INFOMSCIP 2019-2020
lecture 15
Oct 24, 2019

Sources used for the following slides (also recommended for further reading):

- "How (and How Not) to Write a Good Design Paper: A Metaphrase of Roy Levin's and David D. Redell's Evaluation of the Ninth SOSP Submissions" http://www.ida.liu.se/~matar/designpaper.pdf
- Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (p. 38). ACM.
- Zobel, Justin. *Writing for computer science*. Springer, 2015
- Singer, J., & Vinson, N. G. (2002). Ethical issues in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, *28*(12), 1171-1180.

Plus several others (see also references on slides).

Scientific perspectives on GMT (INFOMSCIP)

Literature reviews & related work sections

- Why do a literature review?
  Identifying state-of-the-art, motivating your research and establishing the actual relation
- How to read scientific papers?
  Some tips for doing a literature review
- How to find all relevant literature?
  Snowballing approach, resources, tips & tricks

Why do a literature review & have a related work section?

To make you (review) and reader (related work section) familiar with all necessary knowledge needed to understand your research

⇒ Introduces general context and knowledge that you can build on

Hint: understand your readership.
(E.g., community: AR interaction at CHI vs. ISMAR.)

To identify gaps & shortcomings in state-of-the-art and put your work in context

⇒ Motivates and justifies your research (incl. relevance), establishes that your research is new and adds to scientific knowledge (and how)

⇒ Helps defining and specifying your research question(s)

Hint: Don't just describe the related work,
but also address the actual relation of this work for your research

Identify existing and established aspects related to your work

⇒ Further helps defining and specifying your research question(s)

⇒ Helps with your research methodology and different aspects thereof (e.g., specification of established parameters that can be taken for granted)

Hint: notice the different purpose of citing related work.
Don't just list related works (randomly) but group or categorize it
(rule of thumb: one aspect per paragraph), e.g., based on purpose, relation, …

General tip: check related work sections of good, well-written papers.

Some comments from the paper "**How (and How Not) to Write a Good Design Paper**: A Metaphrase of Roy Levin's and David D. Redell's Evaluation of the Ninth SOSP Submissions" that give hints for good scientific writing. http://www.ida.liu.se/~matar/designpaper.pdf.

Original ideas

Can you state the new idea concisely? If your paper is to advance the state of knowledge, your reader must be able to find the new ideas and understand them.

Hint: it often makes for a strong paper or thesis, if you finish the introduction with a list of contributions.

Are comparisons with previous work clear and explicit?
You cannot simply say: "Our approach differs somewhat from that adopted in the BagOfBits design [3]." Be specific: "Our user interface approach uses tangible pieces rather than a touchscreen as in the BagOfBits design [3], with the expected improvements in co-operation and engagement of several users."

We will come back to this when talking about *literature reviews* next time.

What is the oldest paper you referenced? The newest? Have you referenced similar work at another institution? Have you referenced technical reports, unpublished memoranda, non-reviewed online material, personal communications? The answers to these questions help alert you to blind spots in your knowledge or understanding.

This is also a helpful advice for your literature study.

From lecture 14 (the part on "scientific writing")

Some good examples from thesis projects (that have also been published):

Nina Rosa (thesis published as paper at ICMI 2015)

According to the model by Steuer [30], presence in VR is a human experience and a consequence of immersive technologies. It has two determining dimensions: vividness with the two contributing factors breadth and depth of included modalities; and interactivity with the three contributing factors speed, range, and mapping. Many studies target increasing interactivity by focusing on improving task performance [2, 24, 5]. Yet, when the goal is to improve experiences that users can take part in passively, vividness becomes a vital part of the system, even more urgent than interactivity. Here, we focus on this less studied aspect by specifying the sub-goal of our research as investigating the experience of passive touch under varying visual and auditory conditions.

…

It has been shown that a first person perspective of a life-sized virtual human body alone is sufficient to generate a body transfer illusion [27]. In our experiments we thus apply this methodology focusing on first person body experiences. …

Wendy Bolier (thesis published as paper at ACM MM 2018)

Since the emergence of head-mounted displays (HMDs) on the consumer market, many VR painting tools have been released. Some well-known examples are Tilt Brush [3], Quill [2] and A-Painter [1]. For this study, the choice was made to use A-Painter, as it is open source and thus allows us to make the changes necessary for our research.

