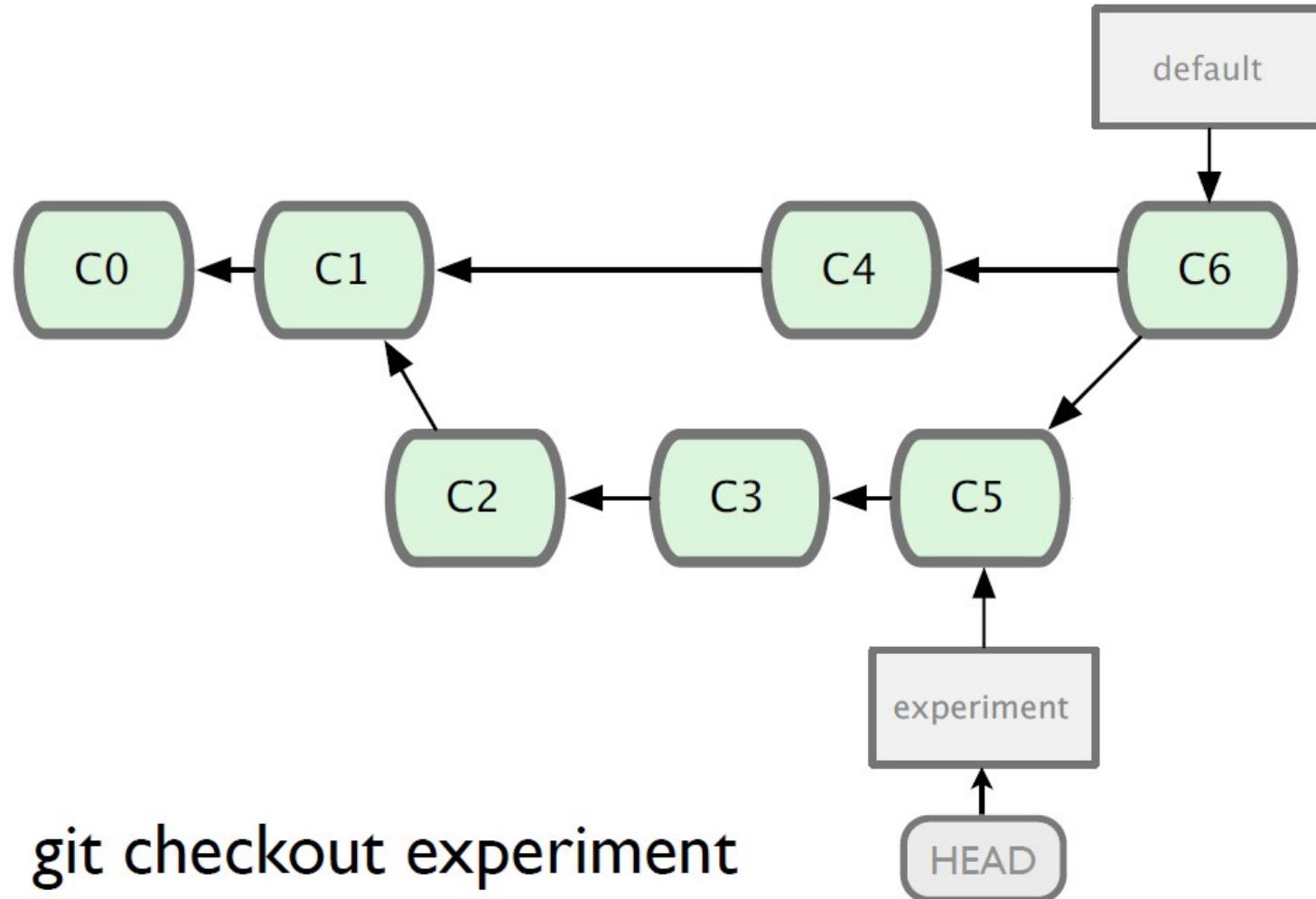
- Steps to merge two branch
  - Checkout the branch you want to merge **onto**
  - Merge the branch you want to merge

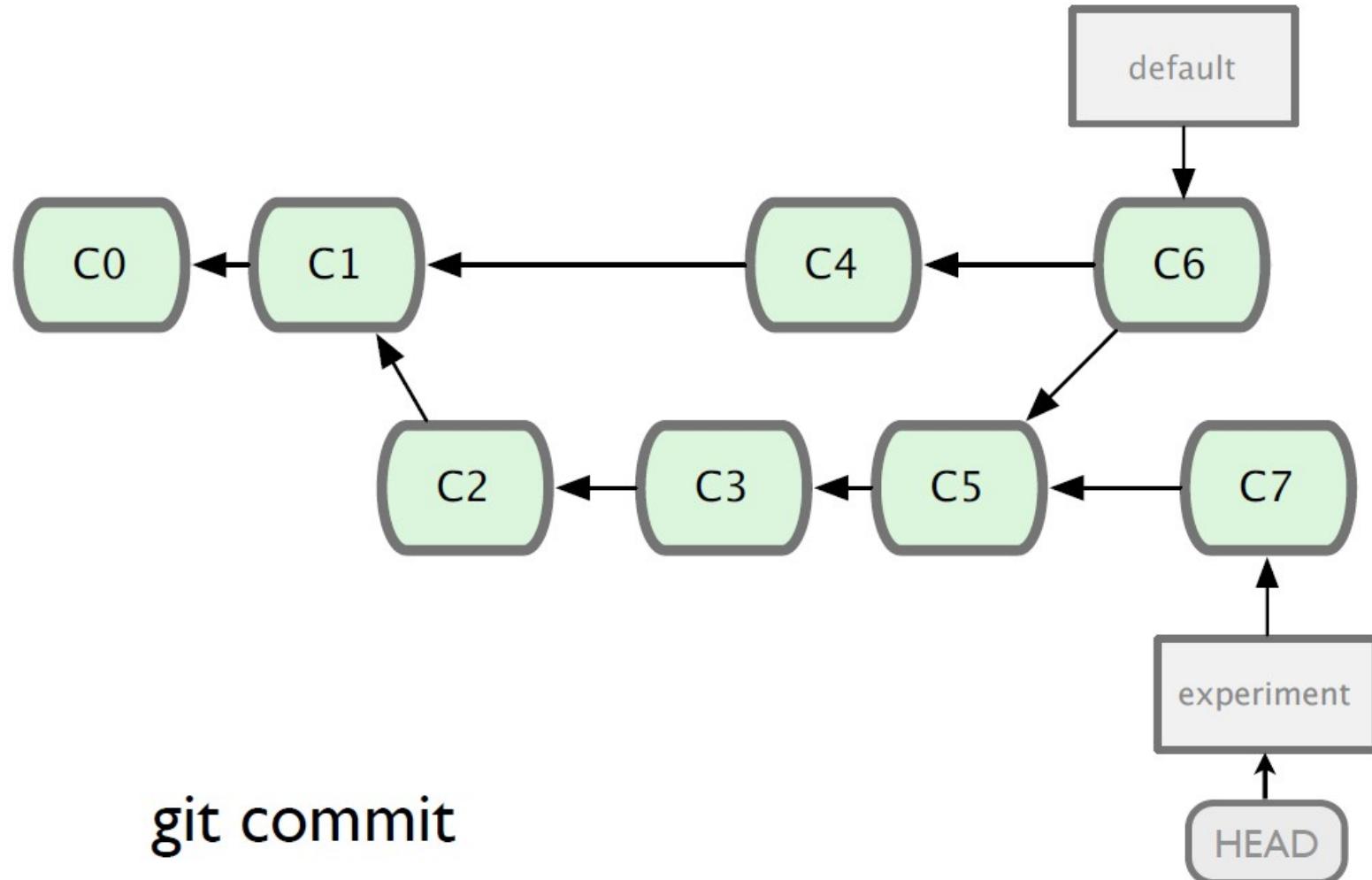


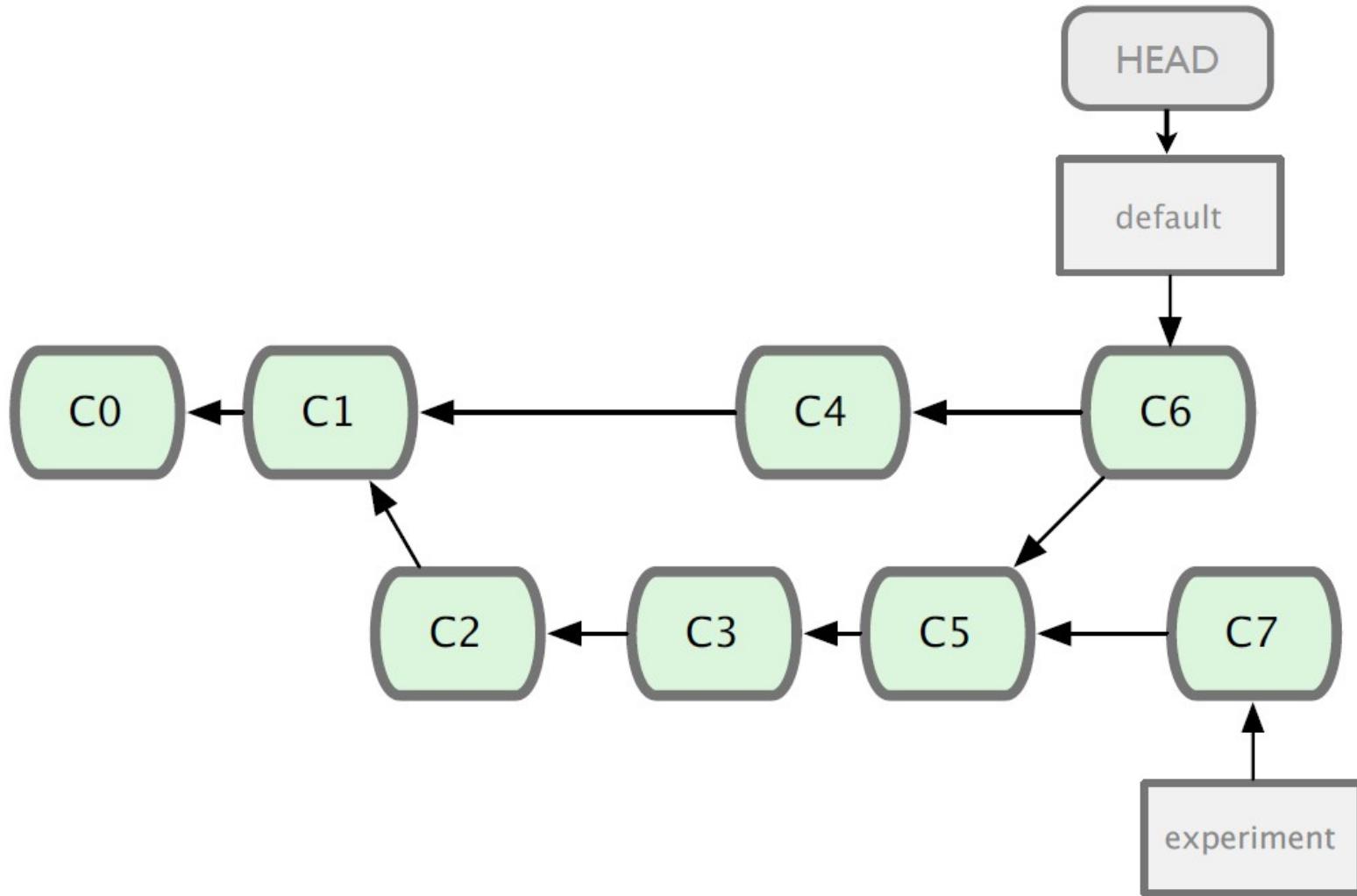
**git checkout default**



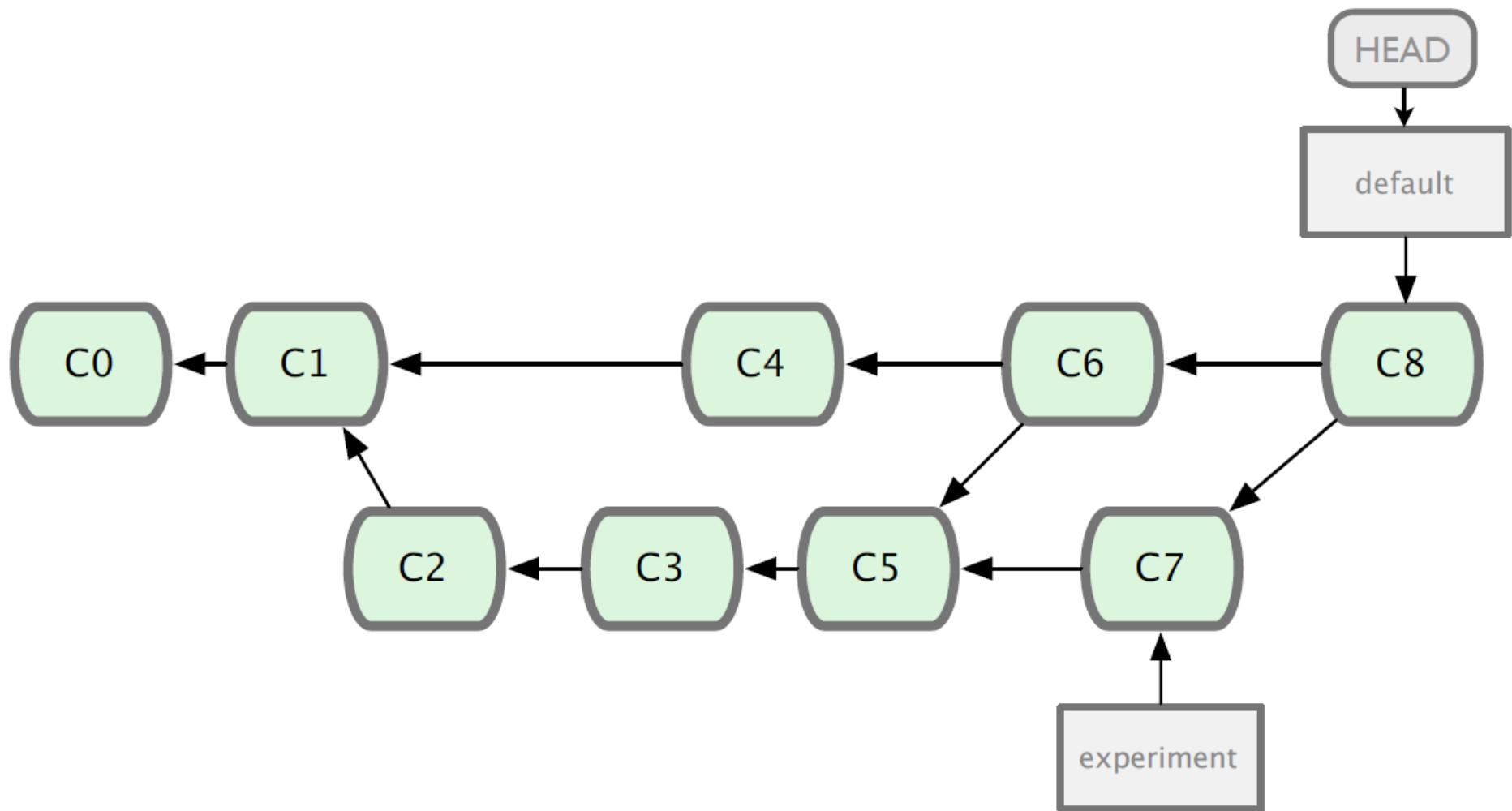
`git checkout default`  
`git merge experiment`







`git checkout default`



git merge experiment

# Branching and Merging

## Why this is cool?

- Non-linear development

```
clone the code that is in production
create a branch for issue #53 (iss53)
work for 10 minutes
someone asks for a hotfix for issue #102
checkout 'production'
create a branch (iss102)
fix the issue
checkout 'production', merge 'iss102'
push 'production'
checkout 'iss53' and keep working
```

# Git and Tagging

- Tags are just human readable shortcuts for hashes
- Branches can be made from any commit
- *git tag <tag-name>*

# git submodule

- Suppose your software has dependencies.
- And that those dependencies are also held in git repositories.
- You can add informations about them to your program.

# git submodule

## How to initialize it:

- `git submodule add git://repo/path/to/project.git`
  - This creates a directory called ‘project’
- `cd project; git checkout <version>`
  - Check out the exact version you wish to use
- `git commit`

# git submodule

- When cloning:
  - git clone <main repo>
    - As usual
  - git submodule init
    - Updates the submodules to the correct checkout version as set by the origin repository.
    - This *\*will\** require connectivity to the submodule repositories.

# **git submodule -- notes**

- Inside the submodules, you do not get a branch.
  - I.e: commits in there are lost
- Whenever you do a ‘git pull’ you should also do a ‘git submodule update’

# Working with remote

- Use git clone to replicate repository
- Get changes with
  - git fetch
  - git pull (fetches and merges)
  - Git commit –am “after merge”
- Propagate changes with
  - git push
- Protocols
  - Local filesystem (file://)
  - SSH (ssh://)
  - HTTP (http:// https://)
  - Git protocol (git://)

# Working with remote Local filesystem

- Pros
  - Simple
  - Support existing access control
  - NFS enabled
  
- Cons
  - Public share is difficult to set up
  - Slow on top of NFS

# Working with remote SSH

## Pros

- Support authenticated write access
- Easy to set up as most system provide ssh toolsets
- Fast
  - Compression before transfer

## Cons

- No anonymous access
  - Not even for read access

# Working with remote GIT

## Pros

- Fastest protocol
- Allow public anonymous access

## Cons

- Lack of authentication
- Difficult to set up
- Use port 9418
  - Not standard port
  - Can be blocked

# Working with remote HTTP/HTTPS

## Pros

- Very easy to set up
- Unlikely to be blocked
  - Using standard port

## Cons

- Inefficient

# Working with remote

- One person project
  - Local repo is enough
  - No need to bother with remote
- Small team project
  - SSH write access for a few core developers
  - GIT public read access

# Working with remote

- Use git remote add to add an remote repository

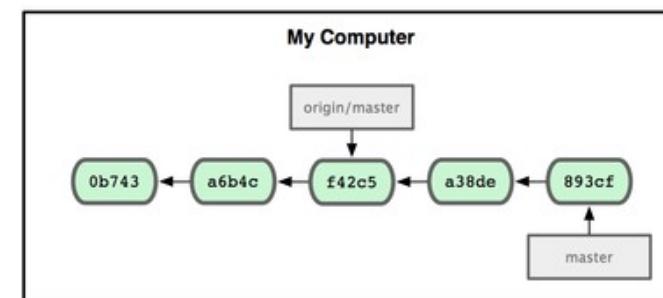
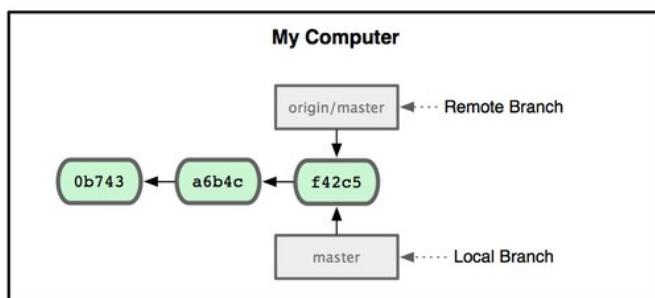
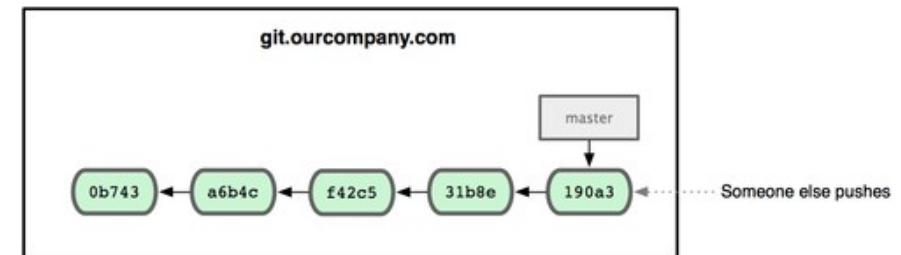
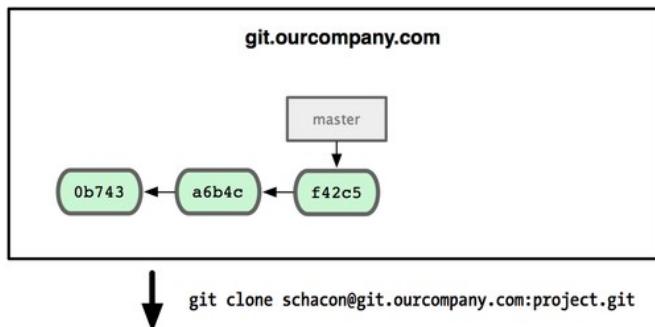


```
Git remote add origin git@github.com:FreezingGod/vimcfg.git  
tboult@tboult-desktop:~/vim_runtime$ git remote  
origin
```

# Working with remote

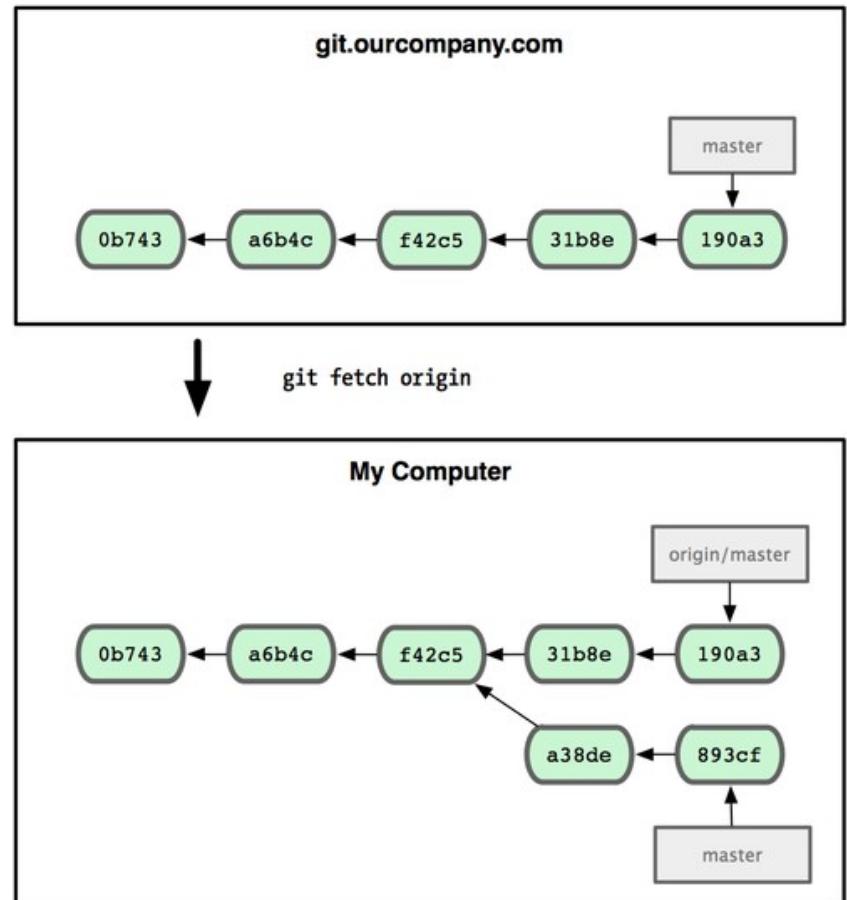
## Remote branching

- Branch on remote are different from local branch



# Working with remote

- Remote branching
  - Branch on remote are different from local branch
  - Git fetch origin to get remote changes
  - Git pull origin try to fetch reomte changes and merge it onto current branch



# Cloning our Repository

- *git clone first-git-repo*
  - Now have a full git repository to work with
- Changes are pushed back with *git push*
  - Pushing changes WILL NOT change working copy on the repository being worked on
- Branches can be based off of remote branches
  - *git branch --track new-branch remote/branch*
- Remote configuration information stored in *.git/config*
  - Can have multiple remote backends!

