The Research Process & Research Ethics

Molly O'Neil
REU Summer 2015

(Excerpted/adapted from Paul Wagner's "Research Methodology", David Touretzky's "Ethics and Etiquette in Scientific Research", and several other sources [cf. last slide])

TEXAS★STATE ®
UNIVERSITY

The rising STAR of Texas

Who Am I?

- Recent TxState Computer Science graduate (M.S., May 2015)

  - Research interests: computer architecture, hardware & software design for irregular parallelism, general-purpose GPU programmability

  - 2 conference papers, 2 workshop papers, and a Master's thesis

  - NSF Graduate Research Fellowship recipient (more on this in July)

- Longer ago…

  - B.S., Electrical & Computer Engineering, Carnegie Mellon, May 2005

  - Design Engineer at ARM until 2010 (high-performance mobile CPUs)

What Is Research?

Adapted from
Paul Wagner's "Research Methodology"
http://www.cs.uwec.edu/~wagnerpj/talks/researchmethodology.ppt
Student ACM Meeting Fall 2003

What Is Research?

- Merriam-Webster Online Dictionary:

  - "a 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"

- So research is…

  - Systematic, rigorous, and thorough

  - Aimed at finding *new* (novel) theories, solutions, or applications

What Is Research?

- Still lots of scope within this definition

  - Small steps vs. large insights

  - Entirely new discovery vs. new application

  - New method vs. better/faster method

  - New result vs. better understanding of old observation

- Always standing on the shoulders of prior work, but…

  - "What we have now isn't quite good enough — we can do better!"

What Is Research **NOT**?

- Programming project

- Product/software development

- Playing with computers, robots, and other cool technology

- Writing a book report

- Re-implementing what's already been done

What Is Research **<u>NOT</u>**?

- Programming project

- Product/software development

- Playing with computers, robots, and other cool technology

- Writing a book report

- Re-implementing what's already been done

However, research often involves doing some or all of these things!

Who Does CS Research?

- Industry R&D teams

- Government labs

- University researchers

  - Faculty

  - Post-docs

- Graduate students

  - Ph.D. students (world-level contribution)

  - Masters students

- Undergraduate students (i.e., you!)

Why Research?

- The large-scale answer

  - Find something that changes the world

  - The thrill and satisfaction of discovery

  - Develop "high risk, high payoff" ideas

- A bit more personal…

  - Work closely with faculty mentors

  - Determine your technical interests & career goals

  - Strengthen your resume (even for non-academia)

  - Get really good at time-management and self-motivation

The Research Process

Adapted from
Paul Wagner's "Research Methodology"
http://www.cs.uwec.edu/~wagnerpj/talks/researchmethodology.ppt
Student ACM Meeting Fall 2003

An Overview

1. Initial Idea

2. Background Investigation

3. Refinement of Idea

4. Core Work

   - Investigation, Development, and Analysis

   - Documentation

5. Evaluation

6. Identification of Future Work

7. Presentation of Results

1.) Initial Idea

- Be on the lookout for problems
  - Inelegant or inefficient solutions
  - Gaps in understanding
  - Slightly-different behavior that can be generalized or abstracted

- Requires critical thinking and constant questioning
  - "Can this be done better/faster/more accurately?"
  - "Do we understand why we see this behavior?"
  - "Can I apply this technique somewhere else?"

- Read lots of papers!
  - (…and pretend you're a reviewer)

Some Notes on Choosing a Topic

- Your topic must…

  - Inspire *enthusiasm* (in you)

  - Inspire *interest* (in others)

- Use your mentor as a sounding board

  - They might already have ideas for you to work on

- You're going to spend a *lot* of time thinking about your topic

  - Won't always be fun or easy

  - If you don't like thinking about it now, imagine how you'll feel by end of summer/college/Ph.D….

2.) Background Investigation

- Related Literature Search

  - Has this problem already been solved?

    - If so, how is your idea better?

  - What similar work will yours be building on?

  - What will be your work's impact and contribution?

  - You will use all of this later when you present or publish your work

- Tools

  - Library resources (e.g., ACM & IEEE journals, conference proceedings)

  - WWW search (e.g., http://scholar.google.com/)

How to Do a Literature Search

- Note the complete citation (IEEE, ACM formats)

- One short paragraph summary of main focus
  - What is the paper about?
  - What do the authors see as the paper's contributions?
  - Sometimes your'e lucky and can pull this straight from the abstract — but usually not

- What did you like/dislike about the paper?
  - How does it impact your project? Can you build on it?
  - In what ways is it transformative?
  - What is wrong with or missing from the authors' approach, analysis, or findings?
  - What would you have done differently?

