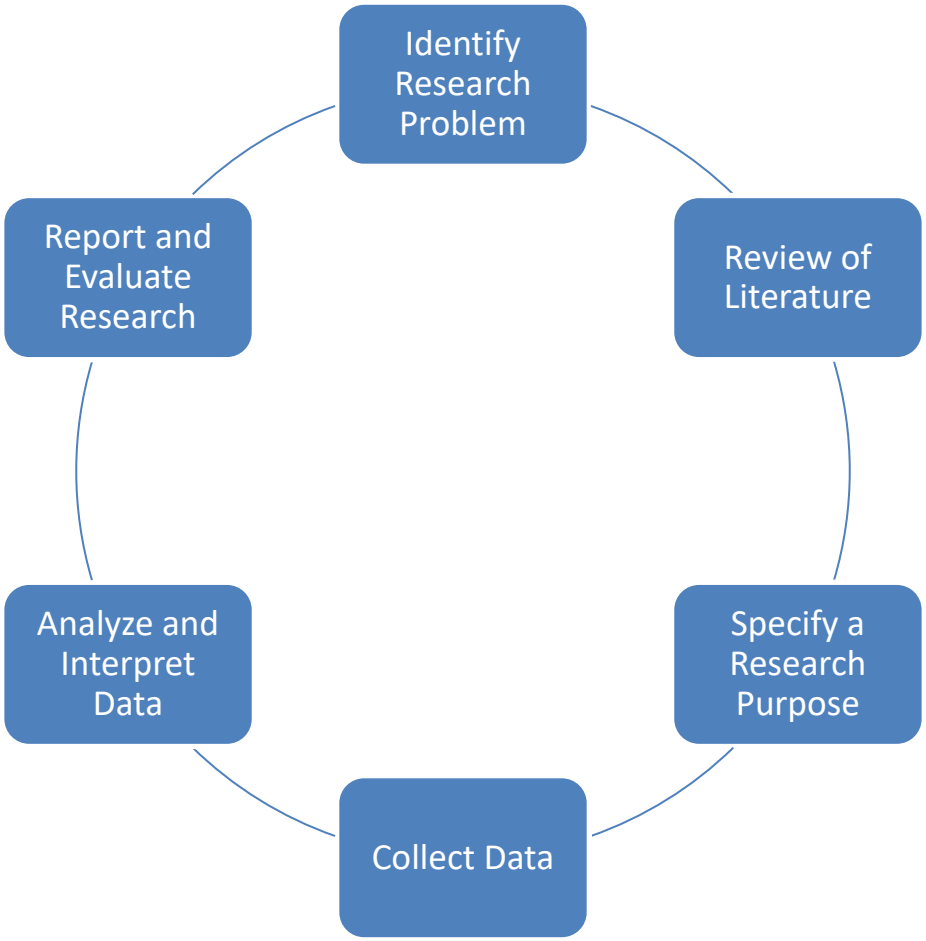
Research and Ph.D. Journey

CS 7123, Spring 2025

Maryam Tabar, Ph.D.