# Git for Software Versioning

- Create convention to define default server
- Developers clone from central server
- Lots of tools for transmitting patches between developers
- Being used for
  - Linux (obviously)
  - Ruby On Rails
  - Check out <http://github.com> for a variety of hosted projects

# Git for Backups

- Example: Directory needs regular backups
  - Could use rsync but unwieldy in size
- Create Git repository for appropriate directory
  - Regular local commits
  - Regular push to backup location
  - Get simple revision history

# Git for Configuration Management

- Example: Apache configurations
  - Multiple environments (dev/test/production)
  - Minor differences between environments
    - IP Address
    - Log levels
  - Want to effectively move changes across environments
- Great way to manage different “experiments” where different code versions/data are used-produced. With tags/branches its efficient and much better than commenting code in/out.

# Git for Documents

- Well matched for latex.
  - Best if one-sentence-per-line
- Well matched for multi-person editing
- Authors/Editor clone from central server
- Being used for
  - Overleaf

# Git Assignment For next week

- Checkout out git repo from <https://github.com/tboult/A3-CS6000F19>
- You are to make a new latex file, named after your last name. In that file copy some of your first assignment about your self, including photo, add the associated files to the repo. Then update Assignment3.tex file to \include your file. Make sure all latex compiles properly and shows your content. Then commit/push your changes.
- After at least 2 other people have uploaded their files, you are to edit one of their files to ask them a question at the end of their file. You are to ask at least 2 people questions, and no one can be asked more than 2 questions. When someone asks you a question, you are to update your file to answer. If there are any merge conflicts, deal with them. If there are more than two questions in your file, revert it back to a point where they were only 2.

# CS6000 F18

# Introduction to LaTeX

- Some content taken from Oetiker et al. “The Not So Short Introduction to LaTeX 2e”

# Average Anatomy of a Thesis

- 6 Chapters, 2 Appendixes
- 14 Section
- 18 Subsections
- 196 Charts
- 32 Figures
- 48 Tables
- 95 cited works

# Data for a Thesis

- ~ 40,000 raw data files
- ~ 8,000 processed data files
- 865 charts
- 41 tables
- **Charts and tables change when data changes!**

# Managing the Complexity

- Scripts to process /gnerate data
  - ◆ Python, Shell scripts, awk, bc, grep, sed, wc, etc...
- Non-interactive programs to create charts
  - ◆ Matplotlib, gnuplot, Matlab, latex
- A document preparation system that:
  - ◆ Decomposes and compartmentalizes writing
  - ◆ Integrates with other tools
  - ◆ Automates formatting and numbering

# There exists alternatives to Word, one example is LaTeX - a typesetting language similar to C/HTML

## A Meshless Based Solution to Vectorial Mode Fields in Optical Microstructured Waveguides

D.R. Burke and T.J. Smy  
Carleton University, Department of Electronics,  
Ottawa, ON, Canada  
[drburke@elec.carleton.ca](mailto:drburke@elec.carleton.ca)

**Abstract:** A meshless method for the solution of full vectorial optical mode fields has been applied to micro-structured optical waveguides. The Finite Cloud Method is used to approximate the solution using a point distribution and material definitions. Presented are two methods of defining material interfaces, one which implements a step index and a second which uses a graded index. Coupled field equations are used to solve for both transverse components of the magnetic field as well as the guided wavelength and effective index of refraction. Comparing results for a ridge waveguide, solid core, micro-structured and air core structures with commercial FEM solvers highlight the methods versatility, accuracy and efficiency.

© 2012 Optical Society of America  
OCIS codes: (000,000) General.

### References and Links

1. P. S. Russell, "Photonic-crystal fibers," *J. Lightwave Technol.*, vol. 24, pp. 4730-4740, Dec 2006.
2. P. Luisse, P. Shaw, J. Schulte, and H.-G. Unger, "Analysis of vectorial mode fields in optical waveguides by a new finite difference method," *Journal of lightwave technology*, vol. 12, pp. 487-494, 1994.
3. H. Urano and H. Hoekstra, "Modelling of microstructured waveguides using a finite-element-based vectorial mode solver with transparent boundary conditions," *Optics Express*, vol. 12, no. 12, pp. 2795-2809, 2004.
4. D. Burke, S. Li, and T. Smy, "Simulating anisotropic heterogeneous models using the finite cloud method," *Microwires & Waveguides*, vol. 41, no. 5, pp. 336-340, 2010.
5. J.-S. Chen, S. Yoon, H.-P. Wang, and W. K. Liu, "An improved reproducing kernel particle method for nearly incompressible finite elasticity," *Computer Methods in Applied Mechanics and Engineering*, vol. 181, no. 13, pp. 137 - 143, 2000.
6. COMSOL Multiphysics, <http://www.comsol.com>, 2011, Version 4.1 Comsol Inc.
7. Rsoft, FemSIM, <http://www.rssoftdesign.com/products.php?sub=Component+Design&init=FemSIM>, 2011, Version 3.3 Rsoft Inc.
8. N. Aluru and G. Li, "Finite cloud method: a true meshless technique based on a fixed reproducing kernel approximation," *Journal of Computational Physics*, vol. 227, pp. 551-572, 2008.
9. J. D. Jackson, *Classical Electrodynamics*, New York: Academic Press, 1998.
10. K. Ramm, P. Luisse, and H.-G. Unger, "Multigrid eigenvalue solver for mode calculation of planar optical waveguides," *IEEE Photonics Technology Letters*, vol. 9, no. 7, pp. 967-969, 1997.
11. MathWorks, <http://www.mathworks.com/products/matlab/>, 2011.
12. GNU Octave, <http://www.gnu.org/s/octave/>, 2011.

### 1. Introduction

With the recent development of both optical fiber with complex geometric structure [1] and the integration of a wide variety of geometrically complex optical waveguides into optoelectronic

## A Meshless Based Solution to Vectorial Mode Fields in Optical Microstructured Waveguides

D.R. Burke and T.J. Smy  
December 13, 2011

### Abstract

A meshless solution to vectorial mode fields has been applied to various micro-structured optical waveguides. The Finite Cloud Method, FCM, has been used to solve coupled field equations for both transverse components of the magnetic field as well as the effective index of refraction for the waveguides. Two methods using either a step-index or a graded-index have been implemented and compared. An approximation to the solution is found using a distribution of points and a cloud about each point, with no mesh and minimal geometric linking knowledge between the points. This gives the ability to use a highly irregular point distribution which can be easily modified or tailored to micro-structured fibers in order to accurately represent the vectorial modal solution. In addition, the use of Bayliss-Gunzburger-Turkel-like transparent boundary conditions (TBC) and an iterative process is compared with a perfectly matched layer (PML), both of which allow for the solution of leaky modes for the structures. Results for ridge waveguides and solid core fibers having low index contrast are in high agreement with the solutions from commercial solvers. Further results with high contrast air hole structures are compared with other solution methods giving promising results and highlight the methods versatility, accuracy and efficiency for a wide range of problems.

### 1 Introduction

As optical waveguides become more complex and integrated into optoelectronic circuits there is a need to accurately model their propagation constants and modes. The guided modes for an optical transmission line describe both the field intensities as well as the effective index of refraction ( $n_{eff}$ ) and group velocity of the signals.

Typical vectorial field mode solvers use either the Finite Difference Method and a regular rectangular mesh [3] or the Finite Element Method and an irregular complicated mesh [4]. This paper has adapted the Finite Cloud Method, a meshless method, using an irregular point distribution and two heterojunction boundary methods which can be used for both step-index and graded index solutions, as an alternative to the above methods.

## A Meshless Based Solution to Vectorial Mode Fields in Optical Microstructured Waveguides

D. R. Burke<sup>a</sup> and T. J. Smy<sup>a</sup>  
<sup>a</sup>Carleton University, 1125 Colonel By Drive, Ottawa, Canada

### ABSTRACT

A meshless solution to vectorial mode fields has been applied to various micro-structured optical waveguides. The Finite Cloud Method (FCM), has been used to solve coupled field equations for both transverse components of the magnetic field as well as the effective index of refraction for the waveguides. Two methods using either a step-index or a graded-index have been implemented and compared. An approximation to the solution is found using a distribution of points and a cloud about each point, with no mesh and minimal geometric linking knowledge between the points. This gives the ability to use a highly irregular point distribution which can be easily modified or tailored to micro-structured fibers in order to accurately represent the vectorial modal solution. In addition, the use of Bayliss-Gunzburger-Turkel-like transparent boundary conditions (TBC) and an iterative process is compared with a perfectly matched layer (PML), both of which allow for the solution of leaky modes for the structures. Results for ridge waveguides and solid core fibers having low index contrast are in high agreement with the solutions from commercial solvers. Further results with high contrast air hole structures are compared with other solution methods giving promising results and highlight the methods versatility, accuracy and efficiency for a wide range of problems.

**Keywords:** meshless methods, optical mode solving, transparent boundary condition, perfectly matched layer, leaky modes, guided modes

### 1. INTRODUCTION

As optical waveguides become more complex and integrated into optoelectronic circuits there is a need to accurately model their propagation constants and modes. The guided modes for an optical transmission line describe both the field intensities as well as the effective index of refraction ( $n_{eff}$ ) and group velocity of the signals.

Typical vectorial field mode solvers use either the Finite Difference Method (FDM) and a regular rectangular mesh<sup>1</sup> or the Finite Element Method (FEM) and an irregular complicated mesh.<sup>2</sup> This paper has adapted the Finite Cloud Method, a meshless method, using an irregular point distribution and two heterojunction boundary methods which can be used for both step-index and graded index solutions, as an alternative to the above methods.

The method is briefly described below and has previously been adapted to solve nonhomogeneous partial differential equations, such as the heat transfer equation,<sup>3</sup> and has been used as the basis for the step-index based solutions. A quasi-homogeneous method also described below has been used for the graded-index method. Both solutions are compared for all structures presented, and as well as compared with known or otherwise referenced solutions.

This work utilizes coupled H-field equations, and can be applied to anisotropic heterogeneous graded-index or step-index structures which are invariant in the z-direction. The models in this work are restricted to isotropic materials. In the first two examples Dirichlet boundary conditions are used, thus requiring the solutions to be guided modes with purely real propagation constants. The remaining more complex structures tested are solved for leaky mode eigenvalues and values using both Bayliss-Gunzburger-Turkel-like transparent boundary conditions (TBC),<sup>2</sup> or a perfectly matched layer (PML).<sup>4</sup> The following paper will briefly explain the methods and equations used, and will compare several solutions for complex structures with commercial solvers.

Further author information: (Send correspondence to D. R. Burke)  
D. R. Burke: E-mail: drburke@elec.carleton.ca  
T. J. Smy: E-mail: tjs@elec.carleton.ca

# LaTeX can be used for papers, labs, word documents, assignments, presentations

## Contents

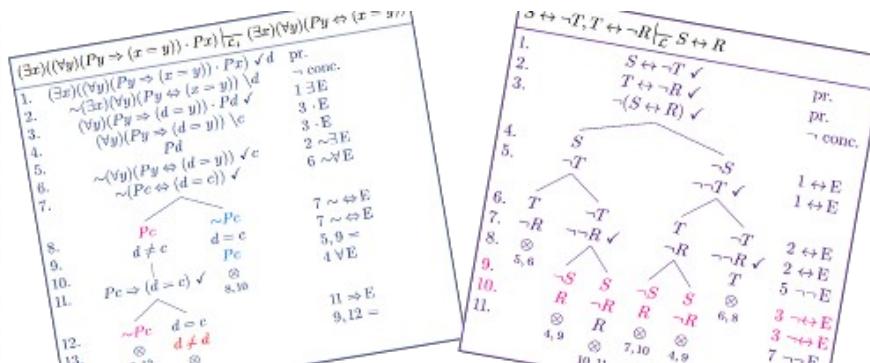
|       |  |    |
|-------|--|----|
| 1     | Introduction                               | 6  |
| 1.1   | Discretization and Finite Element Analysis | 6  |
| 1.2   | Monte Carlo Method                         | 8  |
| 1.3   | Photon Propagation                         | 8  |
| 1.4   | Applications                               | 9  |
| 2     | The Engineering Project                    | 10 |
| 2.1   | Health and Safety                          | 10 |
| 2.2   | Engineering Management                     | 10 |
| 2.3   | Project Management                         | 11 |
| 2.3.1 | Project Schedules                          | 11 |
| 2.3.2 | Team Coordination                          | 12 |
| 2.3.3 | Project Partition                          | 12 |
| 2.3.4 | Team Interactions and Communication        | 12 |
| 3     | Theory                                     | 14 |
| 3.1   | Reflection and Refraction                  | 14 |
| 3.2   | Attenuation and Absorption                 | 16 |
| 3.3   | The Polarization                           | 16 |
| 3.3.1 | Perpendicular Polarization                 | 18 |
| 3.3.2 | Parallel Polarization                      | 18 |
| 3.4   | Diffraction                                | 20 |
| 4     | Geometry Creation                          | 22 |
| 4.1   | Motivation                                 | 22 |
| 4.2   | Elements of Geometry Creation              | 23 |
| 4.2.1 | Testing Environment                        | 23 |
| 4.2.2 | Shapes                                     | 23 |

ELEC 1908 - L<sup>A</sup>T<sub>E</sub>X Assignment

February 15th, 2011

## Assignment :

- Retrieve and unzip <http://www.doe.carleton.ca/~drburke/latex.zip> on your network drive
- In this archive are examples, guides and the assignment
- Open TeXnicCenter and the 'tex' file from the archive. Using TeXnicCenter compile this file PDF. These two files are your starting point
- Goal is to recreate 'final.report.pdf', you may copy and paste the text, but watch for errors and problems, help can be found online and in the provided manuals
- Do not worry if the tex program puts things in a different order, L<sup>A</sup>T<sub>E</sub>X will deal with style and order, you are to deal with the content, links, references, equations, etc.
- Create a new section for the report, include Schrödinger's equation (in any form), and reference this equation in a short sentence or paragraph describing the equation
- In this new section explain how one can open a terminal/command prompt in windows, navigate to the correct folder for your 'tex' file and compile the document using only terminal commands
- Submit only .tex file and .pdf file to drburke@gmail.com with 'ELEC 1908-Latex Assignment' as the subject



$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x, t) + V(x) \Psi(x, t) = i\hbar \frac{\partial}{\partial t} \Psi(x, t)$$

$$\frac{\hbar^2}{2m} \begin{bmatrix} \frac{\partial^2}{\partial x^2} + V & 0 \\ 0 & \frac{\partial^2}{\partial x^2} + V \end{bmatrix} \begin{bmatrix} \Psi_{re} \\ \Psi_{im} \end{bmatrix} = \begin{bmatrix} 0 & -\hbar \\ \hbar & 0 \end{bmatrix} \frac{\delta}{\delta t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

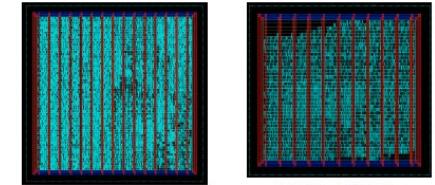


Fig. 7. Cell design 'physical' view  
Fig. 8. Radiogated cell, 'physical' view

taking up significantly less area. A cell layout of the design can be seen in Figure 9. The total dynamic power was estimated at 13.38 mW.

Testing the square root function, a solution is returned in 15 full clock cycles, resulting in an approximate square root data rate of 0.067 MHz. The total clock period was approximately 5.8 MHz, giving a factor 20 increase in the dedicated circuit.

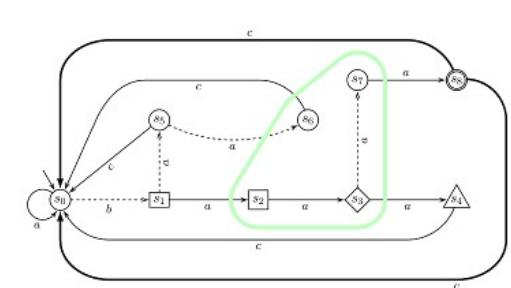
All of the above tests are very approximate in nature and may not represent the actual time savings once other factors are incorporated. They do, however, show that as complexity increases the savings of a dedicated specialized circuit become greater.

## IV. SUMMARY

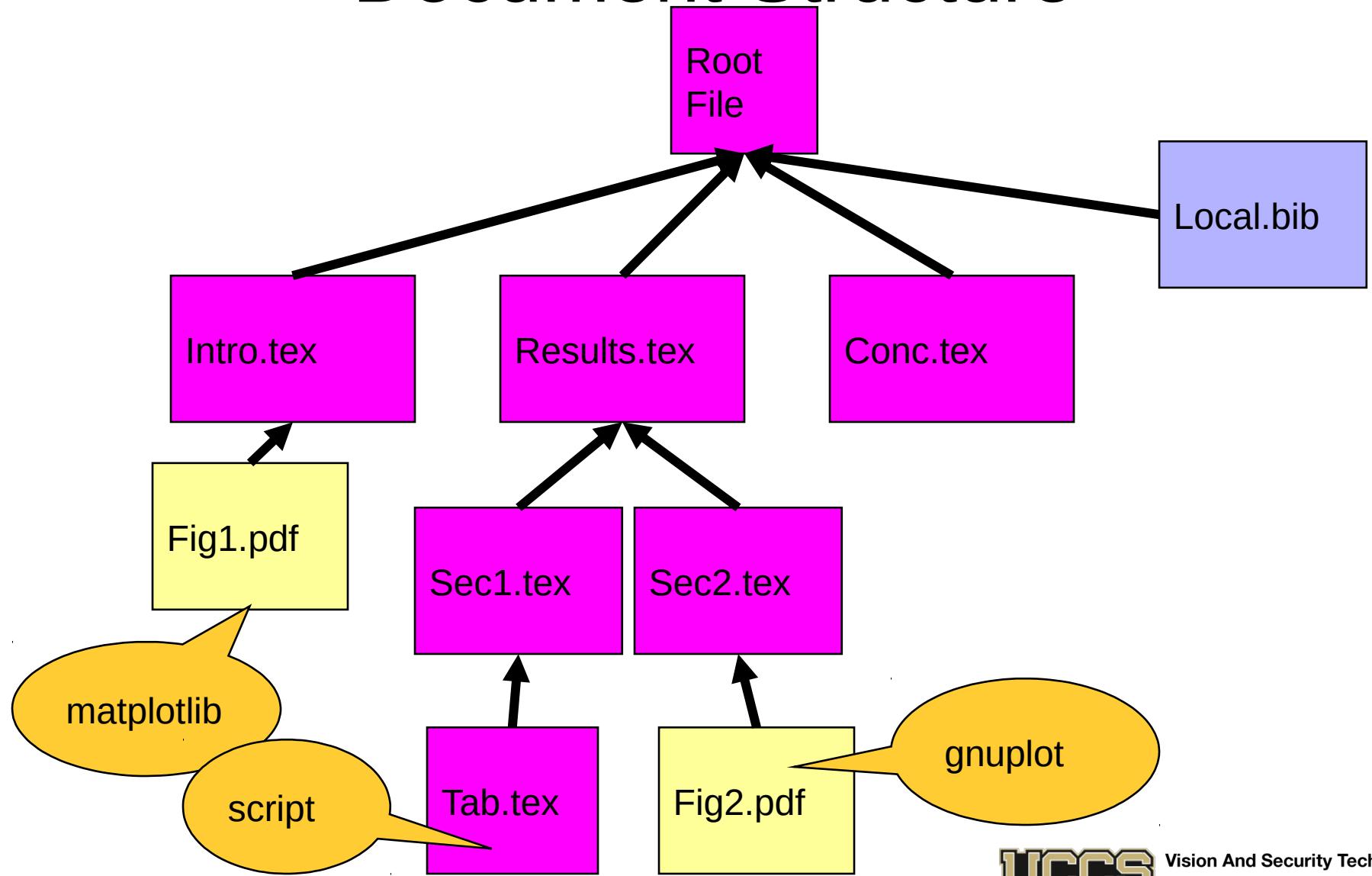
A statistical analysis circuit was designed, simulated and redesigned with improved efficiency for this report. This type of circuit could be used in engineering and scientific applications to greatly speed up data analysis requests.

