Computer Science Research: Asking the Right Questions

Hussein Suleman

PG Research Methods 2021

UCT Computer Science

Introductory Videos

- http://www.youtube.com/watch?v=CkAKBPvDLuo

Graduate degrees

- Masters degree
  - training in how to do research: testing ideas
  - must understand: research methods; analysis; interpretation; appropriate conclusions
- PhD degree
  - training in doing research: coming up with good ideas
  - "significant original contribution to human knowledge"
  - must understand, +masters: how to ask questions; reflection on outcomes; definition of further work
  - should leave with ability to be independent reseacher

Activity 1: Research? Science?

- What is Research?
- What is Science?
- 2 minutes: Think about this, then type your answers in the chat window.

What is Research? (last class)

- systematic investigation of a phenomenon
- systematic way of finding solution to a problem
- investigation to prove/disprove a theory
- process of discovery
- deriving knowledge from experiments
- process in trying to explain a phenomenon
- investigation using scientific methods to find a solution to a problem

What is Science? (last class)

- a collection of proven theories
- systematic way of acquiring knowledge
- method used to solve your problem
- discipline
- proving the correctness of a hypothesis using experiments
- practical activities to do systematic studies through observation and experiments

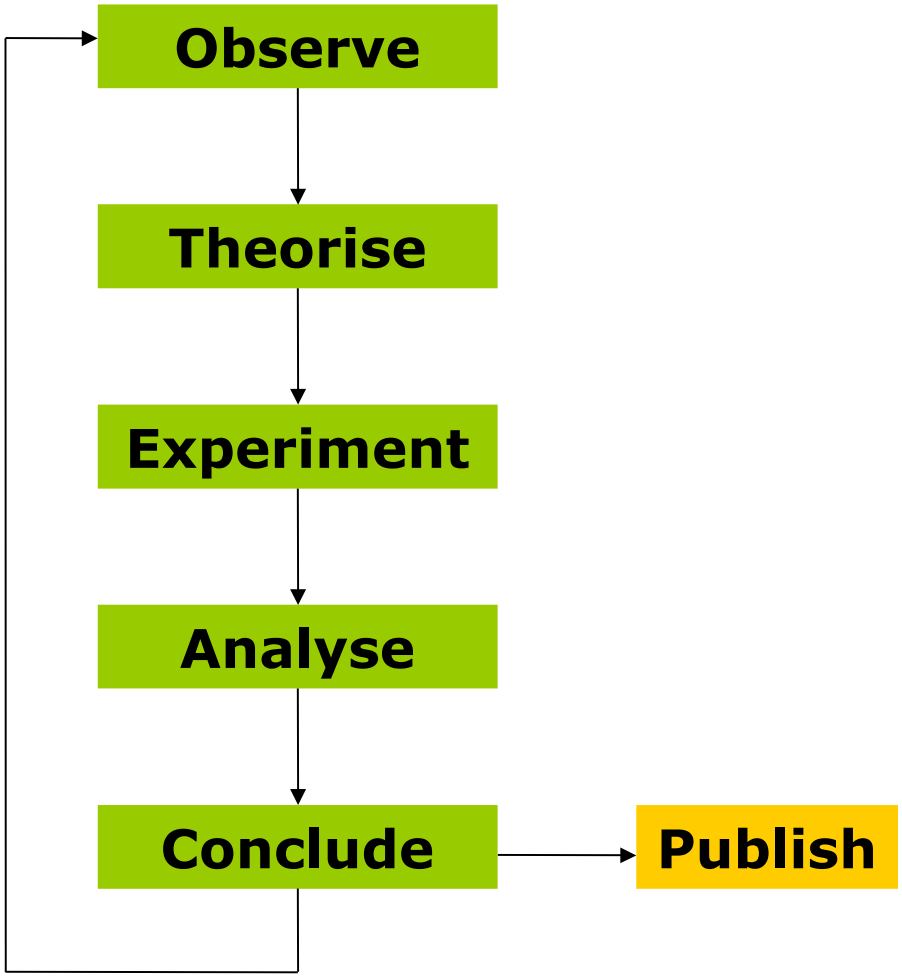
Science and Research

- **Research** is looking for answers.
  - Attempting to understand the nature of being.
- **Original Research** is looking for answers that are not yet known to any other person.
- **Science** is doing research using the scientific method.
- Other types of research methods:
  - the Historical method

What is a Researcher / Scientist?

- A researcher generates/locates knowledge.

- A scientist generates/locates knowledge using the **Scientific Method**.