…

Although many studies suggest that the drawing of 3D objects is beneficial for the improvement of spatial visualization skills, more research is needed to prove the exact influence of S&D by itself. Furthermore, the existing studies all apply to drawing 3D objects onto a 2D surface using graphical projection. No research exists yet evaluating the benefits of drawing 3D objects in a 3D space.

…

Finally it needs to be emphasized that the VR applications in these studies are completely focused on the training of spatial skills. Concerning spatial visualization and mental rotation, no proof exists that an arbitrary VR application will have a positive influence on these abilities. …

How to read relevant papers?

Hint: check review criteria (see preceding & following slides & assignment) and keep them in mind when reading papers

- Identifying contributions & value
- Recognizing flaws & limitations

Hint: establish a "range of modes of reading" (phrase quoted from Zobel 2014, page 21)

Literature research can easily lead to 100s of papers; skim them first and identify which ones to read carefully and which ones "super carefully"

- Browsing to find relevant papers and get an overview
- Background reading
- Thorough, focused reading of key or complex papers

Papers are not textbooks (don't read them in a similar way).

But also: "*… reading about a paper that seems relevant is never a substitute for reading the paper itself. If you need to discuss or cite a paper, read it first.*", Zobel 2015, page 21).

Verifying the references of a paper is an important part of the review process. Lack of essential citations can lead to rejection.

But how to find them all?

How to find all relevant papers?

Approaches:

- Skim journals of relevant communities
- Skim proceedings of relevant conferences
- Keyword searches to get some papers as starting points

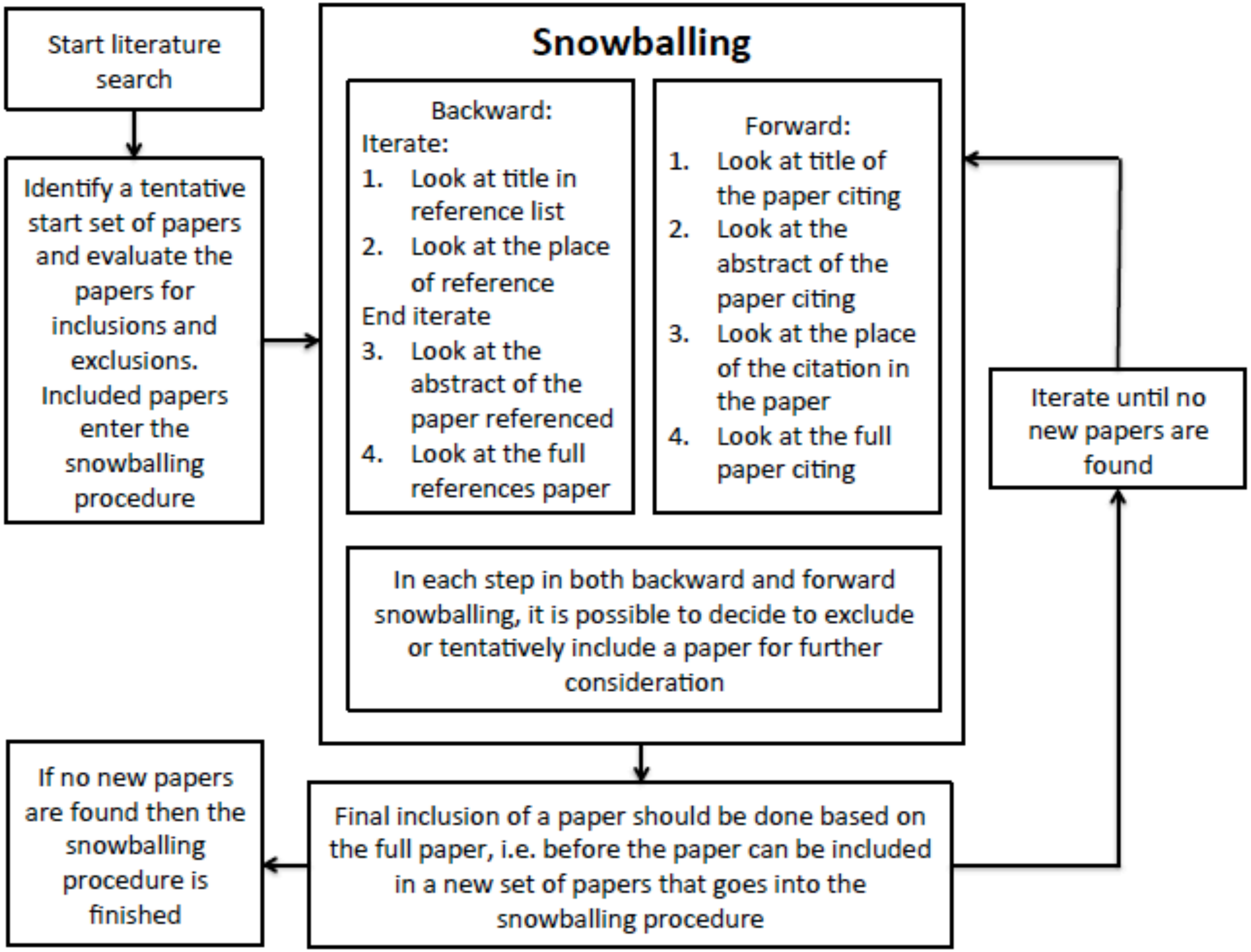
Probably the most common systematic approach is "Snowballing"