- Pointers to other relevant papers

3.) Refinement of Idea

- Incorporate knowledge from background investigation

- Issues:

  - Precision: Identify…

    - The specific problem you're trying to solve

    - Possible *solutions* (plural!)

  - Scope ("fence in" vs. "fence out")

    - What's the essential part of this work?

    - What's tangential, additional, or otherwise best left for later (or for someone else)?

4.) Core Work

- Infrastructure
  - What equipment/software do you need?
    - Try to leverage existing systems: you want to demonstrate the value of your ideas, not re-implement the surrounding toolsets
  - Learn any new skills required (e.g., tools, languages)

- Experimentation & Analysis
  - Run experiments
  - Develop theories based on your results
  - Discuss and brainstorm with others
  - *Re-evaluate* as you go
    - Better ways of solving the problem? Can you generalize/abstract?
    - Never be so committed to your first idea that you can't see what your data are telling you!

4.) Core Work

- Document as you go
  - Don't want to lose knowledge of what you did

- Keep a journal for day-to-day thoughts
  - Paper notebook, Dropbox, emails to yourself, whatever works for you
  - Should be something you can always access
    - You never know when or where you'll have a great idea

- Up-to-date task list
  - It's easy to forget all the threads of ongoing work

- Try to formally write up your work as you go
  - Summarize work, accomplishments, problems, conclusions
  - It's no fun reconstructing whole sections of a project from your notes with a deadline looming

5.) Evaluation

- You have all your data, you've implemented your prototype…

- Still lots of hard questions to answer!

  - Does your prototype work?

  - What did you improve over existing alternatives?

    - How much of an improvement?

  - What can you do now that couldn't be done before?

  - What deeper understanding have you provided?

6.) Identification of Future Work

- Helps organize your own future efforts

- Helps others build on your work

- Sources:
  - Ideas "fenced out" during refinement stage
  - New problems or questions that surfaced during the work

7.) Presentation of Results

- If no one knows about your work or no one can use it, it's not a contribution

- Ways to disseminate your findings:

  - Journals, Conference & Workshop Proceedings

    - e.g., IEEE International Conference on Data Mining

  - Student Research Conferences

    - e.g., National Conference on Undergraduate Research

  - Poster Sessions

    - e.g., WISE Conference at Texas State

  - Web

Research: The Student Experience*

*(*One recent graduate's highly personal and biased opinions...)*

A Dose of Reality

- Research is hard!

    - Easy to burn out

    - Easy to doubt the relevance, novelty, or correctness of your own work

        - Surveys have shown even Nobel Laureates in science report feeling this way!

        - Critical evaluation is a good skill — this is an inevitable side effect

- Lots of risk: "If your project can't fail, it's development, not research"

    - Fear of failure can make it hard to stay motivated

    - <u>Fact</u>: you're going to accumulate some failures

        - Nothing about coursework prepares you for this!

    - Some of them you'll be able to leverage later

- Research always takes longer than you think it will

Tips on a Good Work Process

- Work regularly

  - Doesn't have to be 9-5, but regular work hours help (choose time of day you work best)

- Allocate enough time to make progress

  - It takes awhile to gear back up to speed

- The mental stuff

  - If an idea doesn't work, dust yourself off and keep trying

  - Accept that progress rate will not be constant and getting stuck is OK

How to Deal With Getting Stuck

- It's all about regaining momentum, so switch to something confidence-building
    - Pick a sub-problem that seems less scary (support code is good for this)
    - Catch up on those "nice to have" tasks you never get around to

- What I do when I'm *really* stuck:
    - Tell myself I'm going to work for one hour, then go do something fun
        - Set a timer and make myself work (see ideas above) until it rings
        - Often by the end of the hour, I'm no longer stuck and keep working!
        - If not, really go take a break

- Sometimes those "Aha!" moments that drive research come easiest when not obsessing

How To Be A Bad Research Student

How To Be A Bad Research Student

- Don't make use of your advisor

  - **Reality:** Access to faculty is one of the most valuable aspects of undergrad research. Faculty careers are judged in part on success of students. Your advisor was a student once, too (and probably remembers).