Exercise 1

- How will you apply each step of the scientific method (observe, theorise, experiment, analyse, conclude) to prove or disprove the following statement:

Since our children sit passively while the television dances, their ability to become deeply involved with books, school teachers, and other less frenetic sources of wisdom—their ability to think—atrophies. It should be no wonder that they abandon books, manifest lower intelligence quotients, fail to achieve academically, and have depressed professional aspirations.

Source: http://www.simpletoremember.com/articles/a/dangers-of-television/

Epistemology

- Epistemology is the study of knowledge.
- Many philosophies in knowledge/research.

- Positivism / Empiricism
  - strictly believe what can be calculated/proven
- Post-positivist
  - believe what can be proven imperfectly in the context of the individual
- Pragmatism
  - not about reality - more about practical implications and environment
- Relativism, Skepticism, ...

Inductive vs Deductive Research

- In all cases, research is evidence-based.
- Major schools of thought:

- Inductive:
  - induce conclusions based on probabilistic support from evidence.
  - bottom-up
- Deductive:
  - make conclusions based on logical reasoning, possibly including collected evidence.
  - top-down

Steps in the Research Process

1. Proposal
2. Importance and Significance of Problem
3. Research Questions/Hypotheses
4. Literature Review
5. Proposed Solution
6. Experiments
7. Analysis of Results
8. Conclusions
9. Publications
10. New Hypotheses, Go to 4, Rinse and repeat

1. The Proposal

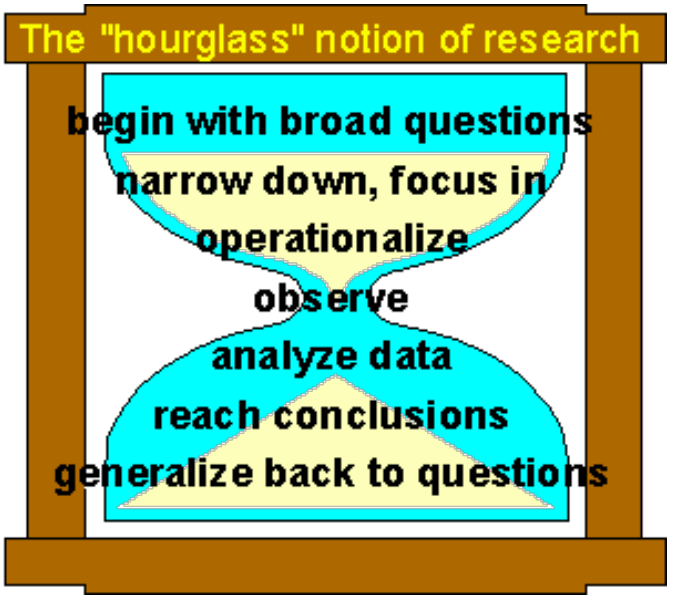
- Formally state the process for the research:
  - To get funding or approval for study.
  - To plan the research.
  - To verify feasibility.
- Structure:
  - Document the problem, solution, approach and possibly expected outcomes.
  - First round of literature review.
  - Experimental methods to be used.
  - Resources needed and possibly risks.
  - Timeframes.

2. Importance of Problem

- Problem must be clearly defined.
  - *grey squirrels are looking less grey over time.*
- If problem is not significant, why do the research?
  - *squirrels could be sick or heading for extinction!*
- Emphasize potential impact of research.
  - *a better understanding of the reasons for less greyness may lead to interventions to prevent extinction.*

Choosing a Research Problem

- Start with broad area of interest.
- Understand existing research and existing practice.
- Narrow down to candidate problems.
- Select viable candidate:
  - interest
  - funding
  - feasibility: time, etc.



http://www.socialresearchmethods.net/kb/strucres.php

3. Research Questions

- Define problem to be solved.
- Guide the research.
- Examples:
  - Is it easier to learn Java, when compared to Python?
  - Is Chrome the fastest Web browser?
  - Does user-centred design improve the usability of scientific software systems?
- Questions must be **answerable, non-trivial, testable** and **falsifiable**.

Answerability

- All research questions should have expected answers.
- You should have some idea about potential answers or types of answers when specifying a question.

- Good RQ: What colour is the dress?

- Bad RQ: How can we improve the speed of Microsoft Windows?
- Why?

Non-triviality

- Is the answer to the research question well-known or easily discovered?
  - If so, it is not a good research question for postgraduate study.

- Trivial RQ: How many days are there in the month of March?
- Trivial RQ: What are the names of Nelson Mandela's children?