The design described here is just the beginning of the process of digital design as well as learn a higher level digital design language, in this case Verilog. This process worked rather well as there is a considerable of documentation on-line for the use of Verilog and many examples and tutorials from which to learn.

Designing a specialized circuit to be more efficient than what it was to be performed to the general processor from within the programs and design language. It was also learned to learn to tailor each specific element to the desired task as opposed to using generic elements of Verilog, such as the switch, which is not very efficient. This tailoring allows for specifications to be met for speed, cost and power, which is highly important for VLSI design.



# Document Structure

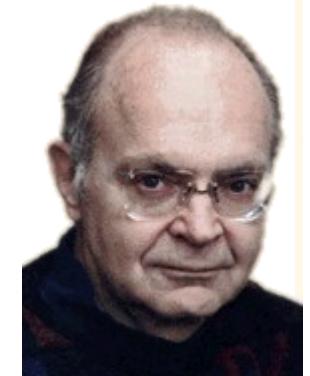


# Resources

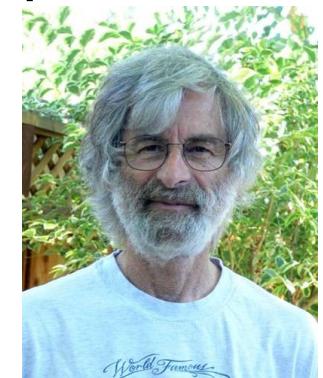
- <https://ctan.org/> the Comprehensive TE  
X Archive Network (Good for packages/tools)
- For TeX (with lots of latex as well) :
  - <http://www.tug.org/>
- LaTex project page
  - <http://www.latex-project.org/>
- Not so short introduction to LaTeX:  
<http://ctan.tug.org/tex-archive/info/lshort/english/lshort.pdf>
- Latex Video Tutorials  
<https://www.youtube.com/watch?v=SoDv0qhysQ&list=PL1D4EAB31D3EBC449>
- Can read the LaTex source code (its written in TeX)
- Google is your friend...

# History

- Late 1970s -> 1989: TeX by Don Knuth
  - Decent typesetting by the **authors!**
  - Same output now as in the future
  - Low level, but powerful
  - Frozen development



- Starting early 1980s: LaTeX by Leslie Lamport
  - High level macros for/in TeX
  - Easy to use! Moderate/difficult to extend.
  - Current version is Latex2e



# LaTeX as in “latex” or...?

- Some say LAY-teck
- Some say LAH-teck
- Some say LAY-tech (as in “Bach”)
- Some say LAY-tex
- ...
- According to Wikipedia: Knuth says /tex/ (as in Bach)
- Lamport says “whatever...”

# Getting started...

- All you need is:
  - LaTeX installation with appropriate interpreters
  - A .tex file (the document)

Can also have, (almost always do have):

- External style files
- TeX editors (more on this later)
- BibTex

# Installation

- Linux
  - Typically comes with LaTeX or has it easily available in the standard repositories
  - “TeX Live”
- Mac
  - <http://www.tug.org/mactex/>
- Windows
  - <http://www.tug.org/protext/>
  - This will install MiKTeX, Ghostscript, and TeXnicCenter.
- Extra packages
  - [http://en.wikibooks.org/wiki/LaTeX/Packages/Installing\\_Extra\\_Packages](http://en.wikibooks.org/wiki/LaTeX/Packages/Installing_Extra_Packages)

# Online tools

## Online tools supporting latex

- <https://www.overleaf.com/>
- [\(they bought https://www.sharelatex.com/\)](https://www.sharelatex.com/)
  
- latexlab.org (integrates with Google Docs)
- <http://quicklatex.com/> , codecogs.com  
or <http://Mathstools.com> (If you must use word/PPT can get quick latex-based images. latter two can also help learn syntax as have GUI for generation)

# Overleaf

- ✖ Web and Browser based latex editor
- ✖ Cloud based storage
- ✖ Large number of templates
- ✖ Collaborative document development

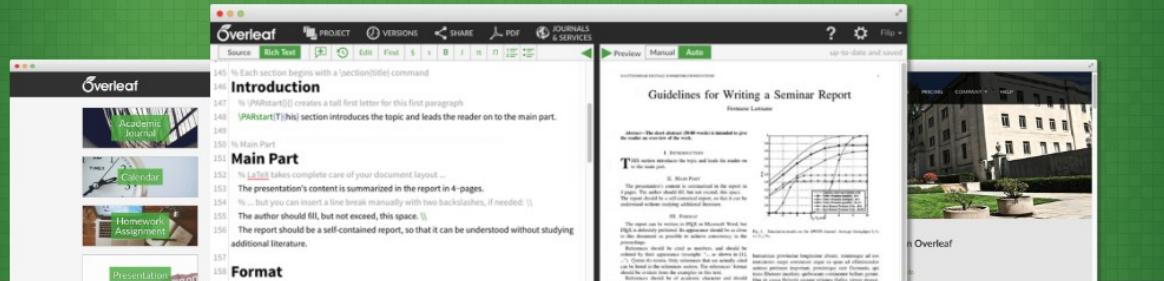


# Collaborative Writing and Publishing

Join 750,000+ authors enjoying the easiest way to create, collaborate and publish online

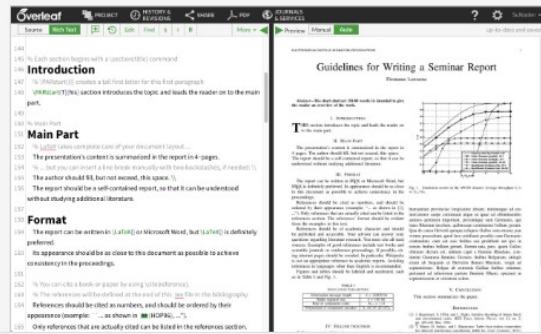
*Start writing now!*

**CREATE A NEW PAPER**



Over 10 million projects, authors from 3600 institutions and more than 2400 templates

Overleaf is an online LaTeX and Rich Text collaborative writing and publishing tool that makes the whole process of writing, editing and publishing scientific documents much quicker and easier.



## Overleaf for Authors

Real-time collaboration in your browser: The convenience of an easy-to-use WYSIWYG manuscript editor, with real-time collaboration and structured, fully typeset output produced automatically in the background as you type.

Prefer to edit directly in LaTeX? Overleaf provides a full collaborative online LaTeX editor you can switch to at any time.

*More on our [benefits page](#)*

Overleaf PROJECT HISTORY & REVISIONS SHARE PDF JOURNALS & SERVICES

Files... Word Count

MD.tex

DOWNLOAD AS ZIP Save to Dropbox

```

1 \documentclass{article}
2 \usepackage[paperwidth=8.5in, top=1.87cm, bottom=1.87cm, left=1.87cm, right=1.87cm]{geometry}
3 \usepackage{graphicx}
4 \usepackage{amssymb}
5 \usepackage{amsmath}
6 \usepackage{xcolor}
7
8 \title{ELEC 4700 Molecular Dynamics PA Assignment}
9 \author{Tom Smy}
10
11 \begin{document}
12 \maketitle
13
14 \text{\em You are encouraged to work in groups of 2-3. Ideally a strong programmer with less experienced people.}
15
16 \begin{description}
17 \item[Goal] In this PA you should familiarize with yourself with Molecular Dynamics modeling.
18 \item[Tasks]
19 \begin{enumerate}
20 \item Basic MD Code simulation:
21 \begin{enumerate}
22 \item if you need to git fork and clone the 4700 code repo. To get the latest code from my repo you can do (I think):\ \
23 \begin{noindent}Open (\texttt{git bash}) and (\texttt{cd}) into the 4700Code folder. Then do:\ \
24 \begin{quote}
25 \begin{itemize}
26 \item git add remote tjs https://github.com/tjssmy/4700Code.git
27 \item git fetch tjs
28 \item git merge tjs/master
29 \item git push origin
30 \end{itemize}\ \
31 \end{quote}
32 \item open Matlab in the (\texttt{MDCode}) folder
33 \item open file (\texttt{MD.m}) and check (\texttt{InitVStream}) is chosen.
34 \item Run MD. Put some breakpoints in and poke around.
35 \end{enumerate}
36 \item Do one of the following:
37 \begin{enumerate}
38 \item Take the existing code and add something to it. For example:
39 \begin{enumerate}
40 \item Add a new geometry to the (\texttt{AddAtomic*.m}). Perhaps an elliptical mass? See (\texttt{AddCirAtomicArray}). Play with it and add other components.
41 \item Add a third atomic mass or make the mass able to be different for each atom. You need to a global search to find where (\texttt{Mass0}) is used and alter the code (\texttt{Everywhere!})
42 \item Generalize (\texttt{AddParticalStream}) to do something cool! Figure out what it can do now and think up something better.
43 \end{enumerate}
44 \item write a simple 1D MD code using a spring like force ( $F = k(x - x_0)^2$ ). See how far you get.
45 \begin{itemize}
46 \item Use only a few atoms 5-10
47 \item Initially space them uniformly and then add a small perturbation.
48 \item Use the force from the two nearest neighbors to calculate the acceleration.
49 \item Play with the spring constant  $k$  can you break it?
50 \end{itemize}
51 \end{enumerate}
52 \end{enumerate}
53
54 when you are finished you should either:
55 \begin{enumerate}
56 \item if you modified the 4700 code you should push the new code to your repo.

```

Preview Manual Auto

up-to-date and saved

ELEC 4700 Molecular Dynamics PA Assignment

Tom Smy

January 22, 2018

You are encouraged to work in groups of 2-3. Ideally a strong programmer with less experienced people.

**Goal** In this PA you should familiarize with yourself with Molecular Dynamics modeling.

**Tasks**

1. Basic MD Code simulation:
  - if you need to git fork and clone the 4700 code repo. To get the latest code from my repo you can do (I think):
    - Open *gitbash* and *cd* into the 4700Code folder. Then do:
      - git add remote [tjs](#) <https://github.com/tjssmy/4700Code.git>
      - git fetch [tjs](#)
      - git merge [tjs/master](#)
      - git push origin
    - open Matlab in the *MDCode* folder
    - open file *(MD.m)* and check *(InitVStream)* is chosen.
    - Run MD. Put some breakpoints in and poke around.
2. Do one of the following:
  - Take the existing code and add something to it. For example:
    - Add a new geometry to the *AddAtomic\*.m*. Perhaps an elliptical mass? See *AddCirAtomicArray*. Play with it and add other components.
    - Add a third atomic mass or make the mass able to be different for each atom. You need to a global search to find where *Mass0* is used and alter the code *Everywhere!*
    - Generalize *AddParticalStream* to do something cool! Figure out what it can do now and think up something better.
  - Write a simple 1D MD code using a spring like force ( $F = k(x - x_0)^2$ ). See how far you get.
    - Use only a few atoms 5-10
    - Initially space them uniformly and then add a small perturbation.
    - Use the force from the two nearest neighbors to calculate the acceleration.
    - Play with the spring constant  $k$  can you break it?

When you are finished you should either:

- If you modified the 4700 code you should push the new code to your repo.
- If you made a new simulator you should create a new empty repo in github, clone it on your computer, add the code to the repo and push it back to github.

*Make sure before you leave you get myself or a TA to review the status.*

# WHERE CAN I FIND IT ?

- ✖ Linux – check documentation for the Latex source  
(suggested IDE: Kile)
- ✖ Mac OS X – use MacTeX, or homebrew or macports  
to install as a package (suggested IDE: TeXShop)
- ✖ Windows – MikTex or ProTeXt  
(suggested IDE: TexWorks or Texniccenter)
- ✖ In the lab click on file and TexWorks should start up.

# Making your first LaTeX documents

1. Start with a skeleton document
2. Write your text
3. Annotate the appropriate parts (math, etc)
4. Interpret through appropriate program
5. Fix errors or modify, and try again...

Older than but not unlike HTML or other markup languages

Interpret via the provided programs:

`pdflatex -> PDF` (older version `latex -> DVI` (`dvi2ps` or `dvi2pdf`)

Or built in features in your IDE.

# LaTeX is written in structured macro markup , then compiled like a ‘C’ program

```
\documentclass[12pt]{article}%
\usepackage{amsmath}%
\usepackage{amsfonts}%
\usepackage{amssymb}%
\usepackage{graphicx}%
\usepackage{setspace}%
\usepackage{cite}%
\usepackage{hyperref}%
%-----
-----%
\begin{document}
\title{}%
\author{Terry
\\ \emph{with} }%
\date{\today}%
\maketitle

\singlespacing

\begin{abstract}%
\end{abstract}

\doublespacing

\section{Theory}
```

compile  
→

file.out  
file.aux  
file.log  
file.pdf

# Structure

- A basic document:
- `\documentclass[11pt, twocolumn]{article}`
- `\usepackage{amsmath, graphicx}`
- `\begin{document}`
- `%document contents go here`
- `\end{document}`
- Notice:
  - `\begin` and `\end` (these define “environments”)
  - `{ }` and `[]` around parameters to commands
  - Commands typically start with backslash
  - Commands “gobble” spaces.. spaces collapse to 1
  - % starts comment, goes to end of line

# PACKAGES ?

➤ Add functionality to your documents

➤ Images

➤ Equations

➤ Links

➤ Etc.

# PACKAGES

\usepackage{amsmath} %

\usepackage{amsfonts} %

\usepackage{amssymb} %

\usepackage{graphicx}

\usepackage{setspace}

\usepackage{cite}

\usepackage{hyperref}

# Formatting Text

- Emphasis and size

`\textbf{some bold text} \emph{some italic text} \underline{some underlined text}`

`\large Some large text.} {\Large Larger text.} {\small Small.}`

- Spacing

- Many spaces = one space
- Use `\n` for newline
- Hit return twice for a new paragraph
- `\newpage`
- “Glue” effects layout.. can stretch or shrink

- Quotes are done with ```` and `''`, not `“`

- Add comments `%comment text until end of line`

- Like any language, some characters are special. For example, `\$ {} %` cannot be written alone. Use `\$` or ...

# Organization

- Indenting text as a quotation (an environment!)

`\begin{quote}`

*A quote.*

`\end{quote}`

- Section headings

- `\section{an arbitrary name}`

- `\subsection{an arbitrary name}`

- `\subsubsection{an arbitrary name}`

# Lists

- Done with an environment
- `\begin{enumerate}`
- `\item some item in the list`
- `\item another item in the list`
- `\end{enumerate}`
  
- Replace `enumerate` with `itemize` for non-numbered
- Various packages or commands to effect layout of lists

# Images/Graphics

- \usepackage{graphicx}
- \includegraphics[height=50%,width=50%]{filename.png}
- \includegraphics[width=.9\columnwidth]{filename.png}
- \includegraphics[width=3.1in]{filename.png}
- \includegraphics[attr1=val1, attr2=val2, ..., attrn=valn]{image.eps}
  
- LaTeX explicitly supports EPS file format, as said in the documentation, but it also supports \*.jpg, \*.png, \*.pdf

# Boxes: Key Concept

- All of tex/latex is based boxes, every char is a box.
- The concept to understand is that the file needs an area to be displayed in.
- This area is called the bounding box. If not in file (or raw image) can specify it with bb argument. Else uses file/paper type.

➤ Example:

```
\documentclass{article}  
\usepackage{graphicx}  
\begin{document}
```

...

```
\includegraphics[bb = 1 2 10 20, width = 4in]{file.eps}
```

...

```
\end{document}
```

# Attributes

- **bb = XXX XXX XXX XXX**  
Bounding box. The first two sets of points, represent the lower-left corner of graphic being the point in the text it was inserted, the second set of points are the upper right corner position.
  - **width = xx**  
Specify the preferred width of the imported image to xx.\*
  - **height = xx**  
Specify the preferred height of the imported image to xx.\*
  - **keepaspectratio**  
This can be set to either true or false. When true, it will scale the image according to both height and width, both will not distort the image, so that neither width or height are exceeded. *Eg. keepaspectratio = true*
  - **scale = xx**  
Scales the image by the desired scale factor. e.g, 0.5 to reduce by half, or 2 to double.
  - **angle = xx**  
This option can rotate the image by xx degrees (anti-clockwise)
  - **trim=l b r t**  
This option will crop the imported image by l from the left, b from the bottom, r from the right, and t from the top. Where l, b, r and t are lengths.
  - **clip**  
For the trim option to work, you must set clip=true
- \*Only specifying one of width/height will scale the image while keeping the aspect ratio.

# Example

```
\documentclass{article}
\usepackage{graphicx}
\begin{document}

\begin{center}
\title{How to insert an image in LaTeX}
\fontsize{15pt}{22pt}\textbf{How to insert an image in LateX}
\end{center}
```

Well lets see how this works. What can you really do with this software, can we insert an image after a paragraph? Are we able to insert an image inline with the text. Are there other features to be had with inserting an image with LaTeX.

```
\par\bigskip
\includegraphics[bb = 0 0 100 150]{images.jpg}
\par
More text to see where it wraps to.

\end{document}
```

# Example

## How to insert an image in LateX

Well lets see how this works. What can you really do with this software, can we insert an image after a paragraph? Are we able to insert an image inline with the text. Are there other features to be had with inserting an image with LaTeX.



More text to see where it wraps to.

# Example

```
\documentclass{article}  
\usepackage{graphicx}  
\begin{document}  
  
\begin{center}  
\title{How to insert an image in LaTeX}  
\fontsize{15pt}{22pt}\textbf{How to insert an image in LateX}  
\end{center}
```

Well lets see how this works. What can you really do with this software, can we insert an image after a paragraph? Are we able to insert an image inline with the text. Are there other features to be had with inserting an image with LaTeX.

```
\par\bigskip  
\includegraphics[bb = 0 125 100 150, trim = 0 0 50 150, clip = true]{images.jpg}  
\par  
More text to see where it wraps to.
```

```
\end{document}
```

# Example

## How to insert an image in LateX

Well lets see how this works. What can you really do with this software, can we insert an image after a paragraph? Are we able to insert an image inline with the text. Are there other features to be had with inserting an image with LaTeX.

More text to see where it wraps to.



---

But just `insertgraphics` is bad..

# Figures

- Latex has a powerful set of “Floats”, where it dose placement

```
\begin{figure}[th]
```

```
    \centering
```

```
    \includegraphics[width=8cm]{actorcritic}
```

```
    \caption{Sutton's Actor-Critic method.}
```

```
    \label{fig:actorcritic}
```

```
\end{figure}
```

```
\begin{figure}[h!]
```

```
    \resizebox{\textwidth}{!}{\rotatebox{-90}{\includegraphics{modules}}}
```

```
    \caption[Systems modules and messages.]{Flat layout of message passing  
between modules.}
```

```
    \label{fig:modules}
```

```
\end{figure}
```

- Latex treats figures as floating objects
- Placement specifiers can be used
  - ◆ **h** place item *here* in the text
  - ◆ **t** place item at the *top* of a page
  - ◆ **b** place item at the *bottom* of a page
  - ◆ **p** place item on a special *page* containing only floats
  - ◆ **!** override other parameters even if (Latex thinks) the result looks awful.

# Labels and references

- Use labels and references to automatically insert reference numbers:

```
\subsection{The first subsection} \label{arbitrarylabel}
```

*Some text.*

```
\subsection{Next subsection}
```

*The previous subsection was \ref{arbitrarylabel}.*

- And for equations:

```
\begin{equation} E = mc^2 \label{einstein}
```

```
\end{equation}
```

*Einstein is associated with Eq.~\eqref{einstein}.*

# Tables

- Table is another “float” environment!, tabular is the actual layout
- Tables are tricky and have a lot of options. An example:
- `\begin{tabular}{|r||l|}`
- `\hline`
- `7C0 & hexadecimal \\`
- `3700 & octal \\ \cline{2-2}`
- `11111000000 & binary \\`
- `\hline \hline`
- `1984 & decimal \\`
- `\hline`
- `\end{tabular}`
- `|r||l|` tells how to setup and align the columns. `&` sets the columns
- See references for more details...

|             |             |
|-------------|-------------|
| 7C0         | hexadecimal |
| 3700        | octal       |
| 11111000000 | binary      |
| <br>        |             |
| 1984        | decimal     |

# Tables

```
\begin{table}[ht]
  \caption{Evaluated conditions.}
  \centering
  \begin{tabular}{| c c |}
    \hline\hline
    & Fixed & Learning \\
    \hline
    Without cancel region & (1)(2) & (4) \\
    With cancel region & (3) & (5) \\
    \hline
  \end{tabular}
  \label{table:conditions}
\end{table}
```

Ideally generate tables from code so can update with any new run.