- Using the reference list of a paper (*forward snowballing*)
  or the citations to the paper to identify additional papers *(backward snowballing).*

Starting points:

- Known articles and papers
  Generally, you have some knowledge (and sources) when starting with your research
- (Dedicated) search engines
  Examples: Google Scholar, DBLP, Citeseer, ..
- Publications of the related community (or communities)
  Examples: ACM DL, IEEE Computer, …

Figure 1. Snowballing procedure.



Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (p. 38). ACM.

The relevance of citations for identifying a paper's relevance and quality

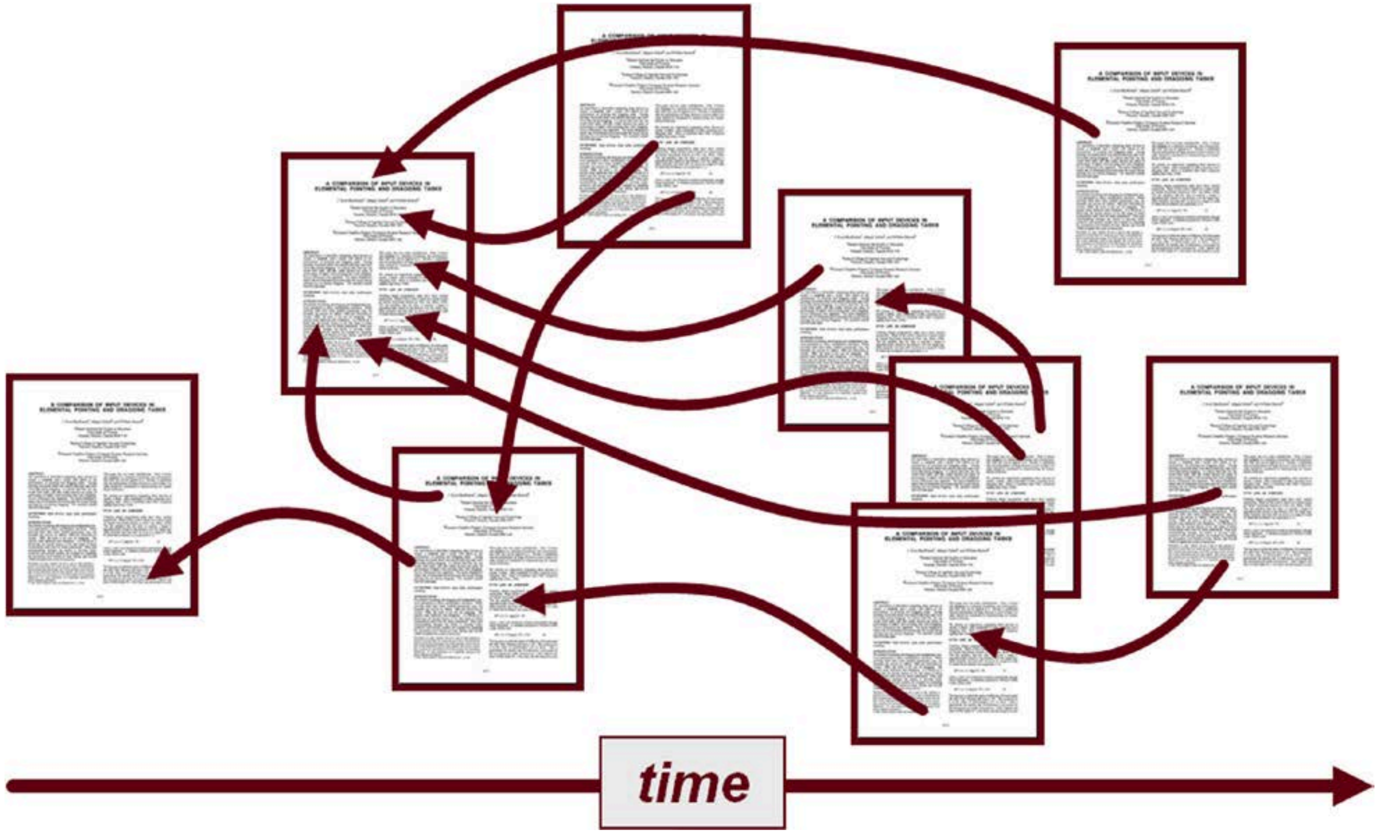


Fig. 4.1: A collection of research papers with citations to earlier papers.

From I.S. MacKenzie: Human-Computer Interaction – An Empirical Research Perspective

Examples for dedicated search engines & systems for scientific research (esp. in computer science)

- Google Scholar, https://scholar.google.com/
  A freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines.
- CiteSeerX, https://citeseerx.ist.psu.edu
  A public search engine and digital library for scientific and academic papers, primarily in the fields of computer and information science.
- DBLP, https://dblp.uni-trier.de/
  A computer science bibliography website.

Also helpful: academic social networking sites

- Academia.edu, https://www.academia.edu/
  A commercial social networking website for academics.
- ResearchGate, https://www.researchgate.net/
  A social networking site for scientists and researchers.

General web search engines:

Can be good to find blogs, Wikipedia pages, news & popular science articles, etc.
Can be relevant for scientific work, too, but only if no related scientific resource exists

Organizations, publishers, communities

- ACM, https://www.acm.org/