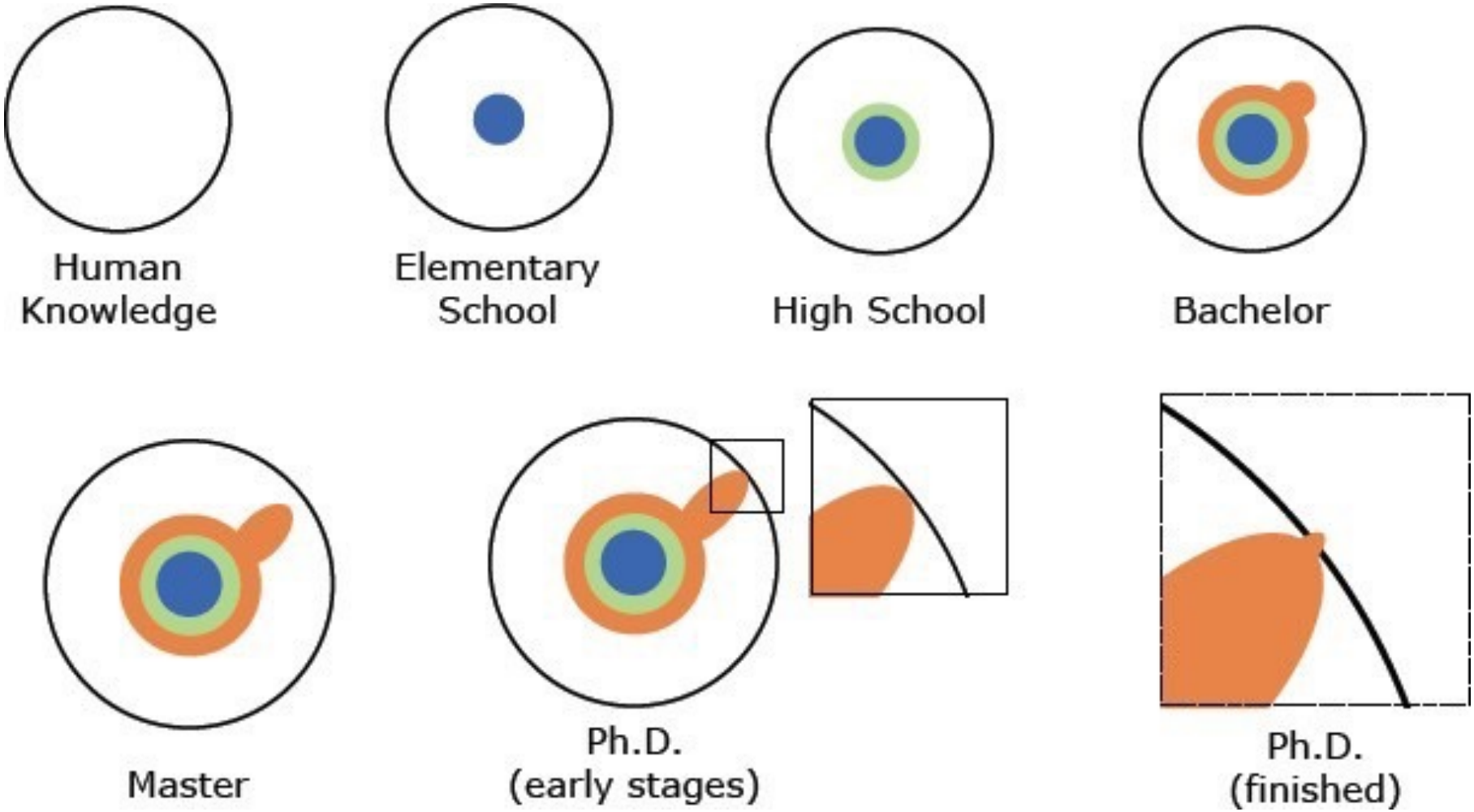
Used/adopted slides of the “Research Design: Interdisciplinary Research Methods for Information Sciences and Technology” course, taught by Prof. Dongwon Lee, at Pennsylvania State University

The Research Process



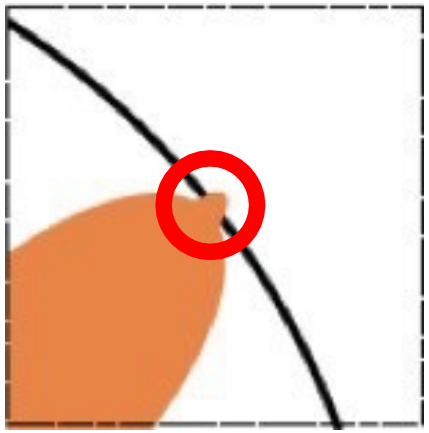
Source: Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, by John W. Creswell and Timothy C. Guetterman

What is “Ph.D.”?