Table 5.2: Evaluated conditions.

|                       | Fixed  | Learning |
|-----------------------|--------|----------|
| Without cancel region | (1)(2) | (4)      |
| With cancel region    | (3)    | (5)      |

```
\documentclass{article}  
\usepackage{csvsimple}
```

```
\begin{document}
```

```
\section*{Easy}  
\csvautotabular{test.csv}
```

```
\section*{Theory}
```

```
\begin{tabular}{r|r|r}%  
% specify table head  
\bf Time (s) & \bf Rel. time (s)& \bf Y  
% use head of csv as column names  
\csvreader{test.csv}{}  
% specify selected coloumns here  
{\\\hline\csvcoli&\csvcolii&\csvcolvi}  
\end{tabular}  
\end{document}
```

### Easy

| Time (s) | Rel. time (s) | X Pos | Rel X Pos | Raw Y Pos | Model Y Pos |
|----------|---------------|-------|-----------|-----------|-------------|
| 43.97    | 0             | 734   | 528       | 14.22624  | 18.26294    |
| 44.01    | 0.04          | 731   | 525       | 14.11335  | 18.14345    |
| 44.04    | 0.07          | 729   | 523       | 14.03819  | 18.06389    |
| 44.07    | 0.1           | 726   | 520       | 13.9256   | 17.9447     |
| 44.11    | 0.14          | 720   | 514       | 13.70096  | 17.70686    |
| 44.14    | 0.17          | 718   | 512       | 13.62624  | 17.62774    |
| 44.17    | 0.2           | 714   | 508       | 13.47704  | 17.46974    |
| 44.21    | 0.24          | 711   | 505       | 13.36535  | 17.35145    |
| 44.24    | 0.27          | 706   | 500       | 13.1796   | 17.1547     |
| 44.27    | 0.3           | 700   | 494       | 12.95736  | 16.91926    |
| 44.31    | 0.34          | 696   | 490       | 12.8096   | 16.7627     |

### Theory

| Time (s) | Rel. time (s) | Y Pos    |
|----------|---------------|----------|
| 43.97    | 0             | 18.26294 |
| 44.01    | 0.04          | 18.14345 |
| 44.04    | 0.07          | 18.06389 |
| 44.07    | 0.1           | 17.9447  |
| 44.11    | 0.14          | 17.70686 |
| 44.14    | 0.17          | 17.62774 |
| 44.17    | 0.2           | 17.46974 |
| 44.21    | 0.24          | 17.35145 |
| 44.24    | 0.27          | 17.1547  |
| 44.27    | 0.3           | 16.91926 |
| 44.31    | 0.34          | 16.7627  |

Can also use  
<http://www.tablesgenerator.com/>

# Math Mode

➤ “Math mode” formats formal notation.

3+1 in normal mode

will **not** look like

3+1 in math mode

➤ To enter inline math mode, use \$ and \$

➤ *When \$x = 3\$, \$f(x,y)\$ reduces to \$y^2+16\$.*

➤ For standalone math lines, use \[ and \]

*The equation can be expressed as follows:*

\[ f(x) = x^2 + 4x + 3 = (x+1)(x+3) \]

➤ or

\begin{equation}

$f(x) = x^2 + 4x + 3 = (x+1)(x+3)$

\end{equation}

# Math Mode

- White space is typically ignored

$\$f(x) = g(x)\$$  will be the same as  $\$f( x )= g ( x )\$$

Math eats spaces, many symbols contain natural spacing for other uses can use specialized spacing commands \; \,, \,

- For text in math mode, use  $\textit{\text{textrm}}$

$\$My\ Fun.\ or\ \textit{\text{textrm}{My\ Fun.\ or\ }}\ \textit{\text{textrm}{My\ Fun.}}\$$

*MyFun.orMy Fun. or My Fun.*

Most math is built in. But the amsmath package predefines more. Just use the package

$\textit{\text{usepackage}{amsmath}}$

# Math Symbols, etc

Subscript and superscripts:  $x^2 \ x_2$

Fractions:  $\frac{a}{b+c}$

Radical:  $\sqrt{x+y}$

Floor and ceiling:  $\lfloor x+y \rfloor \lceil x+y \rceil$

Operators and relations:

$\geq, \leq, \in, \subset,$

$\cap, \cup, \equiv, \sim, \rightarrow$

$\forall, \exists$

Greek letters:  $\lambda, \pi, \Pi$

Sets N, R, Z in amssymb:  $\mathbb{N}, \mathbb{R}, \mathbb{Z}$

$\sum_{i=0}^{\infty} i \quad \prod_{i=0}^n i$

Binomial coefficient:  $\binom{x}{y}$

$$\begin{aligned} & x^2 \quad x_2 \\ & \frac{a}{b+c} \\ & \sqrt{x+y} \\ & \lfloor x+y \rfloor \lceil x+y \rceil \\ & \geq, \leq, \in, \subset, \\ & \cap, \cup, \equiv, \sim, \rightarrow \\ & \forall, \exists \\ & \lambda \pi \Pi \\ & \mathbb{N}, \mathbb{R}, \mathbb{Z} \\ & \sum_{i=0}^{\infty} i \prod_{i=0}^n i \\ & \binom{x}{y} \end{aligned}$$

And much more! See:

<http://www.ift.uib.no/Teori/KURS/WRK/TeX/symALL.html>

|               |             |             |           |             |           |            |          |
|---------------|-------------|-------------|-----------|-------------|-----------|------------|----------|
| $\alpha$      | \alpha      | $\theta$    | \theta    | $\circ$     | \circ     | $\tau$     | \tau     |
| $\beta$       | \beta       | $\vartheta$ | \vartheta | $\pi$       | \pi       | $\upsilon$ | \upsilon |
| $\gamma$      | \gamma      | $\gamma$    | \gamma    | $\varpi$    | \varpi    | $\phi$     | \phi     |
| $\delta$      | \delta      | $\kappa$    | \kappa    | $\rho$      | \rho      | $\varphi$  | \varphi  |
| $\epsilon$    | \epsilon    | $\lambda$   | \lambda   | $\varrho$   | \varrho   | $\chi$     | \chi     |
| $\varepsilon$ | \varepsilon | $\mu$       | \mu       | $\sigma$    | \sigma    | $\psi$     | \psi     |
| $\zeta$       | \zeta       | $\nu$       | \nu       | $\varsigma$ | \varsigma | $\omega$   | \omega   |
| $\eta$        | \eta        | $\xi$       | \xi       |             |           |            |          |
| $\Gamma$      | \Gamma      | $\Lambda$   | \Lambda   | $\Sigma$    | \Sigma    | $\Psi$     | \Psi     |
| $\Delta$      | \Delta      | $\Xi$       | \Xi       | $\Upsilon$  | \Upsilon  | $\Omega$   | \Omega   |
| $\Theta$      | \Theta      | $\Pi$       | \Pi       | $\Phi$      | \Phi      |            |          |

Table 1: Greek Letters

|           |                      |             |                        |                  |                               |            |                       |
|-----------|----------------------|-------------|------------------------|------------------|-------------------------------|------------|-----------------------|
| $\pm$     | <code>\pm</code>     | $\cap$      | <code>\cap</code>      | $\diamond$       | <code>\diamond</code>         | $\oplus$   | <code>\oplus</code>   |
| $\mp$     | <code>\mp</code>     | $\cup$      | <code>\cup</code>      | $\triangle$      | <code>\bigtriangleup</code>   | $\ominus$  | <code>\ominus</code>  |
| $\times$  | <code>\times</code>  | $\oplus$    | <code>\oplus</code>    | $\triangledown$  | <code>\bigtriangledown</code> | $\otimes$  | <code>\otimes</code>  |
| $\div$    | <code>\div</code>    | $\sqcap$    | <code>\sqcap</code>    | $\triangleleft$  | <code>\triangleleft</code>    | $\oslash$  | <code>\oslash</code>  |
| $*$       | <code>\ast</code>    | $\sqcup$    | <code>\sqcup</code>    | $\triangleright$ | <code>\triangleright</code>   | $\odot$    | <code>\odot</code>    |
| $\star$   | <code>\star</code>   | $\vee$      | <code>\vee</code>      | $\lhd^b$         | <code>\lhd^b</code>           | $\bigcirc$ | <code>\bigcirc</code> |
| $\circ$   | <code>\circ</code>   | $\wedge$    | <code>\wedge</code>    | $\rhd^b$         | <code>\rhd^b</code>           | $\dagger$  | <code>\dagger</code>  |
| $\bullet$ | <code>\bullet</code> | $\setminus$ | <code>\setminus</code> | $\unlhd^b$       | <code>\unlhd^b</code>         | $\ddagger$ | <code>\ddagger</code> |
| $\cdot$   | <code>\cdot</code>   | $\wr$       | <code>\wr</code>       | $\unrhd^b$       | <code>\unrhd^b</code>         | $\amalg$   | <code>\amalg</code>   |

|               |                          |               |                          |             |                        |             |                        |
|---------------|--------------------------|---------------|--------------------------|-------------|------------------------|-------------|------------------------|
| $\leq$        | <code>\leq</code>        | $\geq$        | <code>\geq</code>        | $\equiv$    | <code>\equiv</code>    | $\models$   | <code>\models</code>   |
| $\prec$       | <code>\prec</code>       | $\succ$       | <code>\succ</code>       | $\sim$      | <code>\sim</code>      | $\perp$     | <code>\perp</code>     |
| $\preceq$     | <code>\preceq</code>     | $\succeq$     | <code>\succeq</code>     | $\simeq$    | <code>\simeq</code>    | $\mid$      | <code>\mid</code>      |
| $\ll$         | <code>\ll</code>         | $\gg$         | <code>\gg</code>         | $\asymp$    | <code>\asymp</code>    | $\parallel$ | <code>\parallel</code> |
| $\subset$     | <code>\subset</code>     | $\supset$     | <code>\supset</code>     | $\approx$   | <code>\approx</code>   | $\bowtie$   | <code>\bowtie</code>   |
| $\subseteq$   | <code>\subseteq</code>   | $\supseteq$   | <code>\supseteq</code>   | $\approxeq$ | <code>\approxeq</code> | $\Join^b$   | <code>\Join^b</code>   |
| $\sqsubset^b$ | <code>\sqsubset^b</code> | $\sqsupset^b$ | <code>\sqsupset^b</code> | $\cong$     | <code>\cong</code>     | $\smile$    | <code>\smile</code>    |
| $\sqsubseteq$ | <code>\sqsubseteq</code> | $\sqsupseteq$ | <code>\sqsupseteq</code> | $\neq$      | <code>\neq</code>      | $\frown$    | <code>\frown</code>    |
| $\in$         | <code>\in</code>         | $\ni$         | <code>\ni</code>         | $\doteq$    | <code>\doteq</code>    | $=$         | <code>=</code>         |
| $\vdash$      | <code>\vdash</code>      | $\dashv$      | <code>\dashv</code>      | $<$         | <code>&lt;</code>      | $>$         | <code>&gt;</code>      |
| :             | :                        |               |                          |             |                        |             |                        |

|                      |                                 |                       |                                  |                |                           |
|----------------------|---------------------------------|-----------------------|----------------------------------|----------------|---------------------------|
| $\leftarrow$         | <code>\leftarrow</code>         | $\longleftarrow$      | <code>\longleftarrow</code>      | $\uparrow$     | <code>\uparrow</code>     |
| $\Leftarrow$         | <code>\Leftarrow</code>         | $\Longleftarrow$      | <code>\Longleftarrow</code>      | $\Uparrow$     | <code>\Uparrow</code>     |
| $\rightarrow$        | <code>\rightarrow</code>        | $\longrightarrow$     | <code>\longrightarrow</code>     | $\downarrow$   | <code>\downarrow</code>   |
| $\Rightarrow$        | <code>\Rightarrow</code>        | $\Longrightarrow$     | <code>\Longrightarrow</code>     | $\Downarrow$   | <code>\Downarrow</code>   |
| $\leftrightarrow$    | <code>\leftrightarrow</code>    | $\longleftrightarrow$ | <code>\longleftrightarrow</code> | $\Updownarrow$ | <code>\Updownarrow</code> |
| $\Leftrightarrow$    | <code>\Leftrightarrow</code>    | $\Longleftrightarrow$ | <code>\Longleftrightarrow</code> |                |                           |
| $\mapsto$            | <code>\mapsto</code>            | $\longmapsto$         | <code>\longmapsto</code>         | $\nearrow$     | <code>\nearrow</code>     |
| $\hookleftarrow$     | <code>\hookleftarrow</code>     | $\hookrightarrow$     | <code>\hookrightarrow</code>     | $\searrow$     | <code>\searrow</code>     |
| $\leftharpoonup$     | <code>\leftharpoonup</code>     | $\rightharpoonup$     | <code>\rightharpoonup</code>     | $\swarrow$     | <code>\swarrow</code>     |
| $\leftharpoondown$   | <code>\leftharpoondown</code>   | $\rightharpoondown$   | <code>\rightharpoondown</code>   | $\nwarrow$     | <code>\nwarrow</code>     |
| $\rightleftharpoons$ | <code>\rightleftharpoons</code> | $\leadsto^b$          |                                  |                |                           |
| $\ldots$             | <code>\ldots</code>             | $\cdots$              | <code>\cdots</code>              |                |                           |
| $\aleph$             | <code>\aleph</code>             | $\prime$              | <code>\prime</code>              | $\vdots$       | <code>\vdots</code>       |
| $\hbar$              | <code>\hbar</code>              | $\emptyset$           | <code>\emptyset</code>           | $\forall$      | <code>\forall</code>      |
| $\imath$             | <code>\imath</code>             | $\nabla$              | <code>\nabla</code>              | $\exists$      | <code>\exists</code>      |
| $\jmath$             | <code>\jmath</code>             | $\surd$               | <code>\surd</code>               | $\neg$         | <code>\neg</code>         |
| $\ell$               | <code>\ell</code>               | $\top$                | <code>\top</code>                | $\flat$        | <code>\flat</code>        |
| $\wp$                | <code>\wp</code>                | $\bot$                | <code>\bot</code>                | $\natural$     | <code>\natural</code>     |
| $\Re$                | <code>\Re</code>                | $\backslash$          | <code>\backslash</code>          | $\sharp$       | <code>\sharp</code>       |
| $\Im$                | <code>\Im</code>                | $\angle$              | <code>\angle</code>              | $\backslash$   | <code>\backslash</code>   |
| $\mho^b$             | <code>\mho^b</code>             |                       |                                  | $\partial$     | <code>\partial</code>     |
|                      | .                               | .                     | .                                |                |                           |

|           |                      |             |                        |              |                         |
|-----------|----------------------|-------------|------------------------|--------------|-------------------------|
| $\sum$    | <code>\sum</code>    | $\prod$     | <code>\bigcap</code>   | $\bigodot$   | <code>\bigodot</code>   |
| $\prod$   | <code>\prod</code>   | $\bigcup$   | <code>\bigcup</code>   | $\bigotimes$ | <code>\bigotimes</code> |
| $\coprod$ | <code>\coprod</code> | $\bigsqcup$ | <code>\bigsqcup</code> | $\bigoplus$  | <code>\bigoplus</code>  |
| $\int$    | <code>\int</code>    | $\bigvee$   | <code>\bigvee</code>   | $\biguplus$  | <code>\biguplus</code>  |
| $\oint$   | <code>\oint</code>   | $\wedge$    | <code>\bigwedge</code> |              |                         |

Table 7: Variable-sized Symbols

|                      |                    |                   |                   |                      |                      |                   |                    |
|----------------------|--------------------|-------------------|-------------------|----------------------|----------------------|-------------------|--------------------|
| <code>\arccos</code> | <code>\cos</code>  | <code>\csc</code> | <code>\exp</code> | <code>\ker</code>    | <code>\limsup</code> | <code>\min</code> | <code>\sinh</code> |
| <code>\arcsin</code> | <code>\cosh</code> | <code>\deg</code> | <code>\gcd</code> | <code>\lg</code>     | <code>\ln</code>     | <code>\Pr</code>  | <code>\sup</code>  |
| <code>\arctan</code> | <code>\cot</code>  | <code>\det</code> | <code>\hom</code> | <code>\lim</code>    | <code>\log</code>    | <code>\sec</code> | <code>\tan</code>  |
| <code>\arg</code>    | <code>\coth</code> | <code>\dim</code> | <code>\inf</code> | <code>\liminf</code> | <code>\max</code>    | <code>\sin</code> | <code>\tanh</code> |

Table 8: Log-like Symbols

Table 9: Delimiters

```
\rmoustache } \lmoustache ) \rgroup ( \lgroup  
|\arrowvert || \Arrowvert | \bracevert
```

**Table 10: Large Delimiters**

|             |                        |             |                        |           |                      |                  |                             |             |                        |             |                        |
|-------------|------------------------|-------------|------------------------|-----------|----------------------|------------------|-----------------------------|-------------|------------------------|-------------|------------------------|
| $\hat{a}$   | <code>\hat{a}</code>   | $\acute{a}$ | <code>\acute{a}</code> | $\bar{a}$ | <code>\bar{a}</code> | $\dot{a}$        | <code>\dot{a}</code>        | $\ddot{a}$  | <code>\ddot{a}</code>  | $\breve{a}$ | <code>\breve{a}</code> |
| $\check{a}$ | <code>\check{a}</code> | $\grave{a}$ | <code>\grave{a}</code> | $\vec{a}$ | <code>\vec{a}</code> | $\ddot{\vec{a}}$ | <code>\ddot{\vec{a}}</code> | $\tilde{a}$ | <code>\tilde{a}</code> |             |                        |

Table 11: Math mode accents

|                       |                                  |                        |                                   |
|-----------------------|----------------------------------|------------------------|-----------------------------------|
| $\widetilde{abc}$     | <code>\widetilde{abc}</code>     | $\widehat{abc}$        | <code>\widehat{abc}</code>        |
| $\overleftarrow{abc}$ | <code>\overleftarrow{abc}</code> | $\overrightarrow{abc}$ | <code>\overrightarrow{abc}</code> |
| $\overline{abc}$      | <code>\overline{abc}</code>      | $\underline{abc}$      | <code>\underline{abc}</code>      |
| $\overbrace{abc}$     | <code>\overbrace{abc}</code>     | $\underbrace{abc}$     | <code>\underbrace{abc}</code>     |
| $\sqrt{abc}$          | <code>\sqrt{abc}</code>          | $\sqrt[n]{abc}$        | <code>\sqrt[n]{abc}</code>        |
| $f'$                  | <code>f'</code>                  | $\frac{abc}{xyz}$      | <code>\frac{abc}{xyz}</code>      |

Table 12: Some other constructions

# Equations: Example 1

- You type:

```
\begin{equation}
P_{\mathrm{gal}}(k) = \frac{1+Qk^2}{1+Ak} P_{\mathrm{lin}}(k),
\end{equation}
```

- You get:

$$P_{\mathrm{gal}}(k) = \frac{1 + Qk^2}{1 + Ak} P_{\mathrm{lin}}(k), \quad (3.1)$$

# Lists of equations

- Aligning lists of equations is tricky (similar to tables)
- *eqnarray* uses & alignment hooks into the equations
- Don't need to use \$\$ or  $\begin{bmatrix} \end{bmatrix}$ . Already in math mode.

```
\begin{eqnarray}
f(x,y) &=& x + y \\
&=& 4x - 3x + y \\
&=& y (4 \frac{x}{y} - 3 \frac{x}{y} + 1)
\end{eqnarray}
```

Use *eqnarray\** to suppress numbering

# Equations: Example 2

- You type:
- Refer to it as \ref{e:3}

```
\begin{eqnarray}
1 - n_{\rm s} & = & 2 \epsilon_1 + \epsilon_2 \label{e:3} \\
r & = & 16 \epsilon_1 \label{eq4} \\
\end{eqnarray}
```



- You get:

$$1 - n_s = 2\epsilon_1 + \epsilon_2 \quad (3.3)$$

$$r = 16\epsilon_1. \quad (3.4)$$

# Algorithms

- \usepackage[option]{algorithm2e}
- <http://www.ctan.org/tex-archive/macros/latex/contrib/algoritm2e/algoritm2e.pdf>
- (there are many other packages too)

```
\begin{algorithm}[H]
  \SetLine
  \KwData{this text}
  \KwResult{how to write algorithm with \LaTeX2e }

  initialization;
  \While{not at end of this document}{
    read current;
    \eIf{understand}{
      go to next section;
      current section becomes this one;
    }{
      go back to the beginning of current section;
    }
  }
  \caption{How to write algorithms}
\end{algorithm}
```

which gives

```
Data: this text
Result: how to write algorithm with \LaTeX2e
initialization;
while not at end of this document do
  read current section;
  if understand then
    | go to next section;
    | current section becomes this one;
  else
    | go back to the beginning of current section;
  end
end
```

Algorithm 1: How to write algorithms



# Algorithms

```
\begin{algorithm}
\dontprintsemicolon
\KwSty{Module: I-Give-Turn.Timeout.Region}
\;
\Indp
\KwSty{Context: DiP.I-give-turn}
\;
\Indp
\KwSty{Start: }
\ArgSty{Psyclone.Context:DiP.I-give-turn}
\;
\KwSty{Stop: }
\ArgSty{Psyclone.Context:DiP.*}
\;
\KwSty{Trigger: }
\;
\KwSty{TimeOut:}
\ArgSty{3000}
\;
\KwSty{PostOnTimeOut:}
\ArgSty{Unim.Sense.Timeout.I-Give-Turn}
\;
\KwSty{PostOnTrigger:}
\;
\BlankLine
\BlankLine
\caption{Configuration for I-Give-Turn Timeout Region}
\label{alg:regionIGT}
```

**Module: I-Give-Turn.Timeout.Region**  
**Context: DiP.I-give-turn**

**Start:** *Psyclone.Context:DiP.I-give-turn*

**Stop:** *Psyclone.Context:DiP.\**

**Trigger:**

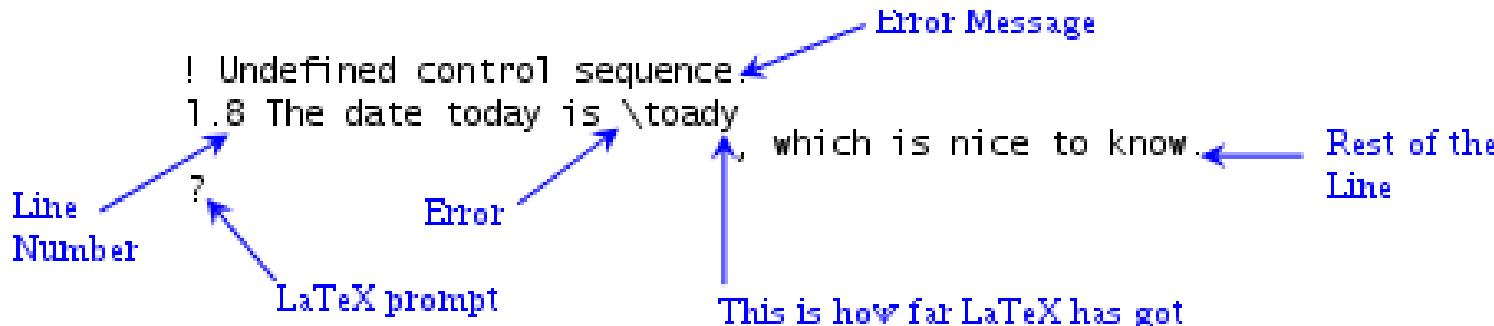
**TimeOut:** *3000*

**PostOnTimeOut:** *Unim.Sense.Timeout.I-Give-Turn*

**PostOnTrigger:**

Figure 3.6: Configuration for Other-Gives-Turn-Decider (OGTD)

# Debugging Latex problems



Note the line number (and file) and look to see if you can understand what went wrong. Check for “typos” as in the above

If not, Google is your friend, ask the exact error message.  
<http://tex.stackexchange.com/> has a lot of answers..