  ACM Digital Library, https://dl.acm.org/
- IEEE, https://www.ieee.org/
  IEEE Computer Society Digital Library, https://www.computer.org/csdl/home
  IEEE Xplore, https://ieeexplore.ieee.org/Xplore/home.jsp

Also check the related special interest groups (ACM) or technical committees (IEEE), and websites of relevant groups and researchers in the field you are working in

Open access initiatives and related archives

- E.g., arXiv.org, https://arxiv.org/
  A repository of electronic preprints (known as e-prints) approved for posting after moderation, but not full peer review.

Some comments on access to papers via UU resources

https://www.uu.nl/en/university-library/help-in-searching/online-access

University Library

Searching for literature | Help in searching | Services | Practical information | About the library | Contact

Searching for literature

Help in searching

- **Online access**
  - Get Access button
  - UBUlink
  - For guests
- Search engines explained
- Search strategy
- Types of literature
- By discipline
- LibGuides

Services

Practical information

About the library

Contact

Online access

When you are searching for literature online, the university library offers access to scientific sources.

CONNECTED TO A UU NETWORK

Are you logged in with your Solis ID and password or UMC account on a desktop computer or connected to the Eduroam Wi-Fi network of Utrecht University on your own device? Then you have access to all subscripton-based online resources of the library.

NOT CONNECTED TO A UU NETWORK

Log in by clicking on the button below to get access from home to licenced sources via the library. If you consult the search engines via the links which you see after logging in, you have access to the publications, **provided that it is licenced material** or freely available material.