Source: http://gizmodo.com/5613794/what-is-exactly-a-doctorate

Publish or Perish



- ○ has to be written first
- ○ has to be validated as novel
- ○ has to be peer-reviewed & published
- To get a good job, must have **good** and/or **many** ○

The Goal of Research Papers

- Disseminate your ideas to others so that people appreciate/use/cite them
- **Graduate**… Of course
  - MS: need to write a thesis to graduate…
  - Ph.D.: “Publish or Perish”
- Without good publications…
  - No good job, no good career
- GPA: rarely care about PhD’s GPA
  - Creating new knowledge >> learning existing knowledge
  - Maintaining reasonable GPA is sufficient

Difference by Discipline

- Pure Sciences (eg, Math, Physics)
  - Pre-print at arXiv.org
  - Rigorous reviews for journals
  - journals >> conferences
- Social Sciences
  - Rigorous reviews for journals
  - Conferences are mostly for gathering or short abstract based screening
  - journals >> conferences
- Computer Science
  - Peer-reviewed conferences
  - Top conferences have 10-20% acceptance rates
  - Specialized and small conferences (attendance of 400+)
  - conferences >> journals

Attending Conferences

- Full paper > Short/Poster/Demo paper

- Networking
- Goal of presentation is to have others get interested in your ideas (so that they read your paper later)

- Find interesting topics and papers
  - Make your presentation as SIMPLE as possible
  - make it INTERESTING

Facts on Paper Reviews

(adopted from J. Cho's slides @ UCLA)

- 3-5 reviewers per paper
- 5-15% acceptance rates for top-tier venues
  - Very competitive
- Criteria
  - Strong Accept/Weak Accept
  - Neutral
  - Weak Reject/Strong Reject
- One weak/strong reject can kill a paper
  - Minimum Weak Accept

About Reviewers

- 10-15 papers per reviewer (for top conferences)
- Reviewer cannot spend 5-10 hours per paper
  - 20 X 10 = 200 hours = (40 hours X 5) = 5 weeks!
- Give a good impression in 1-2 hours!
  - Impression matters a lot
- Reviewer do NOT get paid

WARNING: Of course, to start with, your main idea must be good to get into top-tier venues…

Good Impression in 1-2 hours?

1. Good introduction
   - Everyone reads it
   - If not interesting, people stop reading
2. Easy to read
   - People should understand what you say
3. Build an excitement and a strong case
   - What is good?
4. Broad reference
   - Sometimes kills a paper

Good Introduction Sells

I have often said reviewers make an initial impression on the first page and don't change 80% of the time

Mike Pazzani

This idea, that first impressions tend to be hard to change, has a formal name in psychology, *Anchoring*.

Excerpt from "How to do good research, get it published" by Eamonn Keogh

Good Introduction Sells

The Most Important Part of Your Paper: the Introduction

- The 1/3 – 2/3 Rule from a reviewer's perspective:
  - 1/3 time to read your introduction and make a decision
  - Remaining 2/3 time to find evidence for the decision
- [Take-Home Message #6] **A good introduction with a good motivation is half of your success!**

Xindong Wu

Excerpt from "How to do good research, get it published" by Eamonn Keogh

Good Introduction Sells

The First Page as an Anchor

The introduction acts as an anchor. By the end of the introduction the reviewer *must* know.

- What is the problem?
- Why is it interesting and important?
- Why is it hard? why do naive approaches fail?
- Why hasn't it been solved before? (Or, what's wrong with previous proposed solutions?)
- What are the key components of my approach and results? Also include any specific limitations.
- A final paragraph or subsection: "Summary of Contributions". It should list the major contributions in bullet form, mentioning in which sections they can be found. This material doubles as an outline of the rest of the paper, saving space and eliminating redundancy.

This advice is taken almost verbatim from Jennifer.

Jennifer Windom

If possible, an interesting figure on the first page helps

Excerpt from "How to do good research, get it published" by Eamonn Keogh

DARPA Heilmeier Catechism

- What are you trying to do? Articulate your objectives using absolutely no jargon
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you are successful, what difference will it make?
- What are the risks?
- How much will it cost?
- How long will it take?
- What are the mid-term and final “exams” to check for success?

https://www.darpa.mil/about/heilmeier-catechism

Easy-to-Read Paper ➔ Good Paper

You can always make it complex later

1. Use examples
2. Figures & Tables – Figure speaks !!
   - Summary of notations
3. Define assumptions/models/architecture precisely
   - Explicitly write down assumptions
   - Input, output, property, goal function
4. Make a connection
   - Why this experiment?