How To Be A Bad Research Student

- Don't make use of your advisor

  - **Reality:** Access to faculty is one of the most valuable aspects of undergrad research. Faculty careers are judged in part on success of students. Your advisor was a student once, too (and probably remembers).

- Don't read on your own

  - After all, if something is important, your advisor will know about it and make you read it, right?

  - **Reality:** It's a huge and fast-moving field and your mentor can't know everything about everything. Learn something new, then teach your advisor.

How To Be A Bad Research Student, continued

How To Be A Bad Research Student, continued

- Don't ask for help when you're stuck

  - Worse, hide that you're stuck; over-promise results or overrepresent progress

  - **Reality:** This isn't a class project; collaboration is encouraged. Your advisor wants you making progress, not stalled forever on something you just can't see. Earn and keep your advisor's trust.

  - **Flip side:** Make a real attempt to get there by yourself before running for help.

How To Be A Bad Research Student, continued

- Don't ask for help when you're stuck

  - Worse, hide that you're stuck; over-promise results or overrepresent progress

  - **Reality:** This isn't a class project; collaboration is encouraged. Your advisor wants you making progress, not stalled forever on something you just can't see. Earn and keep your advisor's trust.

  - **Flip side:** Make a real attempt to get there by yourself before running for help.

- Consider your advisor just your manager ("working for" vs. "working with")

  - **Reality:** You are collaborating at the leading edge of the field. You weren't selected just to implement someone else's vision. Contribute ideas! Challenge assumptions! Don't be afraid to ask stupid questions (they might not actually be stupid) or argue that your way is better (of course, be open to understanding why you're wrong, too).

How To Be A Good Research Student

How To Be A Good Research Student

- Read lots
  - You're likely going to spend a lot of the next few weeks doing this
  - Set aside part of each week for paper reading

How To Be A Good Research Student

- Read lots
  - You're likely going to spend a lot of the next few weeks doing this
  - Set aside part of each week for paper reading

- Work hard and consistently
  - Lots of us procrastinate about class projects and have learned we can get away with it: ***research does not work this way!***

How To Be A Good Research Student

- Read lots
  - You're likely going to spend a lot of the next few weeks doing this
  - Set aside part of each week for paper reading

- Work hard and consistently
  - Lots of us procrastinate about class projects and have learned we can get away with it: ***research does not work this way!***

- Trust your advisor and go talk to him/her…
  - When you're stuck or confused or spinning your wheels
  - When you need advice or motivation
  - To share ideas, get feedback, and brainstorm

How To Be A Good Research Student

- Read lots
  - You're likely going to spend a lot of the next few weeks doing this
  - Set aside part of each week for paper reading

- Work hard and consistently
  - Lots of us procrastinate about class projects and have learned we can get away with it: ***research does not work this way!***

- Trust your advisor and go talk to him/her…
  - When you're stuck or confused or spinning your wheels
  - When you need advice or motivation
  - To share ideas, get feedback, and brainstorm

- Be open-minded
  - Don't be afraid of criticism or argument
  - Don't be afraid to ask questions and throw out crazy ideas
  - Don't get locked into one way of seeing a problem

Research Ethics

Excerpted & Adapted from
David S. Touretzky's "Ethics and Etiquette in Scientific Research"
Carnegie Mellon University, April 2007

"Research Ethics" Means Lots of Things

- Use of human subjects

  - Informed consent, data retention/confidentiality, IRB oversight

- Use of animals

  - Appropriate care/use, IACUC oversight

- Moral and ethical debates

  - Stem cell research, impact of new technology (genetic screening, nuclear weapons), etc.

- Professional issues (today's topic)

  - Authorship, intellectual property (IP) rights, data management, etc.

How to Avoid Ethical Dilemmas

1. Know the rules

- How should researchers behave?
- Who says so?

2. Know your rights and responsibilities

- Co-authorship
- Ownership of intellectual property (IP)
- Conflicts of interest

3. Know and recognize common ethical mistakes

- Misappropriation of text or ideas
- Deceptive reporting of results

4. Learn from others' mistakes

Know the Rules: Education & Resources

- NSF funding requires Responsible Conduct of Research training

- Booklets:
  - *On Being a* Scientist, National Academy of Sciences
  - *Honor in Science*, Sigma Xi
  - *Good Science and Responsible Scientists*, AAAS

- Web:
  - *http://ori.hhs.gov*
  - *http://onlineethics.org*

- Case studies

Know the Rules: Official Policies

- Professional/scientific societies have codes of ethics (e.g., ACM)

- U.S. Office of Science and Technology Policy defines misconduct as "fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting results"

  - Must be committed intentionally, knowingly, or recklessly

  - This definition has been adopted by most research funding agencies

- Many scientific journals impose ethical requirements for authors

- University academic honor codes

  - Usually similar to the above, and may add abuse of confidentiality, failure to allocate credit appropriately in publications, failure to report misconduct

Issue #1: Allocation of Credit

- Two forms of credit in a paper:

  - Co-authorship

  - Acknowledgments

- Who gets listed as a co-author?