Package mismatches can be difficult so if you post asking question, explain what system you are on and what packages.

# Simple macros

- `\newcommand{\myname}{MySystemName}`
- `\def\long\longcomment#1{}`
  
- `\newcommand{\bt}[1]{\textbf{#1}}`  
    `\bt{if}`
- `\newcommand{\vfrac}[2]{\ensuremath{\frac{\#1}{\#2}}}`  
    `\vfrac{a}{b}`

%raw tex macro

`\def\long\longcomment#1{}`

# LaTeX in PowerPoint

- TeXPoint – A LaTeX add-on for ppt and word
  - <http://texpoint.necula.org/> (\$)

$$\sum_{i=1}^{n^2} x^i \frac{2}{i+2}$$

- TeXclip – LaTeX to image (there are many ~tools)
  - <http://maru.bonyari.jp/texclip/texclip.php>

Full Latex instead of PPT.)

- Beamer slides using LaTeX
  - <http://bitbucket.org/rivanvx/beamer/wiki/Home>

# And more...

- Fonts, spacing, margins, bibliographies,
- footnotes, book formatting, chapters, ...
- LaTeX is powerful!
- See resources page and google around for help...

# Referencing in BiBtex/ LaTex

---

Making references directly in LaTeX  
(fish.tex)

```
\documentclass{article}
\begin{document}
\emph{My mother} is a \underline{fish} \cite{WF}.
\begin{thebibliography}{99}
\bibitem{WF}
William Falkner, \emph{As I Lay Dying}
\end{thebibliography}
\end{document}
```

Compiling the ‘tex’ file - program reads through your file ONCE

1

See equation \\\eqref{simpleMath}.

```
\begin{equation}
1+2=3 \label{simpleMath}
\end{equation}
```

?

?

# file.aux

Adds  
‘simpleMath’

# Compiling the ‘tex’ file - program reads through your file ONCE

2

```
See equation \  
eqref{simpleMath}.  
  
\begin{equation}  
1+2=3 \label{simpleMath}  
\end{equation}
```

Found  
it !

file.aux

Adds  
'simpleMath'

# Need to compile the document twice/thrice

- Because we are using symbolic references, e.g., \texttt{\\cite{WF}},
  - ◆ a second pass is necessary
- The **second pass** will resolve references but can change pages.. So might need 3 passes
- Personally I recommend using Make
- Our standard paper makefile

all:

```
pdflatex -interaction=nonstopmode -synctex=1 paper.tex  
bibtex paper  
pdflatex -interaction=nonstopmode -synctex=1 paper.tex  
pdflatex -interaction=nonstopmode -synctex=1 paper.tex
```

clean:

```
rm paper.pdf *.aux *.bbl *.blg *.log
```

```
\documentclass{article}
\begin{document}
\emph{My mother} is a \underline{fish} \
cite{WF}.
\begin{thebibliography}{99}
\bibitem{WF}
William Falkner, \emph{As I Lay Dying}
\end{thebibliography}
\end{document}
```

*My mother* is a fish [1].

## References

- [1] William Falkner, *As I Lay Dying*

# Problem with “raw bibliography” or BBL method

- User is burdened with deciding how to format article titles, journal names, proceeding references
- Difficult to reuse references in other documents

# Quick BibTex Tutorial

---

# BibTex Tutorial

- Why do you want to learn BibTex?
- Nuts and Bolts of BibTex
- Examples
  
- BibTeX is...
  - ◆ A plain-text file format and a program designed to work with LaTeX
  - ◆ bibliographical information (author name, journal title, date, etc) stored in the file
  - ◆ incorporates BibTeX file (.bib) into LaTeX documents

# Why you want to learn BibTex

- Complements LaTeX documents by managing bibliography and references
- Minimizes the drudgery of formatting, numbering, and referencing
- Disadvantage:
  - ◆ Steep Learning Curve if you start from scratch
- but plenty of tools/source make it easy to use

# Sample Bibliographic Entry

```
@article{Hemmendinger07,
author = {David Hemmendinger},
title = {The ACM and IEEE-CS guidelines for
undergraduate CS education},
journal = {Commun. ACM},
volume = {50},
number = {5},
year = {2007},
issn = {0001-0782},
pages = {46--53},
doi = {http://doi.acm.org/10.1145/1230819.1230838},
publisher = {ACM},
address = {New York NY USA}
```

➤ @article lets BibTeX know that the bibliographical entry is an article.

## ➤ Other Types:

- ◆ book
- ◆ phdthesis
- ◆ unpublished
- ◆ misc
- ◆ and others

➤ Hemmendinger07 identifies the entry – used when citing it in a LaTeX document



# The BibTex Process

1. Create a **BibTex** file with Reference entries (or get from Zotero etc.. But be careful about Bad entries from google scholar)
2. Get a **\* .bst** file (bibliographic style file)
3. “Compile” or “Build” your **LaTeX** document to create an **\* .aux** file
4. Run **BibTeX** on your **LaTeX** file
5. Run **LaTeX** twice on your updated file
6. View the **dvi** or **pdf** file

# Types of Documents BibTex can handle

- ARTICLE
- BOOK
- BOOKLET
- INBOOK
- INCOLLECTION
- INPROCEEDINGS
- MANUAL
- MISC
- PHDTHESES
- PROCEEDINGS
- TECHREPORT
- UNPUBLISHED

# Each Document type can have the following entries

- address
- author
- booktitle
- chapter
- crossref
- edition
- editor
- howpublished
- institution
- journal
- key
- language
- month
- note
- number
- organization
- pages
- publisher
- school
- series
- title
- type="Ph.D. dissertation"
- volume
- year

Sample BibTex  
Book Entry  
(mybib.bib)

```
@BOOK{Press,  
author="W.H. Press",  
title="Numerical Recipes in C: The Art  
of Scientific Computing",  
publisher="Cambridge University Press",  
year=1992,  
}
```

## Sample BibTex: Article Entry

```
@ARTICLE{Vodacek,  
author="Anthony Vodacek and F.E. Hoge and R.N. Swift and  
J.K. Yungel and E.T. Peltzer and N.V. Blough",  
title="The use of in situ and airborne fluorescence  
measurements to determine {UV} absorption coefficients  
and {DOC} concentrations in surface waters",  
journal="Limnology and Oceanography",  
volume=40,  
number=2,  
year=1995,  
pages="411--415",  
}  
% article is the most commonly abused type If  
you get one from somewhere check if its valid
```

Sample BibTex  
Technical Report Entry

```
@TECHREPORT{Berk,
author="Lex A. Berk and L.S.
        Bernstein and D.C. Robertson",
title="MODTRAN: a moderate
        resolution model for LOWTRAN 7",
number="GL-TR-89-0122",
institution="Spectral Science",
address = "Burlington, MA",
year = 1989
}
```

Sample BibTex  
Ph.D. Dissertation Entry

```
@PHDTHESES{Kuo,  
author="Jan-Tai Kuo",  
title="The Influence of  
Hydrodynamic Transport on  
Phytoplankton Dynamics in  
Homogeneous Lakes",  
school="Cornell University",  
address="Ithaca, NY",  
year=1981,  
}
```

# Sample BibTex Master's Thesis Entry

```
@MASTERSTHESIS{Knobelspiesse,  
author="Kirk D. Knobelspiesse",  
title="Atmospheric Compensation for SeaWiFS  
Images of Lake Superior Utilizing Spatial  
Information",  
school="Rochester Institute of Technology",  
address="Rochester, NY",  
month=Sep,  
year=2000,  
}
```

Sample BibTex  
Proceedings Entry

```
@INPROCEEDINGS{Stoermer,  
author="E.F. Stoermer",  
title="Nearshore phytoplankton  
populations in the Grand Haven,  
Michigan vicinity during thermal bar  
conditions",  
booktitle="Proceedings of the 11th  
Conference on Great Lakes Research",  
pages="137--150",  
year=1968,  
}
```

# Different Formatting Styles

- To invoke a particular style, go to
  - ◆ <http://www.ctan.org>
- Document styles are defined in **\*.sty** files
  - ◆ **mla.sty**
  - ◆ **bmsplain.sty**
- Bibliography styles are defined in **\*.bst** files
  - ◆ **mla.bst**
  - ◆ **amsplain.bst**
  - ◆ **ieee.bst**

# To invoke these styles in your document

- Copy them to your current working directory with your **LaTeX** and **BibTex** document
- Edit your **LaTeX** file to appropriately reference these style guides

Sample **\LaTeX** document listing Bibliography in  
MLA Style (`text.tex`)

```
\documentclass{report}
\usepackage{mla}
\begin{document}

\bibliographystyle{mla}
\bibliography{mybib}
\nocite{*}

\end{document}
```

Commands to Build and  
View Document

**To generate pdf**

```
% pdflatex test.tex  
% bibtex test  
% pdflatex test.tex  
% pdflatex test.tex  
%
```

# Works Cited

- Berk, Lex A., L.S. Bernstein, and D.C. Robertson. MODTRAN: a moderate resolution model for LOWTRAN 7. Technical Report GL-TR-89-0122, Spectral Science, Burlington, MA, 1989.
- Knobelspiesse, Kirk D. Atmospheric Compensation for SeaWiFS Images of Lake Superior Utilizing Spatial Information. Master's thesis, Rochester Institute of Technology, September 2000.
- Kuo, Jan-Tai. *The Influence of Hydrodynamic Transport on Phytoplankton Dynamics in Homogeneous Lakes*. Ph.D. dissertation, Cornell University, Ithaca, NY, 1981.
- Press, W.H. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- RSI. *ENVI User's Guide*. Research Systems Incorporated, Boulder, CO, September 2001.
- Sherwood, D.A. Phosphorus Loads Entering Long Pond, A Small Embayment of Lake Ontario near Rochester, New York. USGS Fact Sheet 128-99, November 1999.
- Stoermer, E.F. "Nearshore phytoplankton populations in the Grand Haven, Michigan vicinity during thermal bar conditions." *Proceedings of the 11th Conference on Great Lakes Research*. 1968, 137–150.
- Vodacek, Anthony, F.E. Hoge, R.N. Swift, J.K. Yungel, E.T. Peltzer, and N.V. Blough. "The use of in situ and airborne fluorescence measurements to determine UV absorption coefficients and DOC concentrations in surface waters." *Limnology and Oceanography* 40 (1995): 411–415.

Sample **\LaTeX** document listing Bibliography in  
AMS Style (test2.tex)

```
\documentclass{report}
\begin{document}

\bibliographystyle{amsplain}
\bibliography{mybib}
\nocite{*}

\end{document}
```

# Bibliography

- [1] Kirk D. Knobelspiesse, *Atmospheric compensation for seaWiFS images of lake superior utilizing spatial information*, Master's thesis, Rochester Institute of Technology, September 2000.
- [2] Jan-Tai Kuo, *The influence of hydrodynamic transport on phytoplankton dynamics in homogeneous lakes*, Ph.D. thesis, Cornell University, Ithaca, NY, 1981.
- [3] W.H. Press, *Numerical recipes in c: The art of scientific computing*, Cambridge University Press, 1992.
- [4] RSI, *Envi user's guide*, Research Systems Incorporated, Boulder, CO, September 2001.
- [5] D.A. Sherwood, *Phosphorus loads entering long pond, a small embayment of lake ontario near rochester, new york*, USGS Fact Sheet 128-99, November 1999.
- [6] E.F. Stoermer, *Nearshore phytoplankton populations in the gulf haven, michigan vicinity during thermal bar conditions*, Proceedings of the 11th Conference on Great Lakes Research, 1968, pp. 137–150.
- [7] Anthony Vodacek, F.E. Hoge, R.N. Swift, J.K. Yungel, E.T. Peltzer, and N.V. Blough, *The use of in situ and airborne fluorescence measurements to determine uv absorption coefficients and doc concentrations in surface waters*, Limnology and Oceanography **40** (1995), no. 2, 411–415.

To include only cited articles we remove the `\nocite{*}` command

```
\documentclass{report}
```

```
\begin{document}
```

We want to get the documents `\cite{RSI}` and `\cite{Kuo}` for our collection

```
\bibliographystyle{amsplain}
```

```
\bibliography{mybib}
```

```
\end{document}
```

We want to get the documents [2] and [1] for our collection.



## Bibliography

- [1] Jan-Tai Kuo, *The influence of hydrodynamic transport on phytoplankton dynamics in homogeneous lakes*, Ph.D. thesis, Cornell University, Ithaca, NY, 1981.
- [2] RSI, *Enviro user's guide*, Research Systems Incorporated, Boulder, CO, September 2001.

# Cross-referencing in Proceedings

```
@PROCEEDINGS{Narayanan,  
editor="Ram Mohan Narayanan and James E.  
Kalshoven, Jr.",  
title="Advances in Laser Remote Sensing for  
Terrestrial and Oceanographic  
Applications",  
booktitle="Proceeding of SPIE",  
publisher="SPIE",  
volume=3059,  
year=1997,  
}
```

# Article in Proceedings

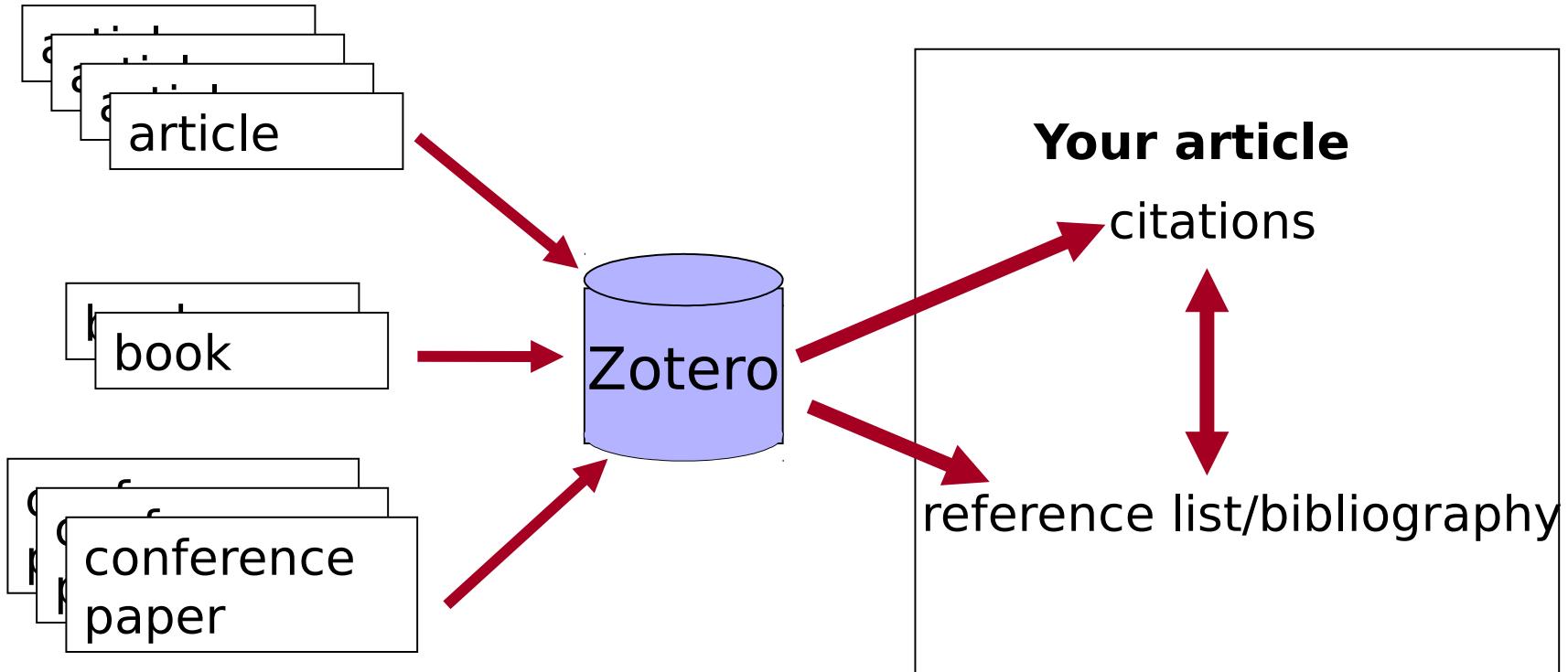
```
@INPROCEEDINGS{Ulrich,  
author="B.L. Ulrich and P. Lacovara and S.E.  
Moran and M.J. DeWeert",  
title="Recent results in imaging lidar",  
crossref="Narayanan",  
pages="95--108",  
}
```

# Works Cited

Narayanan, Ram Mohan and Jr. James E. Kalshoven, editors. *Advances in Laser Remote Sensing for Terrestrial and Oceanographic Applications*, SPIE, 1997, Volume 3059 .

Ulrich, B.L., P. Lacovara, S.E. Moran, and M.J. DeWeert. "Recent results in imaging lidar." See Narayanan and James E. Kalshoven (1997), 1997, 95–108.