Paper Organization (9 pages)

1. Introduction (1 and half pages)
2. Related Work (1 page)
3. Main Ideas (2 pages)
4. Experiments (3 pages)
5. Conclusion (half page)
6. References (1 page)

Short Main Idea

- Watson & Crick's Nature paper on double helical structure of DNA is only 1 page (+ 1 paragraph) long

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1929).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxy-ribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxy-ribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

What *NOT* to do

- Research Misconduct
  - Plagiarism
  - Research manipulation
  - Double submission/publication

https://www.publichealthnotes.com/research-ethics-definition-principles-and-advantages

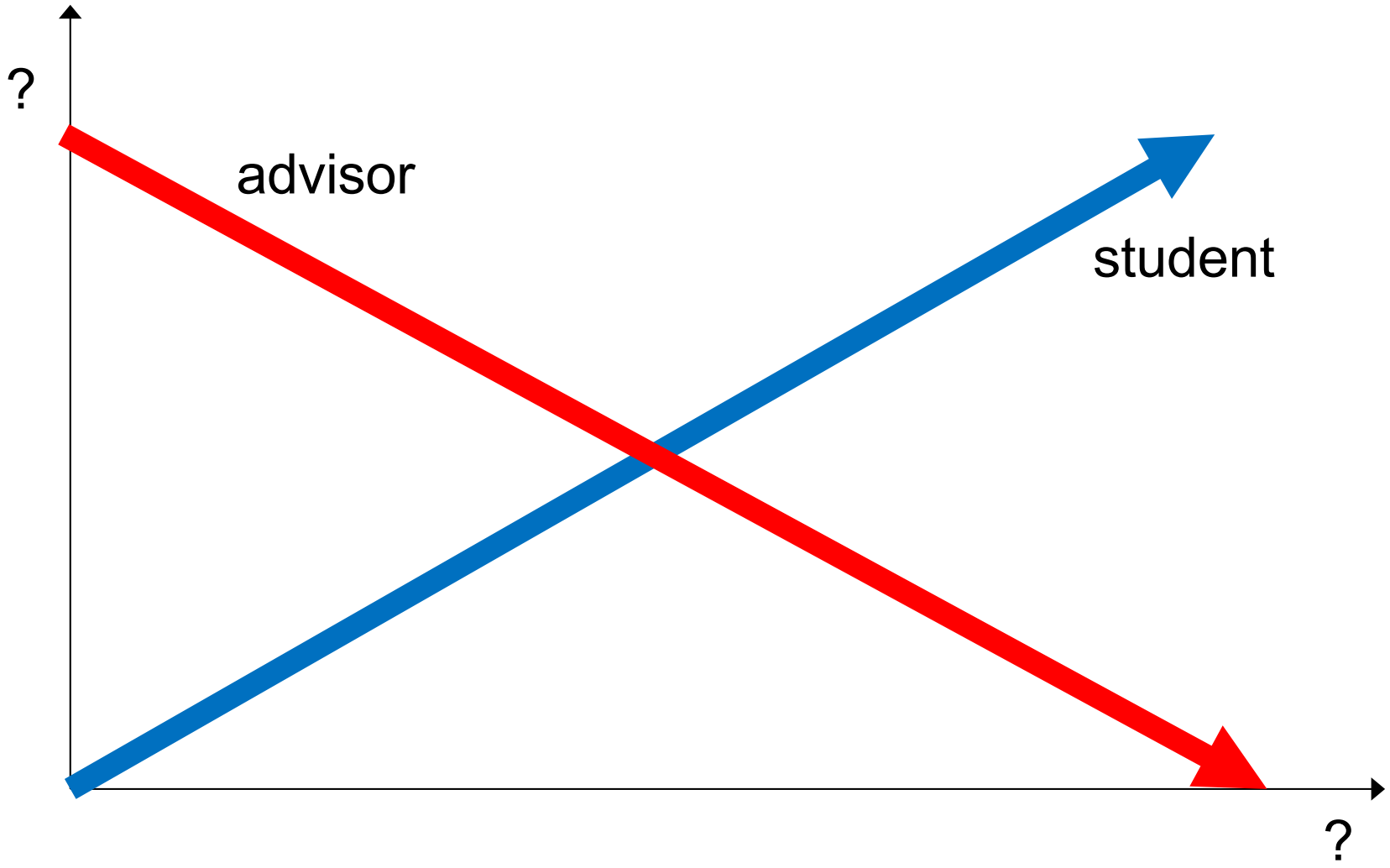
Start Writing Early On…

- Even if you feel you are NOT ready yet

- HOWEVER, you get:
  - Invaluable experiences and learn from that

    Having such experiences early is important
  - Writing sharpens your ideas and gives more ideas