Direct login

In class activity:
Literature research

Ethical aspects related to publications

Zobel, Justin. *Writing for computer science*. Springer, 2015.
Chapter 17: **Ethics**

Quotes from the introduction:

Science is built on trust. Researchers are expected to be honest, and research is assumed to have been undertaken ethically.

…

The major societies of science have **codes of conduct** that scientists are expected to adhere to. Breaches of these codes are regarded as extremely serious; even senior, respected academics have lost their positions after having been found to commit misconduct. Familiarity with these codes and their implications for day-to-day work is essential for a practicing scientist.

Intellectual Creations

Many of the ethical issues that arise in the context of research are related to the ***ownership of both ideas and descriptions*** of them. Loosely, these might be described as intellectual creations, that is, the stuff that knowledge workers make, including concepts, inventions, discoveries, designs, documents (text, images, or video), or code. People might own their intellectual creations, or the creations might belong to their employer, or a publisher.

Created content, legal issues

- Copyright of produced material
- Ownership of intellectual property (e.g., patents from ideas)

Plagiarism

A central element of the process of science is that each paper is an original contribution of new work. Scientists' reputations are built primarily on their papers: both the work and how it is reported.

Self-plagiarism

Authors who re-use their own text may well be plagiarizing. Using the same text in two papers is a step in the direction of publishing the same work twice.

Misrepresentation

Misrepresentation occurs when a paper does not accurately reflect the outcomes that were observed or the contributions of previous research. When presenting results, researchers are expected to ensure that they are accurate, describe any experimental issues or limitations that could have affected the outcome, provide enough detail to enable reproduction or verification, be fair in description of other work, report negative as well as positive results, not state falsehoods, and take the effort to ensure that statements are complete and accurate. …
However, an honest mistake is not misconduct. …
In its clearest form, misrepresentation is fraud: the making of claims that are outright false. Another form of misrepresentation is when authors imply that they have high confidence in their results when in fact the experiments were preliminary or were limited in some way.

Example for plagiarism

"Another form of misjudgement is inappropriate or inadequate citation. Suppose that Barlman and Trey (2001) wrote the following:

> The impact of viruses has become a major issue in many large organizations, but most still rely on individual users maintaining virus definitions, with no internal firewalls to protect one user from another. However, any structure is only as strong as its weakest link; these organizations are highly vulnerable.

It would then be considered plagiarism to write the following:

> Viruses have become a major issue in many large organizations, but most organizations still rely on users maintaining virus definitions on their individual computers, with no internal firewalls to protect one computer from another. However, any structure is only as strong as its weakest link; these organizations are highly vulnerable to infection (Barlman and Trey 2001).

In this example, a citation is given, but it isn't made clear that the citation refers to the whole block of text. Also, there is nothing to indicate that the wording is unoriginal—despite a few small changes, the text is copied. If the wording or the sense of the original text is required, it would instead be appropriate to write something like the following:

> As discussed by Barlman and Trey (2001), who investigated the impact of viruses in large organizations, "most still rely on individual users maintaining virus definitions, with no internal firewalls to protect one user from another. However, any structure is only as strong as its weakest link; these organizations are highly vulnerable."

Alternatively, the essence of the original can be concisely summarized, with clear attribution:

> Barlman and Trey (2001) investigated the impact of viruses in large organizations. They found that organizations are vulnerable if individuals fail to keep virus definitions up to date, as internal firewalls are rare."

Interesting question: is this text (copied from the referenced book) properly cited here?

Authorship

Deciding who has merited authorship of a paper can be a difficult and emotional issue. A broadly accepted view is that each author must have made some significant contribution to the intellectual content of the paper. Thus directed activities such as programming do not usually merit authorship, nor does proofreading. An author should have participated in the conception, execution, or interpretation of the results, and usually an author should have participated to some degree in all of these activities. The point at which a contribution becomes "significant" is impossible to define, and every case is different, but neither code-cutting under the direction of a researcher nor management roles such as obtaining funding justify authorship. Nor is it appropriate to give authorship as a reward or favour.