# Use a reference management tool!



# Bigger documents/Better team work

- Latex support “include/input” which can improve workflow
- \input{introduction}
- \input{related}
- \input{approach}
- \input{experiments}
- \input{discussion}
- \input{conclusion}

RingLoss RingLoss

Ab Review History Chat

Menu

Source Rich Text

Recompile

Overfull \hbox (9.99998pt too wide) in our\_approach.tex, line 222

Overfull \vbox (12.9501pt too high) detected at line 57

BibTeX there's a number but no volume in chow1957optimum

BibTeX empty publisher in krizhevsky2012alexnet

BibTeX can't use both volume and number fields in netzer2011SVHN

BibTeX can't use both author and editor fields in phillips2011evaluation

BibTeX empty journal in sermanet2013overfeat

View Raw Logs

Other logs & files

```
1 \section{Introduction and Problem Formulation}
2 %MG{Generally: when adding new citations, please use
3 % the standard way for citation keys: lastnameYEARkeyword;
4 % I have changed all of the current citation keys}
5 
6 Ever since a convolutional neural network (CNN)
7 \cite{krizhevsky2012alexnet} won the ImageNet Large
8 Scale Visual Recognition Challenge (ILSVRC) in 2012
9 \cite{russakovsky2013avocados}, the extraordinary
10 increase in the performance of deep learning
11 architectures has contributed to the growing application
12 of computer vision algorithms.
13 
14 Many of these algorithms presume detection before
15 classification or directly belong to the space of
16 detection algorithms, ranging from object detection
17 \cite{girshick2014rcnn,girshick2015fast,ren2015faster,li
18 u2016ssd,redmon2016yolo},
19 
20 %MG{\cite{yolov3} is an arxiv paper stating \emph{an
21 %incremental improvement}. Do you really want to cite
22 %that?},
23 
24 face detection \cite{hu2017tinyfaces}, pedestrian
25 detection \cite{zhang2018pedestrian} %,
26 %MG{\cite{wang2018sface,zhu2018seeing} are arxiv
27 %papers. Do you really want to cite them?}
28 
29 %text detection \MG{If you do not have a citation for
30 %text detection, you might want to remove it from the
31 %list}
32 
33 and alike.
34 
35 Interestingly, though each year new
36 state-of-the-art-algorithms emerge from each of these
37 domains, a crucial component of their architecture
38 spacebelow=6pt,%
```



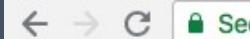
Menu



## GitHub Sync



&gt; Images



Download



Source

Actions

Copy Project

Word Count

Sync

Dropbox

GitHub

Mendeley

Settings



This project is synced with the GitHub repository at [Vastlab/RingLoss](#)

### Recent commits in GitHub

[TB adds PR-AUC plot](#)

96540a

by Terry Boult <[tboult@vast.uccs.edu](mailto:tboult@vast.uccs.edu)>[TB adds revised PR-AUC plot \(without fusion\)](#)

5ba7cf

by Terry Boult <[tboult@vast.uccs.edu](mailto:tboult@vast.uccs.edu)>[TB adds revised PR-AUC plot \(without fusion\)](#)

d06b0e

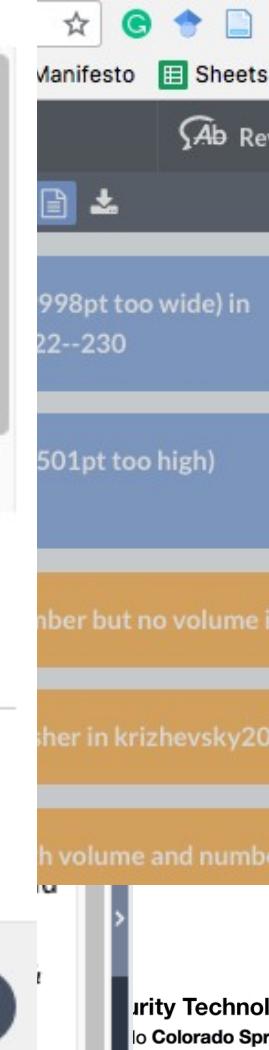
by Terry Boult <[tboult@vast.uccs.edu](mailto:tboult@vast.uccs.edu)>[Increase legend add spac and kaggle for figure](#)

da8657

by Terry Boult <[tboult@vast.uccs.edu](mailto:tboult@vast.uccs.edu)>[Pull GitHub changes into Overleaf v2](#)[Push Overleaf v2 changes to GitHub](#)

f142a0

b142a0

[Close](#)

# Resources

➤ For TeX (with lots of latex as well) :

➤ <http://www.tug.org/>

➤ LaTex project page

➤ <http://www.latex-project.org/>

➤ Not so short introduction to LaTeX:

<http://ctan.tug.org/tex-archive/info/lshort/english/lshort.pdf>

➤ Latex Video Tutorials

<https://www.youtube.com/watch?v=SoDv0ghyysQ&list=PL1D4EAB31D3EBC449>

➤ Can read the LaTex source code (its written in TeX)

➤ Google is your friend

# Your personal research assistant

Zotero is a free, easy-to-use tool to help you collect, organize, cite, and share research.

[Download](#)

Available for Mac, Windows, and Linux

New: Just need to create a quick bibliography? Try [ZoteroBib](#).

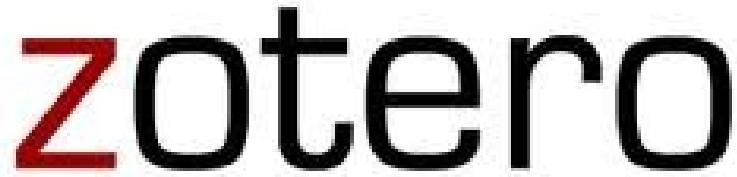


# Reference management tools

- Tools that help scholars to create and manage their lists of references for research projects.
- Most tools are designed to organize citations into specific formats for the preparation of manuscripts and bibliographies.
- Many search tools provide ways to download references into reference management tools.

# Software options

|              |   |
|--------------|---|
| Free         | <br>  |
| Subscription | <br>Thomson Reuters<br><br><br> |



- Zotero is a free (open source) reference management tool, that is available as an add-on for the Firefox web browser.
- Standalone version for Chrome, Safari and Firefox browsers
- To download and install Zotero go to:  
<http://www.zotero.org>

# Zotero functionality

- Zotero recognizes bibliographic information on books, journal articles, and other resources from websites and databases and extracts metadata from these sources.
- Store related PDFs, files, images, and links in your library
- Create a bibliography in Word or OpenOffice or export bib file for latex.
- Publish shared libraries for your organization, class, or research group

[Home](#) > Downloads



## Zotero 5.0 for Mac

Your personal research assistant

[Download](#)

Other platforms

[Windows](#) · [Linux 32-bit](#) · [Linux 64-bit](#)



## Zotero Connector

Save to Zotero from your browser

[Install Chrome Connector](#)

Zotero Connectors automatically sense content as you browse the web and allow you to save it to Zotero with a single click.

Note that there are Download options: [Zotero/Firefox](#) and [Zotero Standalone](#) version for [Firefox](#), [Chrome](#) and [Safari](#).



## Plugins

Install one of the many third-party plugins and become even more productive.  
[Browse Plugins](#)



## My Library

- diabetes and developing countries
- Information Literacy AND physicians
- Duplicate Items
- Unfiled Items
- Trash

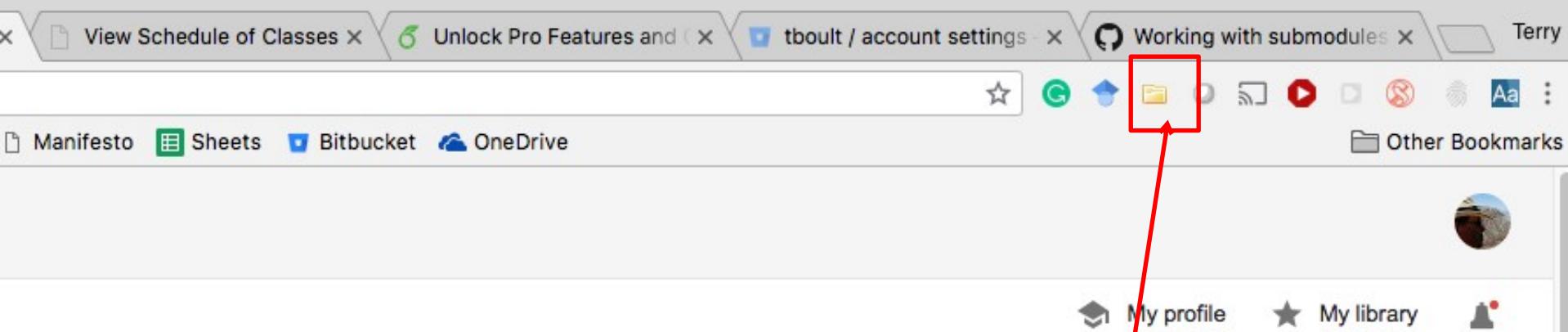
| Title                             | Creator           |
|-----------------------------------|-------------------|
| The Architecture of Risk for T... | Abdullah et al.   |
| Noncommunicable diseases ...      | Amara and Aljunid |
| A Demonstration of Peer Su...     |                   |
| Depression and type 2 diabe...    |                   |
| Present and future associat...    |                   |
| Ectopic Fat, Insulin Resist...    |                   |
| Randomised controlled tria...     |                   |
| High Prevalence of Abdomin...     | Ezenwaka et al.   |
| Beta cell imaging - a key too...  | Gotthardt et al.  |
| Global estimates of diabetes ...  | Guariguata et al. |
| Suboptimal glycaemic and bl...    | Kibirige et al.   |
| Predominance of multi-resist...   | Moremi et al.     |
| Cluster randomised controlle...   |                   |

Open the Standalone version by clicking on the icon that, for this computer, is listed on the bottom taskbar.

13 items in this view

No tags to display





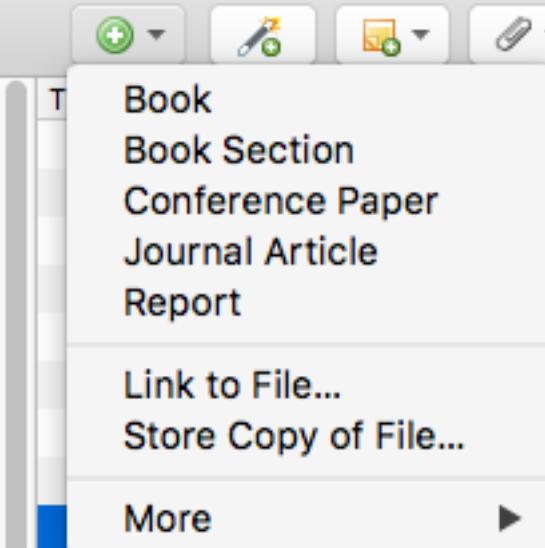
Zotero has been loaded and is accessible by clicking on the folder in chrome toolbar.

Registering for Zotero is optional.

[PDF] [researchgate.net](#)

# Adding citations to Zotero

- You can add citations to your library in several ways:
  - ◆ Manually enter citation information
  - ◆ Import citation information for a single item
  - ◆ Import citation information for multiple items
  - ◆ Automatically cite web pages
  - ◆ Add an item by identifier (ISBN, DOI or PMID)



Info Notes Tags Related

Item Type Conference Paper

Title

Author (last), (first)

Abstract

Date

Proceedings Title

Conference Name

Place

Publisher

Volume

Pages

Series

Language

DOI

ISBN

Short Title

URL

Accessed

Archive

Loc. in Archive

Library Catalog

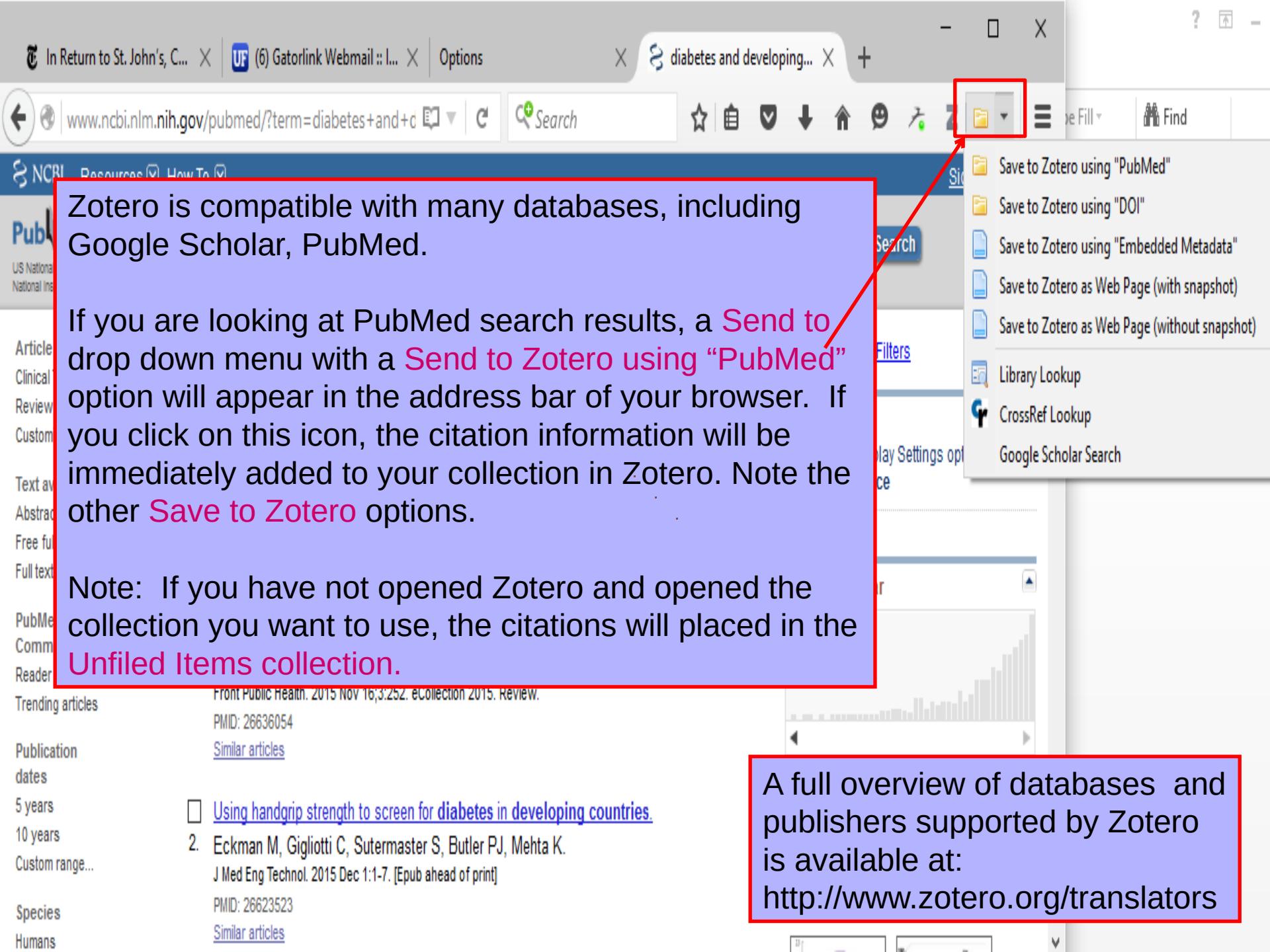
Call Number

Zotero is compatible with many databases, including Google Scholar, PubMed.

If you are looking at PubMed search results, a **Send to** drop down menu with a **Send to Zotero using "PubMed"** option will appear in the address bar of your browser. If you click on this icon, the citation information will be immediately added to your collection in Zotero. Note the other **Save to Zotero** options.

Note: If you have not opened Zotero and opened the collection you want to use, the citations will placed in the **Unfiled Items collection**.

A full overview of databases and publishers supported by Zotero is available at:  
<http://www.zotero.org/translators>



The screenshot shows a web browser window with several tabs open. The active tab is displaying search results from PubMed for the query "diabetes and developing". A context menu is open on the right side of the screen, triggered by a Zotero extension icon in the toolbar. The menu is titled "Zotero" and includes the following options: "Save to Zotero using 'PubMed'", "Save to Zotero using 'DOI'", "Save to Zotero using 'Embedded Metadata'", "Save to Zotero as Web Page (with snapshot)", "Save to Zotero as Web Page (without snapshot)", "Library Lookup", "CrossRef Lookup", and "Google Scholar Search". A red arrow points from the text in the main content area to the "Save to Zotero using 'PubMed'" option in the menu. A red box also highlights the "Save to Zotero using 'PubMed'" option and the Zotero extension icon in the toolbar. The left sidebar of the browser shows various search filters and settings.

## User profiles for author:T.Boult



Terrance E. Boult

El Pomar Prof. of Innovation and Security, University of Colorado Colorado Springs

Verified email at vast.uccs.edu

Cited by 8610

### Constraining object features using a polarization reflectance model

LB Wolff, TE Boult - IEEE Transactions on Pattern Analysis & Machine ..., 1991 - computer.org

ABSTRACT< p> The authors present a polarization reflectance model that uses the Fresnel reflection coefficients. This reflectance model accurately predicts the magnitudes of polarization components of reflected light, and all the polarization-based methods presented ...

☆ 99 Cited by 437 Related articles All 11 versions

Cracking

WJ Scheir

This paper

biometrics

published

☆ 99

Separat

SK Nayar,

Specular r

These high

binocular s

☆ 99

## Is Zotero Running?

The Zotero Connector was unable to communicate with the Zotero desktop application. The Connector can save some pages directly to your zotero.org account, but for best results you should make sure Zotero is open before attempting to save.

You can [download Zotero](#) or [troubleshoot the connection](#) if necessary.

Enable Saving to Online Library

Cancel

Try Again

Quality

LP Jain, W

Abstract! Use of binary model helps discriminate positive class from known negatives! Use of single-class probability model prevents misclassification of unknowns! Does not directly use negative data for probability estimation! Does not assume any particular distribution ...

☆ 99 Cited by 196 Related articles All 4 versions

### Registration of planar film radiographs with computed tomography

JMG Brown, TE Boult - mmbia, 1996 - computer.org

[PDF] researchgate.net

[PDF] springer.com

[PDF] uccs.edu

CLICK folder to import to zotero

[https://scholar.google.com/scholar?hl=en&as\\_sd=0%2C6&q=author%3AT.Boult&btnG=](https://scholar.google.com/scholar?hl=en&as_sd=0%2C6&q=author%3AT.Boult&btnG=)

gle 17 Cal Scholar Contacts myUCCS Canvas Onelook Maps CoTrip WMeet Manifesto Sheets Other Bookmarks

olar

author:T.Boult



About 604 results (0.04 sec)

My profile

My library



### User profiles for author:T.Boult



Terrance E. Boult

El Pomar Prof. of Innovation and Security, Univ

Verified email at vast.uccs.edu

Cited by 8610

### Constraining object features using a polar

LB Wolff, [TE Boult](#) - IEEE Transactions on Pattern An

ABSTRACT< p>The authors present a polarization reflection coefficients. This reflectance model accurate polarization components of reflected light, and all the p

☆ 99 Cited by 437 Related articles All 11 versio

### Cracking fuzzy vaults and biometric encry

WJ Scheirer, [TE Boult](#) - Biometrics Symposium, 2007,

This paper is a security analysis of leading privacy enh biometrics including biometric fuzzy vaults (BFV) and b published attacks, combined with various" proven" sec

☆ 99 Cited by 339 Related articles All 11 versio

Zotero Item Selector

Select which items you'd like to add to your library:

- Constraining object features using a polarization reflectance model
- Cracking fuzzy vaults and biometric encryption
- Separation of reflection components using color and polarization
- Quality of experience
- Registration of planar film radiographs with computed tomography
- Vision of the unseen: Current trends and challenges in digital image and video forensics
- Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor environments
- Error analysis of background adaption
- Separable image warping methods and systems using spatial lookup tables
- Factorization-based segmentation of motions

[Select All](#) [Deselect All](#) [Cancel](#) [OK](#)

### Separation of reflection components using color and polarization

[PDF] [springer.com](#)

SK Nayar, XS Fang, [T Boult](#) - International Journal of Computer Vision, 1997 - Springer

Specular reflections and interreflections produce strong highlights in brightness images.

These highlights can cause vision algorithms for segmentation, shape from shading, binocular stereo, and motion estimation to produce erroneous results. A technique is ...