  - Lab director is a co-author on all papers?

  - What's the least contribution that results in a spot on the author line?

  - What order do author names appear in?

- Authorship conventions vary greatly by discipline

Co-Authorship

- Rule of thumb:

  - A co-author should have made direct and substantial contributions to the work (not necessarily to the writing):
    - Design of the research
    - Interpretation of the data
    - Drafting of the paper

- As well as establishing *credit*, the author list establishes *accountability* for the scientific integrity of the paper

  - Penalties may apply!

  - David Baltimore case: Nobel Laureate at MIT was co-author on a paper where primary investigator was later accused of fraud

Ordering of Authors

- How is the ordering of authors determined?

  - First and last are usually the key positions

    - In some fields, group leader's name goes last; in some fields, it goes first

  - Some disciplines order authors alphabetically, some by contribution amount

- Authorship is best discussed openly and frankly *and as far in advance as possible*

Co-Authorship in Computer Science

- Generally: authors ordered by the amount of their contribution
  - But in some sub-communities, author list is alphabetical
  - Senior author (i.e., lab director) is often (but not always!) last
  - No general consensus on lab directors always getting co-authorship

- Papers typically have 1-4 authors

  - Rarely see the large author lists of, e.g., physics

- Contributions may include…
  - Providing key ideas, theory, or design
  - Doing the implementation
  - Running experiments and/or collecting data
  - Analyzing/interpreting the data
  - Writing up the results

Ask Your Advisor

- What are the authorship conventions in our field?

- What are the authorship conventions in your lab?

- Are students prohibited from submitting papers (even sole-authored) without your approval?

- Who owns the code/data/manuscript? (Also see university policy on intellectual property)

Acknowledgments

- Usually a small section at the end of the paper

- People who made contributions that don't merit co-authorship may (sometimes *must*) be acknowledged elsewhere in the paper

- Not as good as co-authorship, since it doesn't appear on a CV

- But it's good manners, and costs nothing (but a line or two of paper space)

Acknowledge People Who...

- Contribute a good idea

- Coin a useful term

- Provide pointers to citations

- Help with debugging tricky code

- Help with typesetting or illustrations

- Provide significant resources, e.g., equipment

- Provide funding!

Issue #2: Misappropriating Text

- Borrowing "just a sentence or two" without attribution is *plagiarism*

- Very easy to avoid: just give a citation!

Example: Misappropriation vs. Attribution

Example: Misappropriation vs. Attribution

- Smith: *The parrot is a remarkable bird in many respects. In terms of intelligence, humor, and manual dexterity, it is unequalled in the avian kingdom.*

Example: Misappropriation vs. Attribution

- Smith: *The parrot is a remarkable bird in many respects. In terms of intelligence, humor, and manual dexterity, it is unequalled in the avian kingdom.*

- Jones, wrong way: Parrots are excellent mimics. But <u>*the parrot is a remarkable bird in many*</u> other <u>*respects. In terms of intelligence, humor, and dexterity, it is*</u> unparalleled <u>*in the avian kingdom*</u>.

Example: Misappropriation vs. Attribution

- Smith: *The parrot is a remarkable bird in many respects. In terms of intelligence, humor, and manual dexterity, it is unequalled in the avian kingdom.*

- Jones, wrong way: Parrots are excellent mimics. But *the parrot is a remarkable bird in many* other *respects. In terms of intelligence, humor, and dexterity, it is* unparalleled *in the avian kingdom*.

- Jones, right way: Parrots are excellent mimics. But in addition, as Smith (2005) observes, "*in terms of intelligence, humor, and manual dexterity, [they are] unequalled in the avian kingdom*."