Ideal Grad Student



Habit #0. Reading Papers

- #1 most important skill for researchers
- Survey paper
- Reviews for conferences or journals

Habit #1. Personal Research Log

- Maintain personal research log or diary
  - Sketch your research ideas into writing
  - Update your ideas as time passes
  - Occasionally *go back to old writings*
- Prepare the informative review for each paper that you read
  - Summary
  - Pros and cons
  - Limitations or problems
  - If needed, contact authors and ask questions

    Usually authors are very happy to discuss with their readers

Habit #2. Practice Writing

- Writing skill is critical in research
- Research paper is a long logically-written fact-based "essay" to convince readers
  - Logical argument and critical finding are must

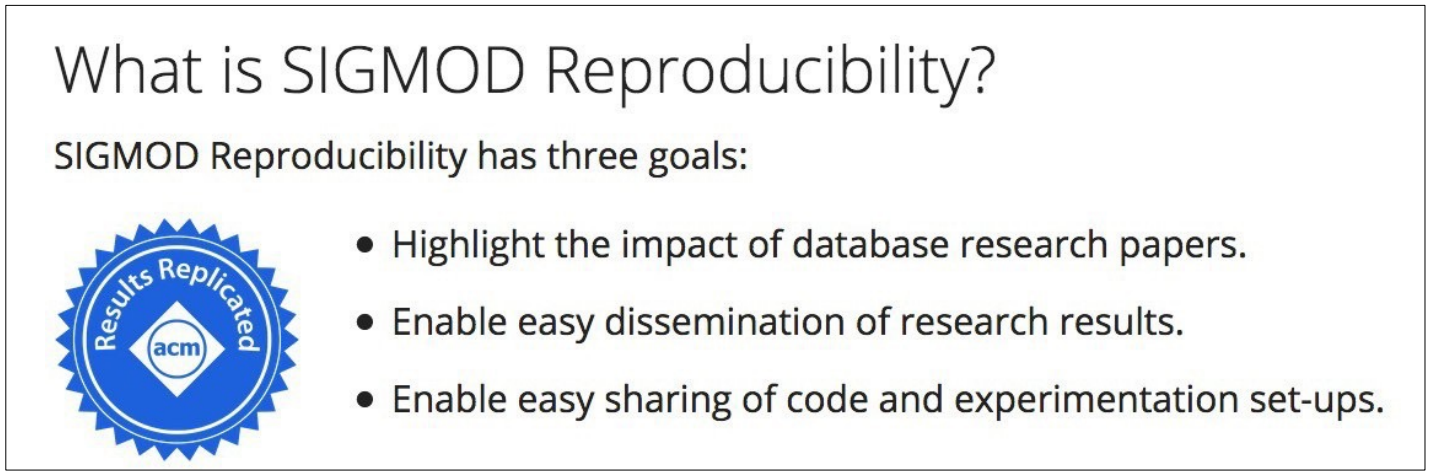
- Equally, or probably more, important ➔ storytelling
  - How to present your findings in a more convincing and approachable way to both experts and lay readers

Habit #3. Practice Presentation

- Critical skill for any jobs
  - Video record your presentation for self-check
- Practice your presentation whenever possible
  - Lab/class presentation
  - Ask feedback from peers
- Attend others' presentations whenever possible
  - Faculty job talk
  - Research group talk
  - Give questions and approach speakers after talks

Habit #4. Dissemination

- Have personal homepages
  - http://john-doe.github.com/
  - Share your works
- Make datasets (and codes) publicly available
  - Replicable research becomes more important
  - By default, after a publication, share datasets and codes



Habit #5. Learning, not Grades

- Have to have different mindset as PhD students
  - Be proactive, not passive, learners
  - Grades from classes or overall GPA becomes less important
- You take courses because they will be useful for your research, not to get good grades
- You work hard in classes so that you gain the most in learning, not to get good grades