Testability

- Can we find the answer to a research question?
- Specifically, can we construct an experiment or conduct an analysis to determine the answer?

- Good RQ: Do first year CS students at UCT in 2021 know who Alan Turing was?
- How will we test this?

- Bad RQ: Did first year CS students at UCT in 1975 know who Alan Turing was?

Falsifiability

- Falsifiability means that it MUST be possible to prove a deductive research question to be false. Some condition must exist that, if it is true, implies a negative answer to the question.

- RQ: Do all first year students understand how a for loop works?
  - underlying hypothesis: they do understand!
- What condition/event will provide a negative conclusion to this question?

Reproducibility

- The answer to a research question should not be vague or applicable to specific data or a specific time (not specified in the question itself).

- Bad RQ: How many white cars are parked outside the CS building?
- Why?

- Good RQ: Why did UCT staff buy more white cars than black cars in 2020?

Good research questions are …

- **Relevant**: Arising from issues raised in literature and/or practice, the question will be of academic and intellectual interest.
- **Manageable**: You must be able to access your sources of data (be they documents or people), and to give a full and nuanced answer to your question.
- **Substantial** and **original**: The question should showcase your imaginative abilities, however far it may be couched in existing literature.
- Fit for **assessment**: Remember, you must satisfy the learning outcomes of your course. Your question must be open to assessment, as well as interesting.
- **Clear** and **simple**: A clear and simple research question will become more complex as your research progresses. Start with an uncluttered question then unpeel the layers in your reading and writing.
- **Interesting**: Make your question interesting, but try to avoid questions which are convenient or flashy. Remember, you will be thinking about this question for an entire year.

Source: *http://www.socscidiss.bham.ac.uk/research-question.html*

Broad vs Specific

- Avoid questions that too broad and not answerable/testable.
- Bad RQ: Why did the chicken cross the road?

- Avoid questions that are too specific and not useful.
- Bad RQ: How many chickens crossed the main road outside Pizza Hut in Rondebosch between 12h00 and 12h05 on 16 March 2021?

Types of Questions

- How can...?
  - can almost never be fully answered...
  - partial answers are not necessarily bad, but can you reword the question instead?
- Can...? What is...?
  - Has a yes/no or trivial answer.
  - partial answers (yes, sometimes) are not necessarily bad, but can you reword the question instead?
- Determining relationships among variables
  - What is the difference between X and Y?
  - Is X related to Y?
  - How fast is algorithm X compared to Y?
- Why do ...? What are the reasons for...?
  - Finding the reasons for an occurrence.

Activity 2A: Research Questions

- Problem X: One of the participants must explain the problem they want to solve (in 2 minutes).

- Devise one **good** research question for this study.
- Submit this to the lecturer by email.

- These will be collected and discussed in the second session.

Activity 2B: Research Questions

- Problem X:

- Assess the research questions for the following aspects:
  - Are the questions clear?
  - Are the questions important?
  - Are they
    - answerable?
    - non-trivial?
    - testable?
    - falsifiable?

Activity 2C: Research Questions

- Problem X:

- Discussion.

4. Literature Review

- Search conference proceedings, journal articles, magazines, etc. for similar work.
  - ACM Digital Library, Google Scholar, Microsoft Academic Search, Citeseer, UCT library, etc.
- **If research questions have already been answered, no need for original research!**

- Create a *synthesis* of related work.
- Relate prior work to current work.
- Ensure that next phases build on known results and do not reinvent the wheel!

5. Proposed Solution

- Typically an experimental software prototype, but not always.
- May involve hardware.
- May be application of existing software.
- May be new process/method/idea/etc.
- Outside of experimental CS, this step is not always present!

6./7. Experiments and Analysis

- Theoretical Proof (Analytical)
- Simulation
- Proof of Concept
- Experimental Evidence
  - User studies (Statistical)
  - Instrumentation and measurement (Statistical)
- Non-experimental Data Collection
  - Surveys
  - Focus groups
- Case Studies
- Body of Evidence

Theoretical Proof / Analytical

- Find answers purely by mathematical calculation.
- Example:
  - Question: What is the fastest sorting algorithm for 5 numbers in reverse order?
  - Solution: Calculate comparisons/swaps and determine answer.

Simulation

- **Discrete Event Simulation** is used where mathematical calculations are not simple because of complexity of scenario, especially multiple interacting bodies.
- Events are created at specific times and a simulator (e.g., NS3, Omnet++) or real system acts on the events.
- Example:
  - Question: Does the Apache Web server fit within 20MB of memory at all times?
  - Solution: Simulate requests based on studying various user patterns and measure memory use.