☆ 99 Cited by 346 Related articles All 15 versions

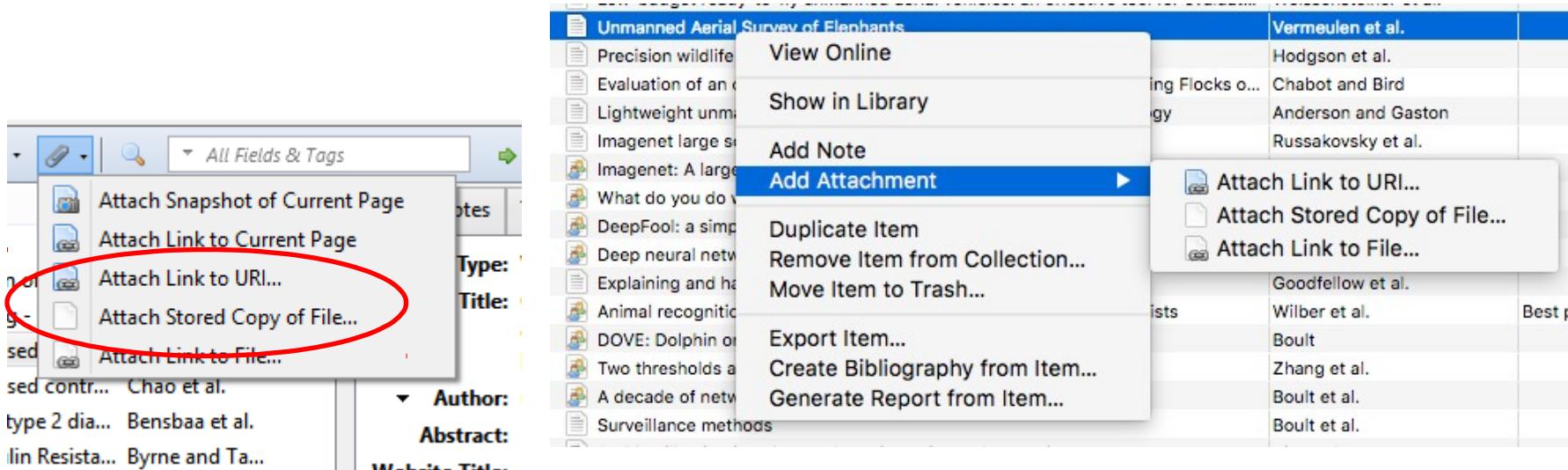
### Quality of experience

**UCCS**

Vision And Security Technology  
University of Colorado Colorado Springs

# Attaching files to items

- In Zotero, any type of file can be attached to an item.
- Items such as the PDF of an article or images can be opened in the Firefox window, while other file-types open in external programs.
- You can add attachments to items by clicking on the button in the middle column that looks like a paperclip:



- All files stored in the Zotero library are kept in a folder within the Zotero directory (unless another location is specified in

# Acknowledgments

Zotero

Everything

| Title  | Creator               | Extra                         |
|--|-----------------------|-------------------------------|
| Focal loss for dense object detection  | Lin et al.            | arXiv preprint arXiv:1708.... |
| Study of dynamic infrared scene projection technology based on Digital Micro...  | Hu et al.             |                               |
| Dynamic infrared scene projectors based upon the DMD                             | Beasley et al.        |                               |
| <b>Finding tiny faces</b>  | <b>Hu and Ramanan</b> | <b>1</b>                      |
| Development/testing on WIDER faces   |                       |                               |
| Joint face detection and alignment using multitask cascaded convolutional net... | Zhang et al.          |                               |
| Single image deblurring for a real-time face recognition system                  | Heflin et al.         |                               |
| A taxonomy of face-models for system evaluation                                  | Iyer et al.           |                               |
| FAAD: face at a distance   | Boult et al.          |                               |
| Long-range facial image acquisition and quality                                  | Boult and Scheirer    |                               |
| A taxonomy of face-models for system evaluation                                  | Iyer et al.           |                               |
| Face and eye detection on hard datasets  | Parris et al.         |                               |

Info Notes Tags Related

1 note: Add Development/testing on WIDER faces

Zotero

Everything

| Title   | Creator               | Extra                         |
|---|-----------------------|-------------------------------|
| Focal loss for dense object detection   | Lin et al.            | arXiv preprint arXiv:1708.... |
| Study of dynamic infrared scene projection technology based on Digital Micro...   | Hu et al.             |                               |
| Dynamic infrared scene projectors based upon the DMD                              | Beasley et al.        |                               |
| <b>Finding tiny faces</b>   | <b>Hu and Ramanan</b> | <b>1</b>                      |
| <b>Development/testing on WIDER faces</b>   |                       |                               |
| Joint face detection and alignment using multitask cascaded convolutional net...  | Zhang et al.          |                               |
| Single image deblurring for a real-time face recognition system                   | Heflin et al.         |                               |
| A taxonomy of face-models for system evaluation                                   | Iyer et al.           |                               |
| FAAD: face at a distance  | Boult et al.          |                               |
| Long-range facial image acquisition and quality                                   | Boult and Scheirer    |                               |
| A taxonomy of face-models for system evaluation                                   | Iyer et al.           |                               |
| Face and eye detection on hard datasets   | Parris et al.         |                               |
| Caffe: Convolutional architecture for fast feature embedding                      | Jia et al.            |                               |
| Going deeper with convolutions  | Szegedy et al.        |                               |
| What's Hiding in My Deep Features?  | Rudd et al.           |                               |
| Large scale unconstrained open set face database                                  | Sapkota and Boult     |                               |
| GRAB: generalized region assigned to binary                                       | Sapkota and Boult     |                               |
| DMD diffraction measurements to support design of projectors for test and eval... | Rice et al.           |                               |
| Considerations for DMDs operating in the infrared                                 | Dupuis and Mansur     |                               |
| Hyperspectral image projectors for radiometric applications                       | Rice et al.           |                               |
| Digital micromirror device (DMD) projector based test bench for vision chips      | Gergelyi and Földesy  |                               |

B I U S X<sub>2</sub> X<sup>2</sup> A A Paragraph T<sub>x</sub> “ ”

Development/testing on WIDER faces

best performing detector on UCCS challenge

integrated into face attribute demo

# Creating bibliographies from Zotero

- To create a bibliography from Zotero, highlight one or more references and then right-click to select **Create Bibliography from Selected Item**. Then select a **Citation Style** for your bibliography and choose one of the following four **Formats** to create your bibliography:
  - ◆ *Save as RTF* will allow you to save the bibliography as a rich text file, which can be opened for example in Word.
  - ◆ *Save as HTML* will allow you to save the bibliography as a HTML file for viewing in a web browser.
  - ◆ *Save to Clipboard* will allow you to save the bibliography to your clipboard to paste into any text field.

Can “export” references into bibtex file.

Zotero

File Edit Tools Window Help

New Item  
New Note  
New Collection...

Close ⌘W  
Import... ⌘I  
Import from Clipboard ⌘I  
Export Library...

My Library.bib

Users/tboult/Zotero/storage/5REEID2U/Reed et al. - 2007 - Impact of depression on re...

Zotero

My Library

- BioCell
- ENGAGE
- ITEST
- JanusUCCS
- MakeFrgile
- NSF-SCH-Proposal
- opennet\_cvpr16

**OpenSet**

- refs
- Thor
- My Publications
- Duplicate Items
- Unfiled Items
- Trash

Group Libraries

IMAGIN\_NSF\_ICURP\_14

- Duplicate Items
- Unfiled Items
- Trash

neuroNEX

RIS import 2D static image

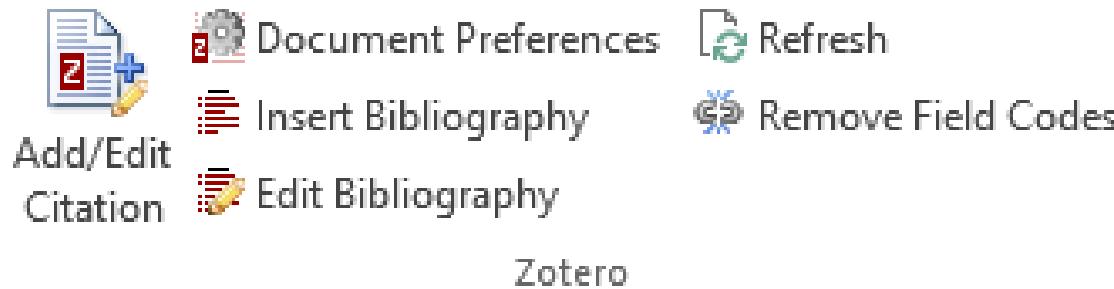
Title Creator

|  |                         |
|--|-------------------------|
| 24/7 security system: 60-FPS color EMCCD camera with integral human recogn...        | Vogelsong et al.        |
| Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A revie...   | Linchant et al.         |
| Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation          | Koh and Wich            |
| The use of conservation drones in ecology and wildlife research                      | Ivosevic et al.         |
| Wildlife research and management methods in the 21st century: Where do unm...        | Chabot and Bird         |
| Precision wildlife monitoring using unmanned aerial vehicles                         | Hodgson et al.          |
| Low-budget ready-to-fly unmanned aerial vehicles: an effective tool for evaluat...   | Weissensteiner et al.   |
| Unmanned Aerial Survey of Elephants  | Vermeulen et al.        |
| Precision wildlife monitoring using unmanned aerial vehicles                         | Hodgson et al.          |
| Evaluation of an off-the-shelf Unmanned Aircraft System for Surveying Flocks o...    | Chabot and Bird         |
| Lightweight unmanned aerial vehicles will revolutionize spatial ecology              | Anderson and Gaston     |
| Imagenet large scale visual recognition challenge                                    | Russakovsky et al.      |
| Imagenet: A large-scale hierarchical image database                                  | Deng et al.             |
| What do you do when you know that you don't know?                                    | Bendale and Boult       |
| DeepFool: a simple and accurate method to fool deep neural networks                  | Moosavi-Dezfooli et al. |
| Deep neural networks are easily fooled: High confidence predictions for unreco...    | Nguyen et al.           |
| Explaining and harnessing adversarial examples                                       | Goodfellow et al.       |
| Animal recognition in the Mojave Desert: Vision tools for field biologists           | Wilber et al.           |
| DOVE: Dolphin omnidirectional video equipment  | Roult                   |
| Hyperspectral image projectors for radiometric applications                          | Rice et al.             |
| Digital micromirror device (DMD) projector based test bench for vision chips         | Gergelyi and Földesy    |
| Generative adversarial network-based synthesis of visible faces from polarimetr...   | Zhang et al.            |
| Learning Face Age Progression: A Pyramid Architecture of GANs                        | Yang et al.             |
| Context-sensitive search using a deep learning model                                 | Guo et al.              |
| Multi-task deep convolutional neural networks for efficient and robust traffic la... | Mei et al.              |
| Training a neural network to detect objects in images                                | Erhan et al.            |

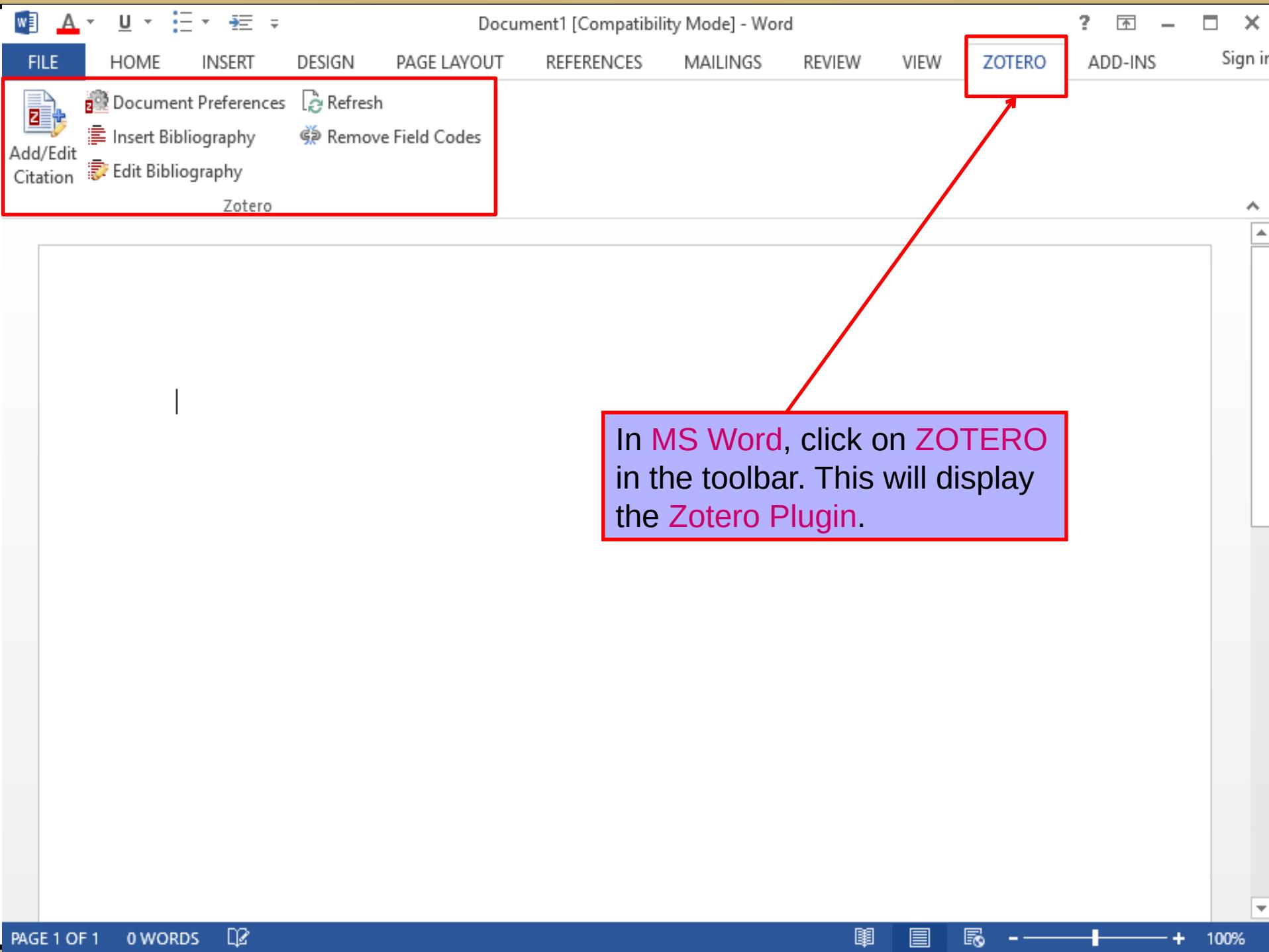
University of Colorado Colorado Springs

# Word Processor Integration

- Zotero's Word and OpenOffice plugins allow users to insert citations directly from their word processing software.
- You can find a link to install the plugin on the Zotero homepage ([www.zotero.org](http://www.zotero.org)).
- After installing the plugin and clicking on Zotero, you should see this row of icons in your Microsoft Word toolbar:



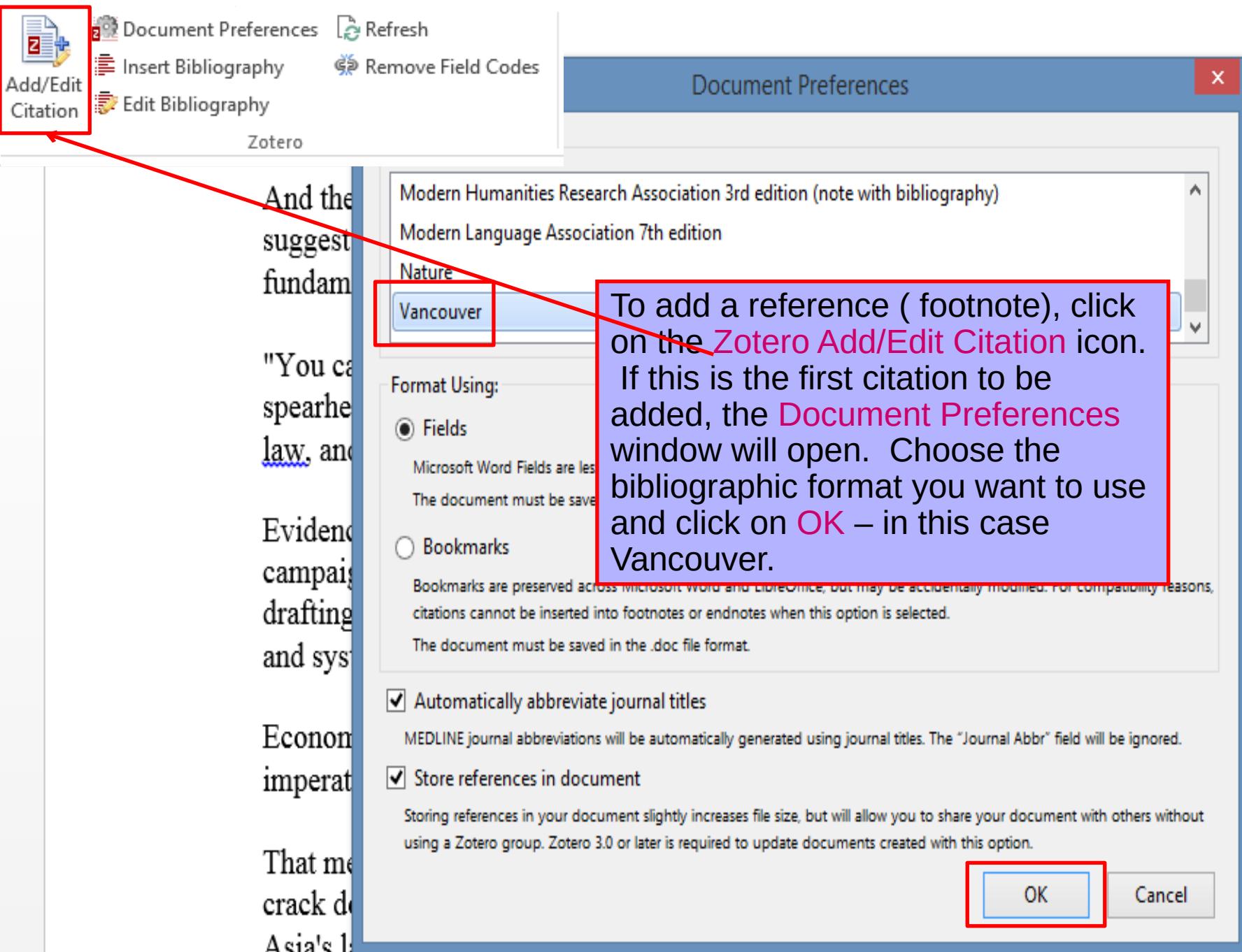
- These six buttons allow you to manage references and bibliographies in your Microsoft Word documents.

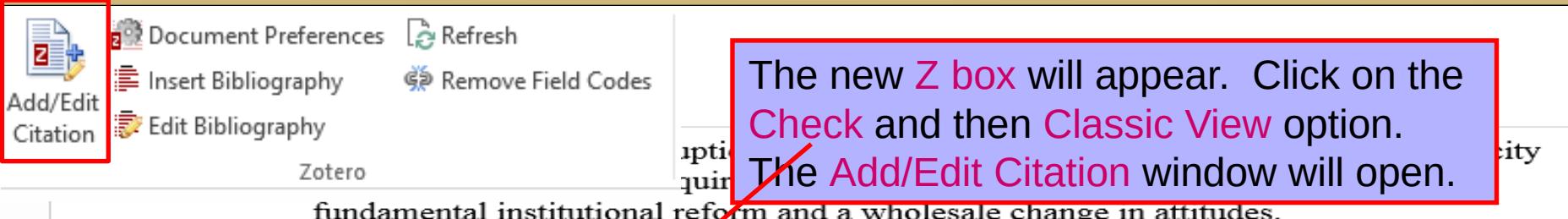




The key Zotero options  
are listed on this page.

- Zotero Add Citation/Edit Citation
- Zotero Insert Bibliography/Edit Bibliography
- Zotero Set Documents Preferences  
(to change the style)





fundamental institutional reform and a wholesale change in attitudes.

"You cannot fight corruption just by fighting corruption," said Daniel Kaufmann, who spearheaded the World Bank's efforts to improve the study of governance and the rule of law, and who estimates that \$1 trillion of bribes are paid every year.

Evidence campaign, drafting of and systemic governance reforms" were

Economists who specialize in governance imperative -- it is essential for promoting

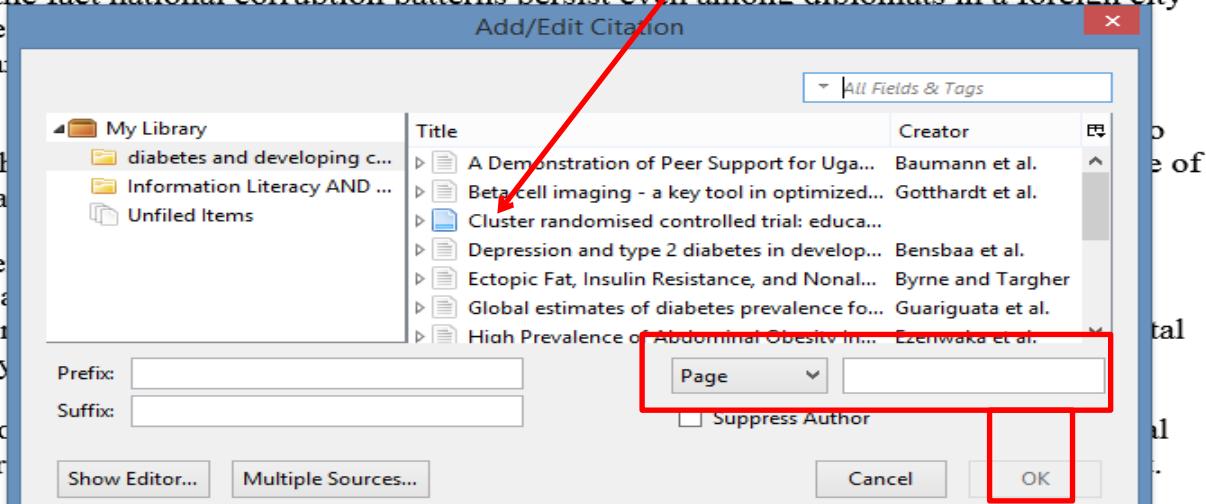
And the fact national corruption patterns persist even among diplomats in a foreign city  
sugge  
funda

"You  
spear  
law, a

Evide  
campa  
draftin  
and sy

Econo  
imper

In the Add/Edit Citation window,  
highlight the citation you want as  
a reference. Add the page # and  
click on OK.



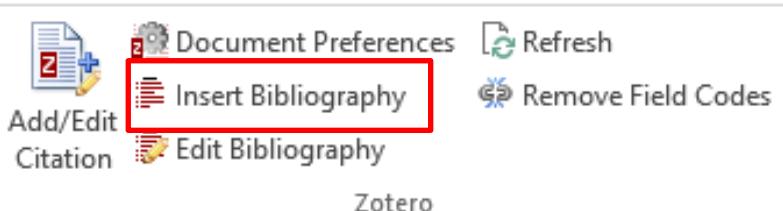
That means economies like Singapore and Hong Kong which have successfully sought to crack down on corruption have received real economic benefits in return. And southeast Asia's laggards have driven investors away because of their poor reputation.

We have added 3 Vancouver style references to this document.