Confidentiality and Conflict of Interest

Researchers need to respect each other's privacy. Sharing of a computer system with other people does not mean that one has the right to use their data without permission, for example, or to disclose their results to other people. Code or executables may be made available under terms such as commercial-in-confidence, and the fact that many people use commercial software they haven't paid for does not mean that it is appropriate for researchers to do so.

Ethics and scientific integrity

Scientific integrity refers to honest and transparent methods and reporting of research. In other words, there is no cooking the books or even subtly influences of bias to privilege some lines of investigation or results over others. There is no falsification or fabrication or plagiarism.

https://en.wikiversity.org/wiki/Scientific_integrity

The role of ethics in science
Ethics is a set of moral obligations that define right and wrong in our practices and decisions.

https://www.visionlearning.com/en/library/Process-of-Science/49/Scientific-Ethics/161

Notice that ethics do not just address your actual research, but also its ethical consequences.

Ethics in Software Experimentation

- Any empirical research activity involving human subjects must take ethical aspects into consideration.
- Singer and Vinson (2002) provided practical guidelines for the conduct of empirical studies.
- They identified four key principles:
  1. Scientific Value
  2. Beneficence
  3. Informed Consent
  4. Confidentiality

Singer, J., & Vinson, N. G. (2002). Ethical issues in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, *28*(12), 1171-1180.

Source: course "Advanced research methods" from the MBI master program at UU, slides on "**Ethics in Research**"

Four Ethical Principles

1. **Scientific Value** (*your win*)
   The study should have *scientific value* in order to motivate subjects to expose themselves to the risks of the empirical study.

2. **Beneficence** (*their win*)
   Weighing risks, harms and benefits, the *beneficence* must overweigh, not only for the individual subjects, but also for groups of subjects and organizations.

3. **Informed Consent**
   Subjects must give *informed consent* to their participation, implying that they should have access to all relevant information about the study, before making their decision to participate or not.

4. **Confidentiality**
   Researchers must take all possible measures to maintain *confidentiality* of data and sensitive information, even when this is in conflict with the publication interests.

Source: course "Advanced research methods" from the MBI master program at UU, slides on "**Ethics in Research**"

2. **Beneficence** (*their win*)
Weighing risks, harms and benefits, the *beneficence* must overweigh, not only for the individual subjects, but also for groups of subjects and organizations.

- Risk of harm to Participants
  - Disrupts participant's work
  - Results of the researcher may devaluate participant's work
  - Publication of study may harm the company's business
- Benefits of study
  - Scientific value: useful to society?
  - Depends on importance of the research topic
  - Note: validity is crucial – invalid results means the study has no benefits
  - May also be specific benefits to participants,
    e.g., training, exposure to state-of-the art techniques, etc.
- Beneficence: Benefits should outweigh the risks
  - Understand and justify any tradeoffs in the design of the study

Source: course on "Empirical Research Methods for Computer Scientists"
by S. Easterbrook and B. Barbosa Neves, University of Toronto,
"**Research Design and Ethics**", http://www.cs.toronto.edu/~sme/CSC2130/

Question: Should you pay your participants?

I once reviewed a paper that was ultimately rejected because (among other things) authors provided a small monetary reward to subjects (1 cent per score point in a game). Any idea why?

- Arguments in favor
  - Can help with recruitment
  - Compensate participants for their time
- Arguments against
  - May induce participants to take risks they otherwise would not take
  - May get expensive (esp. if rates are to be more than a token)
- Issues
  - Institutional review boards might have standard rates; might be too low for professional Software Engineering
- Alternatives:
  - All participants entered into draw for some new gadget

Source: course on "Empirical Research Methods for Computer Scientists" by S. Easterbrook and B. Barbosa Neves, University of Toronto, "**Research Design and Ethics**", http://www.cs.toronto.edu/~sme/CSC2130/

3. Informed Consent

Subjects must give *informed consent* to their participation, implying that they should have access to all relevant information about the study, before making their decision to participate or not.