Proof of Concept

- Show existence only.
- Example:
  - Question: Can we build a mousetrap that is reusable?
  - Solution: Build a reusable mousetrap.
- No need to prove anything else.
- Only applicable if existence not known.

User Studies

- Engage with users in order to answer questions.
- Typically involves getting users to perform some tasks (controlled or uncontrolled) and get feedback directly or indirectly.
- Often involves a post-test survey of users.
- **Usability**, usefulness, **User eXperience**, etc.
- Example:
  - Question: Can first year students at UCT build a more usable DSTV remote control?
  - Solution: All students build a new remote control and test its usability with real users.

Performance Studies

- Test that a system performs within some constraint.
- Can be used for comparative studies.
- Very rigorous and often repeatable.
- May involve **instrumentation** of systems.
- Example:
  - Question: Are Google's search results better than Bing's?
  - Solution: Create list of representative queries and then compare results.
- Example:
  - Question: How does performance of system X compare to system Y?

Surveys / Interviews

- Collect and analyse opinions of users.
- Good to establish problems or issues.
- Not good standalone technique in CS.
- Example:
  - Question: What SE methodology do programmers prefer?
  - Solution: Create survey and get a sample of programmers to answer the survey – then analyse the results.

Focus Groups

- Gather opinions of users working in a group.
- Usually have a facilitator to lead and track discussion.
- Can be used for evaluation but also for requirements gathering.
- Is usually verified by other techniques.
- Example:
  - Question: Is system X usable?
  - Solution: Demonstrate system X to users and/or allow them to use it … then discuss features in detail with users as a group.

Case Studies

- To demonstrate real-world applicability of a theory.
- Used where other proofs fail or to complement other proofs.
- Example:
  - Question: Can we create an efficient and general purpose digital repository toolkit?
  - Solution: Create a toolkit and apply it to some different scenarios. Each application is a case study that provides further evidence of the truth of the research question.

Body of Evidence

- Meta-technique to prove difficult questions.
- Use of a collection of different techniques to provide evidence in as many different ways possible.
- Example:
  - Question: Is Java too slow to be practical?
  - Solution:
    - Show Java to be slow in some applications (Proof of concept)
    - Interview programmers for their opinion (User studies)
    - Measure performance and compare (Performance studies)

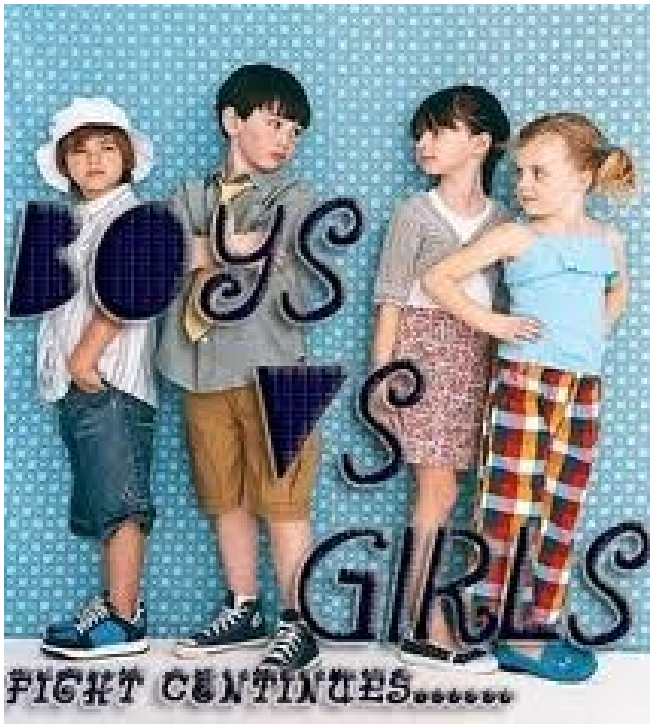
8. Conclusions

- Conclusions must logically follow from results.
- Research questions must be answered:
  - If no answer, then design or execution of research was flawed or encountered problems.
  - Statements could be strong or weak, depending on evidence.
- Make appropriate statements – do not over-generalise!
  - This is not Mythbusters!

Exercise 2

- What experimental techniques can be used to test the hypothesis that girls are better programmers than boys? How?

techniques:
analytical proof
simulation
proof of concept
user studies
performance studies
survey/interview
focus group
case studies
body of evidence



the end ...

to find me, search for ”Hussein Suleman”