Citation Etiquette

- Cite other people's work freely and often! Because…

  - You avoid antagonizing your reviewers by failing to acknowledge their contributions

  - You demonstrate your mastery of the literature

  - You may make new friends! (Scholars love to be cited)

  - Doing so encourages others to cite your work in return

Issue #3: Misappropriating Ideas

- A researcher must never present someone else's ideas as his or her own

  - *Always* cite your source!

  - Even if the originator of the idea doesn't care about credit

- Example:

  Adding "eye of newt" to the mixture produced a higher reaction rate and, ultimately, a far more potent product.[1]

  [1]We are grateful to Mr. A. E. Newman, a high school student who was visiting our lab for the day, for suggesting this important step.

Issue #4: Research Fraud

- A.k.a. *"Painting the Mice"*

  - William Summerlin, dermatologist @ Sloan-Kettering, 1974: announced ability to transplant tissue from unrelated animals, demonstrated with white mice with patches of fur transplanted from a black mouse

    - Problem: his lab assistants discovered the black patches came off with rubbing alcohol…

- When exposed, these cases often generate widespread publicity and end careers. Yet, they keep happening…

Types of Data Fraud (*Honor in Science*, Sigma Xi)

- <u>Trimming</u>: smoothing irregularities to make the data look extremely accurate and precise

- <u>Cooking</u>: retaining only those results that fit the theory and discarding the rest

- <u>Forging</u>: inventing some or all of the research data reported, and/or reporting experiments to obtain those data that were never performed

Favorite (Inexcusable) Excuses for Data Fraud

"Those outlier points must be a measurement error"

"That would only confuse the reader"

"Everybody cleans up their data before publication"

"We're just missing a few data points and we don't have time to run the experiments before the deadline"

Cases & Penalties

- Hwang Woo-Suk (South Korea)

  - Fabricated results to support claim to have cloned human embryonic stem cells

  - Coerced egg donations from female subordinates

  - Co-author at Univ of Pittsburgh found innocent of fraud but reprimanded for "shirking" responsibility to validate results

- Eric Poehlman (U. Vermont)

  - Faked data in 15 NIH grant applications worth $2.9 million over 10 years

  - Sentenced to 366 days in federal prison and fined $200K

- Office of Research Integrity: http://ori.hhs.gov/

  - Annual summaries of scientific misconduct cases and penalties applied

  - Typically: barred for 3-5 years from applying for federal funds or serving on any review/advisory panels. (So…)

Data Management

- Other researchers must have access to data, research materials, and method description required to support the conclusions stated in a publication

  - As much as possible of this should be in the paper

  - Public, published code is good where possible

  - Retain all data (as allowed by IRB, etc.) that appears in or supports a publication

Some Other Issues

- Failure to Disclose

  - Always disclose potential conflicts of interest to avoid accusations of misconduct

  - E.g., don't review papers of people you know

- Intellectual Property [the legal right to control the application (patent) or expression (copyright) of an idea]

  - Most institutions have policies specifying how IP must be handled: how research data are collected/stored, how and when results can be published, etc.

  - May delay publication of results

- Talking to the Public

  - Don't announce discoveries before they're peer reviewed

  - Don't oversell your results

  - Don't announce your project as "The Texas State" study without the university's permission

Dealing with Problems

- Get your advisor's advice

  - If you're having a problem with your advisor, talk to him/her about it before asking for outside opinions

  - Next line of defense: ask to speak confidentially with another senior scientist you trust and respect

- Allegations of misconduct should be handled with as much confidentiality as possible

  - People's careers are at stake

  - Remember that there are two sides to every story

Thank You! Questions?

- Do great work and have a great summer!

- Reminder: Use your resources

  - Faculty advisors

  - Graduate student mentors

  - Me, too!

    - Molly O'Neil, moneil@txstate.edu

    - http://cs.txstate.edu/~mo1162/

Acknowledgments

- This talk was excerpted/adapted from the following presentations:

  - Paul Wagner, "Research Methodology", http://www.cs.uwec.edu/~wagnerpj/talks/researchmethodology.ppt, Student ACM Meeting Fall 2003

  - David S. Touretzky, "Ethics and Etiquette in Scientific Research", Carnegie Mellon University, April 2007

- I also borrowed ideas and advice from:

  - David Chapman, "How to do research in the MIT AI lab", http://dspace.mit.edu/handle/1721.1/41487

  - Christos Kozyrakis, "How to Have a Bad Career as a Stanford Graduate Student", http://csl.stanford.edu/~christos/publications/BadCareer.pdf