"You cannot fight corruption just by fighting corruption," said Daniel Kaufmann, who spearheaded the World Bank's efforts to improve the study of governance and the rule of law, and who estimates that \$1 trillion of bribes are paid every year.(1)

Evidence showed there was little to be gained from "yet another anti-corruption campaign, the creation of more commissions and ethics agencies, and the incessant drafting of new laws, decrees, and codes of conduct," he said, adding that "fundamental and systemic governance reforms" were needed instead.(2)

Economists who specialize in governance say combating corruption is not just a moral imperative -- it is essential for promoting long-term economic growth and investment.(3)



To create a bibliography, click on the **Zotero Insert Bibliography** icon. This automatically adds the bibliography (in the Vancouver format) to the bottom of the document.

"You cannot fight corruption just by fighting corruption," said Daniel Kaufmann, who spearheaded the World Bank's efforts to improve the study of governance and the rule of law, and who estimates that \$1 trillion of bribes are paid every year.(1)

Evidence showed there was little to be gained from "yet a more commissions and ethics agencies, and the incessant conduct," he said, adding that "fundamental and systemic

Economists who specialize in governance say combating essential for promoting long-term economic growth and investment.(3)

1. Bensbaa S, Araab C, Boujraf S, Ajdi F. Depression and type 2 diabetes in developed and developing countries. Indian J Endocrinol Metab. 2014 Jan;18(1):117–8.
2. Ezenwaka CE, Okoye O, Esonwune C, Onuoha P, Dioka C, Osuji C, et al. High Prevalence of Abdominal Obesity Increases the Risk of the Metabolic Syndrome in Nigerian Type 2 Diabetes Patients: Using the International Diabetes Federation Worldwide Definition. Metab Syndr Relat Disord. 2014 Mar 6;
3. Abdullah N, Attia J, Oldmeadow C, Scott RJ, Holliday EG. The Architecture of Risk for Type 2 Diabetes: Understanding Asia in the Context of Global Findings. Int J Endocrinol. 2014;2014:593982.

Note: Do not use your word processor to edit citations. Any changes automatically will revert to the original as Zotero updates your document.

# Citation Styles

- A core feature of Zotero is its ability to automatically format citations and bibliographies.
- By default Zotero comes with a selection of popular styles (such as APA, Harvard , IEEand Vancouver), but many more are available through the online Zotero Style Repository at:  
<http://www.zotero.org/styles>
- Visit the Style Repository and click the “[Install]” link next to the style you wish to install.

# Zotero Citation Style Repository

## Zotero Style Repository

### Default Styles

These styles are distributed automatically to Zotero clients.

- [American Political Science Association](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [American Psychological Association 6th Edition](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [American Sociological Association](#) [\[Install\]](#) (2008-12-23 06:40:00)
- [Chicago Manual of Style \(Author-Date format\)](#) [\[Install\]](#) (2010-03-15 01:35:00)
- [Chicago Manual of Style \(Full Note with Bibliography\)](#) [\[Install\]](#) (2010-03-15 01:35:00)
- [Chicago Manual of Style \(Note with Bibliography\)](#) [\[Install\]](#) (2010-03-15 01:35:00)
- [Chicago Manual of Style \(Note without Bibliography\)](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [Harvard Reference format 1 \(Author-Date\)](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [IEEE](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [Modern Humanities Research Association \(Note with Bibliography\)](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [Modern Humanities Research Association \(Note without Bibliography\)](#) [\[Install\]](#) (2008-12-23 06:40:00)
- [Modern Language Association](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [National Library of Medicine](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [Nature Journal](#) [\[Install\]](#) (2010-03-14 23:40:00)
- [Vancouver](#) [\[Install\]](#) (2010-03-14 23:40:00)

Click on the “[Install]” link next to the style you wish to install.

# Zotero web account (other options)

- On [www.zotero.org](http://www.zotero.org) you can create a Zotero account.
- You need a Zotero account to synchronize your library, participate in groups, or post to the support forums.
- Data synchronization transfers library items, notes, links, tags, etc.—everything except attachment files—between your local computer and the Zotero servers, allowing you to work with your data from any computer with the Zotero extension. It also allows you to view your library online on [www.zotero.org](http://www.zotero.org).
- To synchronize your data with the Zotero web server, open Zotero's Sync preferences tab and enter your login information in the Zotero Sync Server section.



# Zotero Online Library



Welcome inagel | [Inbox](#) | [Profile](#) | [Settings](#) | [Logout](#)



[Donate](#)

[My Library](#)

[Groups](#)

[People](#)

[Support](#) ▾

[Get Involved](#)

Search support

[Search](#)

Home > People > inagel > Library

## Library

### Collections

#### [My Library](#)

ACT Malaria

CEBHA

ILS

Laboratory strengthening

| Title  | Creator              | Date Added       |
|--|----------------------|------------------|
| Artesunate versus quinine for treating severe malaria  | Sinclair et al.      | 2011-03-22 11:11 |
| Lecture notes.   | Gill                 | 2011-03-21 21:57 |
| WHO   Malaria treatment  |                      | 2011-03-21 21:46 |
| Efficacy and safety of a new pediatric artesunate-mefloquine drug formulation for the treatment of uncomplicated falciparum malaria in Gabon                             | Bouyou-Akotet et al. | 2011-03-21 21:42 |
| Artemisinin Antimalarials: Preserving the "Magic Bullet"   | Maude et al.         | 2011-03-21 21:42 |
| The State of the Art in Anti-Malarial Drug Discovery and Development   | Burrows et al.       | 2011-03-21 21:42 |
| Is parasite clearance clinically important after malaria treatment in a high transmission area? A 3-month follow-up of home-based management with herbal medicine or ACT | Willcox et al.       | 2011-03-21 21:42 |

Send me occasional news by e-mail.[UPDATE NOTIFICATIONS](#)[Profile](#)[Editor](#)[New Projects](#)[Subscription](#)[Account](#)[Notifications](#)[Linked Accounts](#)

## Linked Accounts

You can link your Overleaf account with other services to enable the features described below.

### Twitter

Sign in to Overleaf with your Twitter account.

[LINK](#)

### Google

Sign in to Overleaf with your Google account.

[LINK](#)

### ORCID <http://orcid.org/0000-0001-5007-2529>

Securely establish your identity by linking your ORCID iD to your Overleaf account.

[unlink](#)

Submissions to participating publishers will automatically include your ORCID iD for improved workflow and visibility.

### Figshare

Export your Overleaf projects directly to Figshare via the publish menu.

[LINK](#)

### Zotero

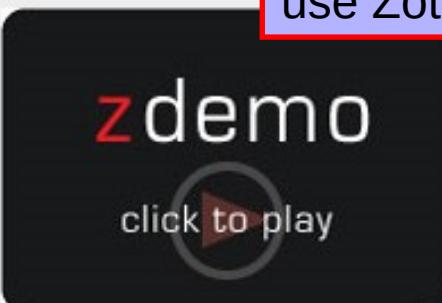
Import your Zotero bibliography into your Overleaf projects with one click.

[LINK](#)

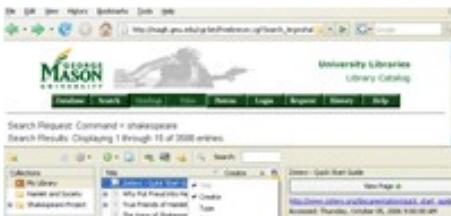
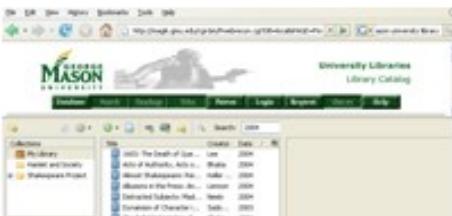
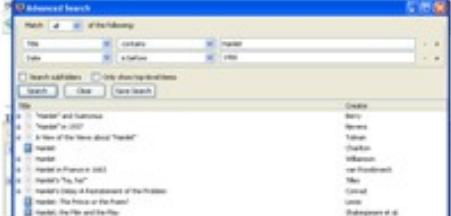
You are here: [start](#) » [screencast\\_tutorials](#)

In an effort to make Zotero as user friendly as possible we have developed these screencasts demonstrating many of the basic functions of Zotero. Click on any of the images to watch screencasts detailing the features named below. Special thanks to Steve Bailey from CU-Boulder for preparing the introductory demo.

## Getting Started

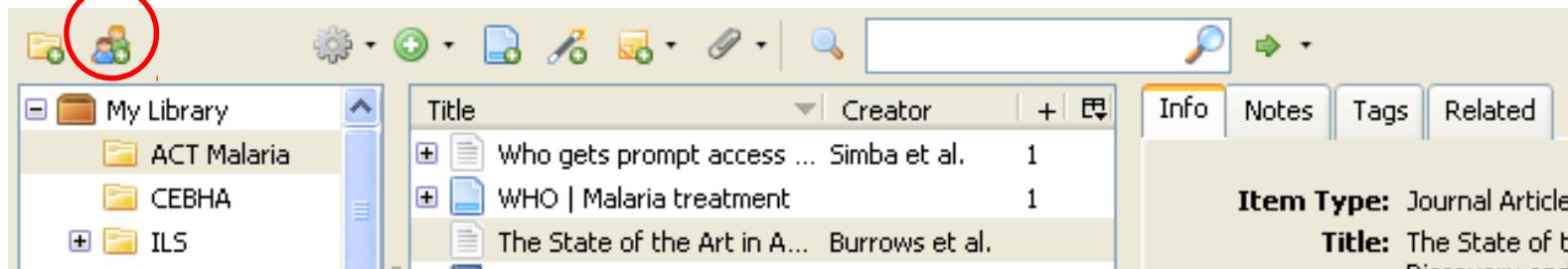
|  |   |  |
|--|---|--|
| <a href="#"><b>Tour of Zotero</b></a><br> | <a href="#"><b>Quick Bibliography</b></a><br> | <p>Zotero <b>Screencasts</b> demonstrate many of the basic functions of Zotero, such as adding references, managing your library, and how to use Zotero in Word.</p>  |
|--|---|--|

## Finding Items

|   |  |   |
|---|--|---|
|  |  |  |
|---|--|---|

# Zotero Groups

- Zotero Groups provide a way to share collections with a class, work with colleagues on a project or keep track of conversations in a specific field of research.
- Groups can be private or public.
- To create a group in Zotero:  
Click the New Group icon located in the top left corner of the left column in Zotero and follow the instructions.



# Zotero Support

On the Support page on the Zotero website (<http://www.zotero.org/support/>), which is available in several languages, you can find documentation on additional features of Zotero as well as Frequently Asked Questions and a number of screencast tutorials.

The screenshot shows the Zotero support page. At the top, there's a navigation bar with links for Home, Groups, People, Support (with a dropdown arrow), Get Involved, Register, Find People, Login, add more file storage, and a Donate button. Below the navigation is a search bar labeled "Search documentation" and a "Search" button. The main content area has a heading "start" and a breadcrumb trail "You are here: start". To the right, it says "Translations of this page:" followed by a row of language codes: en, cs, de, es, fr, hu, it, ja, ko, nl, pl, pt, ru, sl, sv, zh. Under "start", there's a section titled "Quick Links" with a list of links: Installation, Quick Start Guide, Getting Help and Troubleshooting, Frequently Asked Questions (which is highlighted with a red box), Screencast Tutorials, and Developers.



# Data Syncing of Zotero Libraries

- While Zotero stores all data locally on your computer by default, Zotero's sync functionality allows you to access your Zotero library on any computer with internet access. Zotero syncing has two parts: data syncing and file syncing.
- Data syncing syncs library items, but doesn't sync attached files (PDFs, audio and video files, images, etc.). To sync these files, you can [set up file syncing](#) to accompany data syncing, using either Zotero File Storage or WebDAV.

# References

- A Beginner's Guide to Using Zotero - UBC Biomedical Branch Library. Available at:  
<http://www.slideshare.net/giustinid/beginners-guide-to-zotero>
- Zotero Quick Start Guide. Available at:  
[http://www.zotero.org/support/quick\\_start\\_guide](http://www.zotero.org/support/quick_start_guide)

This module initially was developed by:

Ingeborg Nagel, MSc

Royal Tropical Institute (KIT) The Netherlands

# You Tube Resources

- How to insert Zotero citations and references into a Word document; explains Add-Ins (Word) uses APA citation style:
- <https://www.youtube.com/watch?v=imSxa5MbXrc>
- Zotero Overview - covers Chrome, Safari web browsers and standalone option – including how to sync web and stand alone (computer) versions; uses Vancouver citation style:
- [https://www.youtube.com/watch?v=vNfrv9lD\\_TM](https://www.youtube.com/watch?v=vNfrv9lD_TM)
- Zotero Standalone account – to keep info on your computer:
- <https://www.youtube.com/watch?v=H8UTehdF92s>
- Zotero Tutorial – covers options of using Chrome and Safari web browsers:
- <https://www.youtube.com/watch?v=7B1rpXYk7M8>

Updated 2015 12

# **REVIEWING AND GIVING FEEDBACK**

Peer Feedback process  
Conference/Journal Reviewing

Based in part on material from IRA/NCTE.  
Images copyright Microsoft Corporation.

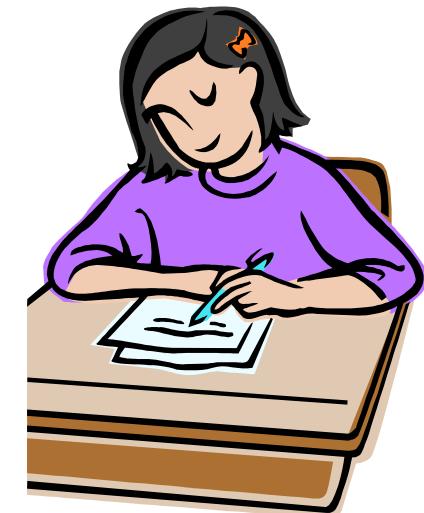


**Vision And Security Technology**  
University of Colorado **Colorado Springs**

# 6 STEPS TO EFFECTIVE PEER FEEDBACK

There are three important steps to remember when you are giving feedback on another student's work.

- Step 1: **Skim**, make a few notes
- Step 2: Read it **Critically & creatively**
- Step 3: **Summarize (3-4 sentence)**
- Step 4 – **Compliments**
- Step 5 – **Suggestions**
- Step 6 – **Corrections**



# STEP 1: SKIM

- Do your 3-5 min skim. Make notes on what you think is the core of the paper and high-level issues and if that made you want to continue.
- You scan should try to find at least 2 positive things to say.
- Note questions you had at this stage.

## STEP 2A: CRITICAL READ

- Max 30min for this assignment (could be 1-2 hours for a real review of “novel” work).
- Read it all, taking notes as in critical read.
- If you use paper or pdf, can mark/circle/highlight areas with issues. Don’t have to describe the issues in details.
- Note questions that need to be answered.
- Your job is not proofreading, but you can markup what you note, but balance time. No point telling them about fixing grammar if the section needs to be rewritten.

## STEP 2B: CREATIVE READ

- Find the core of the story that you think works.
- Find multiple positive things to say
- Include usual creative reading process/ideas.
- Try to think of ways to make the paper “better” in your view but remember you are only 1 reader.

## STEP 3: SUMMARIZE

- Generate a 3-4 summary of what you think the paper is about.
- Do NOT, just cut/paste sentences from the abstract as the summary. Use your notes but don't look at the paper.
- If scan and detailed reads are different have a sentence or 2 about that differences and what might have lead to the confusion.

## COMPLIMENTS

- The first rule of peer feedback is to BE POSITIVE and Constructive!
  - Remember, THIS IS about helping them!
  - Think about how you would want someone to speak to you about your writing.



## STEP 1

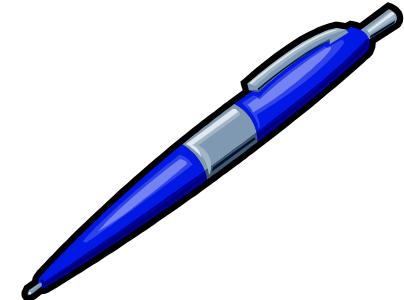
# COMPLIMENTS

Tell the author(a) what you think they did **well**:

- The significant part of the was ....
- I really loved your story because
- I think you used a lot of supporting details for ....

## STEP 2

# SUGGESTIONS



- **Making suggestions means giving the author some specific ideas about how to make his or her writing better.**
- Remember – stay positive and be specific!
  - Instead of, “It didn’t make sense,” say, “If you add more details after this sentence, it might be clearer.”

If just writing comments, clearly specify where is the issue (page, paragraph column line). If marking can highlight sections and let the authors figure out what is the issue.

## STEP 2

# SUGGESTIONS



## Here are some areas that you may want to make suggestions about:

- Perhaps using \_\_\_\_\_ would have helped the reader \_\_\_\_\_.
- I didn't understand the sentence/paragraph \_\_\_\_\_.
- I found your argument about xxx confusing (weak, not convincing..)
- Your conclusion was too short/long/obvious.
- \_\_\_\_\_ statement is an extraneous detail.
- \_\_\_\_\_ statement is unsupported.
- \_\_\_\_\_ statement is wrong.
- A citation might help \_\_\_\_\_

### STEP 3

## CORRECTIONS

- The third step in the peer feedback process is making corrections.
- Corrections means making suggested edits to your peer's paper for:
  - Spelling mistakes
  - Grammar mistakes
  - Missing punctuation
  - Incomplete or run-on sentences



# **HOW TO REVIEW PAPERS FOR CONFERENCES/JOURNALS**

Based on part on slides from [Milos Prvulovic](#)



**Vision And Security Technology**  
University of Colorado **Colorado Springs**

# STANDARD REVIEW ITEMS

- Title and author of paper
- Summary of paper & contributions
- Positive things about the paper (one paragraph). Always find some.
- Weaknesses
  - Major comments
  - Minor comments
- Private Comments to the PC/AC/Editor

# PAPER REVIEWING ALGORITHM

- Read the paper
- Think about it
- Take a look at related work
- Leave it alone for a few days
- Read it again
- Write the review

# THINKING ABOUT IT

- What is the paper trying to do
  - Observation/Motivation
    - Programs often do X...
    - In the future, Y won't work...
  - Implementation
    - Here's how to exploit this observation
  - Evaluation
    - We get great perf/power/reliability/... improvement, or
    - It didn't work and here's why

# CONTRIBUTIONS

- Did you gain an important insight
  - A new problem?
  - A new way to look at a problem?
  - Counter-intuitive or unexpected?
- Did it propose a good solution to a problem?
  - Explanation why solution should work
  - Explanation why it's a better solution
- Did it provide a good quantitative evaluation
  - Should work in real processors?
  - Should it gain or lose relevance in the future?
  - Should work on real applications?

# SUMMARY

- Like a paper abstract
  - Short and self-contained (no undefined acronyms)
  - Short high-level description of contribution
  - DO NOT just copy from their abstract
- But without the “sales pitch”
  - Avoid quoting numbers
  - What is really the contribution

# POSITIVE CONTRIBUTIONS/GOOD POINTS

- Successful contributions
  - If you learned something, that's a good thing
  - If could be useful for real systems, that's good
- List the most important things
  - If you only list “Well written”, paper is really bad
- There are ALWAYS good points ^ find them

# WEAKNESSES (BAD POINTS) MAJOR

- Motivation problems
  - Unclear whether there is a problem
  - Unclear why it's an important problem
- Implementation problems
  - Unclear why solve the problem this way
  - Unclear what tradeoffs are
- Evaluation problems
  - We all know about these
- Organization problems
  - Difficult to follow?
- List the most important bad points
  - If “difficult to read” is the only problem, it’s a great paper

# QUESTIONS TO ANSWER

- Important questions only
  - If answer can't change your recommendation  
don't ask the question here (see next slide)
- Can it be answered in a few days?
  - If lots of work needed to answer, don't ask here
- Unsure if you got something right?
  - If decision depends on something, and  
you are not sure if your understanding is correct  
DO ask it here

# “MINOR/GENERAL COMMENTS” SECTION

- Help the authors improve the paper
- Always be constructive
  - Even if paper is hopeless:  
“The problem of cache conflicts in 1024-way set associative caches is not very important, and the authors should focus on lower-associativity caches”
- All grammar/writing/organization issues go here!
- If it’s a reject, can stop grammar after a few examples..

# CONFIDENTIAL COMMENTS

- Only put things that are important and would disclose your identity or would be questioning ethics
  - “My group did this already in 1999 and we published it in ISCA that year”
  - I think is is a double submission similar to paper XXX at conference yyy
  - I think this paper is plagerizing paper ZZZZ

# SCORING (NUMBERS) IF THEY HAVE THEM

- Compare to other papers in the same journal/conference/class
- So, compare to other papers in this class
- Weak accept
  - Similar in quality to lower half of accepted papers (top 20% of all papers)
- Strong accept
  - Similar to upper 50% of accepted papers (top 10% of submission)
- Neutral/boarderline
  - Would be in bottom 10% of accepted papers
- Weak reject
  - Top half of rejected, but not fundamentally broken
- Strong reject
  - You would argue against accepting it

# ALWAYS THE BIG PICTURE

- Don't focus on minor issues
  - Very easy to find lots of these
- Focus on make-or-break issues
  - If X wasn't there, it would be a much worse paper
  - If X done right, it would be a much better paper
- Always consider the space
  - Don't ask for it if you don't think something else should be removed or made shorter

# LOOKING FOR EXAMPLE REVIEWS

- Ask your advisor (probably want to read the papers/reviews in your area)
- Look at
  - <https://openreview.net/> (reviews of both accepted and rejected papers)
  - <http://papers.nips.cc/> (recent years have reviews of accepted papers)