- Elements
  - Disclosure – participants have full information about purpose, risks, benefits
  - Comprehension – jargon-free explanation, so participants can understand
  - Competence – participants must be able to make rational informed choice
  - Voluntariness – no coercion or undue influence to participate
  - Consent – usually indicated by signing a form
  - Right to withdraw
    - Participant can withdraw from study at any point without having to give reasons
    - Participants can request their data to be excluded (might not be possible!)
- Challenges
  - Student participants: perception that their grade will be affected if they don't participate
  - Industrial participants: perception that the boss/company wants them to participate

We often use our own students for experiments. What are potential problems with that?

Source: course on "Empirical Research Methods for Computer Scientists" by S. Easterbrook and B. Barbosa Neves, University of Toronto, "**Research Design and Ethics**", http://www.cs.toronto.edu/~sme/CSC2130/

4. Confidentiality

Researchers must take all possible measures to maintain *confidentiality* of data and sensitive information, even when this is in conflict with the publication interests.

- Protecting anonymity
  - Do not collect any data (e.g., names) that allow participants to be identified
  - But you need a signed consent form, so …
  - Sever participant's identity from their data before it is stored and analyzed
  - Researcher-subject interactions should be held in private
- Protecting the data
  - Consent form states who will have access to the data, and for what purpose
  - Raw data should be kept in a secure location
  - Reports should only include aggregate data
- Exceptions
  - When it is impossible to identify individuals from the raw data
  - When more harm results from maintaining confidentiality than breaching it

Source: course on "Empirical Research Methods for Computer Scientists" by S. Easterbrook and B. Barbosa Neves, University of Toronto, "**Research Design and Ethics**", http://www.cs.toronto.edu/~sme/CSC2130/

Singer & Vinson – Discussion questions

- How much detail do participants need about the study?
- How can we ensure confidentiality in the era of data mining?
- Is it ever ethical for course instructors to run controlled experiments on different pedagogical techniques?
- How important is it to share the results of your study with the participants?
- What other codes of ethics apply (e.g., ACM?)

Source: course on "Empirical Research Methods for Computer Scientists" by S. Easterbrook and B. Barbosa Neves, University of Toronto, "**Research Design and Ethics**", http://www.cs.toronto.edu/~sme/CSC2130/

Reviewing

Peer-review organization

- Journal articles
- Magazine articles

↓

- Steering committee
- Editor in chief
- Editorial board
- (Reviewers) invited, not fixed

- Conference papers
- Symposia papers
- Workshop papers

↓

- General chairs
- Technical program chairs,
  Poster chairs,
  Demo chairs,
  …
- Technical program committee (reviewers)

Note that this can vary. That's why a look at the organization and review procedure can tell you something about the quality of the publication / event.

Indications for good conferences (and journals):

- Review procedure (number, length of reviews, …)
- Meta reviews
- Reviewer discussion
- Author rebuttal
- In-person meeting of PC, …

From lecture 14 (the part on "scientific publications")

Typical review procedure

	Authors	Organizers	
(authors)	Submit paper		
		Assign to reviewers	*(program chairs)*
		Reviewing (min. 2)	*(program committee)*
(authors)	Rebuttal		
		Discussion	*(PC members)*
		Meta review	*(area or PC chairs)*
		PC meeting	*(PC or chairs)*
		Decisions sent	*(program chairs)*
(authors)	Revise if accepted		
(authors)	Submit final version		
(one author)	Present at event		
		Publish proceedings	*(publication chair)*

Note that not all these steps may apply. Again, this can be an indication for quality (or lack thereof) of an event.

From lecture 14 (the part on "scientific publications")

Peer review process: problems, pitfalls, what can go wrong, …
(or: why can you find bad papers, even at top events?)

- Not all review processes are done that thorough.
- Fairness? Double-blindness cannot always solve this.
- Reviewer bias (remember the importance of citations for one's career!)
- Sloppy, low quality work.
- Mistakes happen.
- …

Other reasons?

Note: While the scientific peer-reviewing process is well established (and works to some degree) it has *many* flaws, too.
Thus, criticism and alternative approaches exist (e.g., open publication initiatives).

From lecture 14 (the part on "scientific publications")

Assignment 7:
Paper review

(implicit learning goal:
improve your reading and writing skills)