Research Conduct and Academic Practice

Dana McKay
dana.mckay@unimelb.edu.au

Today's lecture

- Being a good colleague
- Research ethics
- Plagiarism
- What to do if you have problems

Why is this important

- Think about what is the right thing to do (not always clear cut!)
- How to avoid doing the wrong thing
- How to tell when the wrong thing is being done to you
- How to resolve challenges

Thanks to Justin Zobel for the framework and many examples

Covers

- Authorship
- Fraud
- Plagiarism and intellectual property
- Conflict of interest
- Research on human subjects
- Privacy of computer accounts
- Sexual harassment
- Racial vilification

'It is a basic assumption of institutions conducting research that their staff members are committed to high standards of professional conduct.' (NHMRC/AVCC, 1997.)

Being a good colleague

- Privacy and harassment
- Authorship
- Working with others

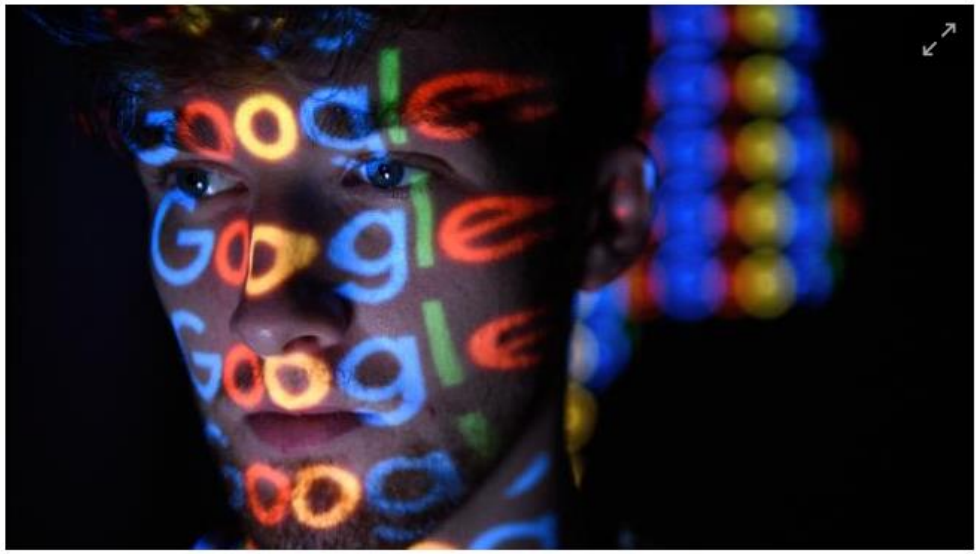
Harassment

Exclusive: Here's The Full 10-Page Anti-Diversity Screed Circulating Internally At Google

Kate Conger
Aug 6, 2017, 8:00am· Filed to: feature

Share

James Damore: Why I was fired by Google



James Damore, the Google engineer fired after circulating a controversial manifesto on diversity, says the company is "like a cult." Picture: Getty

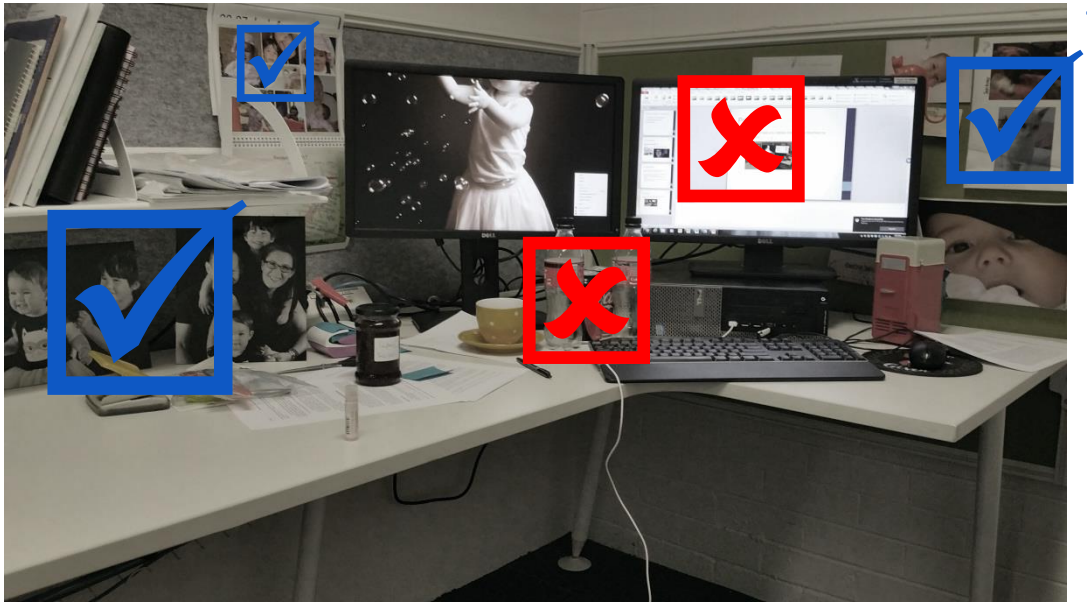
Harassment

Respect others: Age, sex, marital status, sexuality, religion, parental status, disability, race….

- Be conscious of how you share your environment (smells, large objects, speech)
- Be kind to those around you, graduate study is stressful
- You don't have to put up with bullying (even by a supervisor or manager)
  - Buying a colleague a coffee as a nice gesture vs. "where's my coffee?"
  - A supervisor who is too busy to read a paper doesn't get authorship

Privacy: Desks

Justin would tell you desks should be treated as invisible but….



Things you can talk about (with sensitivity)

Things you never talk about

- This will be different for everyone—be cautious
- Never touch or take anything

Privacy: Information

Research data

- If you collected it, keep it private
- If you didn't collect it, it is not yours
  - Even if it is on someone's desk
  - Even if it is on a visible file share
- Sharing can be arranged with consent

Personal data

- Don't share phone numbers, non-uni emails, conversations without permission

Authorship issues

Example (from Touretzky):

Professor Smith, an acknowledged expert in the field, is invited to write an article for an upcoming special issue of Hacker's Monthly.

Smith asks grad student Jones to help.

Some of the most important results in the article are the product of Jones's thesis research.

Authorship issues

What are the possible outcomes?

1. As the invited party, Smith appears as the sole author. Jones' work is referred to as being 'done in my lab' but Jones is not mentioned explicitly
2. Smith appears as the sole author, citing Jones' thesis as 'in preparation'
3. Smith and Jones are joint authors
4. Jones agrees to results appearing in the article, but only if the author order is 'Jones and Smith'
5. Jones had been planning to send an article to Hacker's Monthly, and so declines Smith's offer of collaboration. Instead Jones sends the editor of Hacker's monthly a title and abstract promising a full manuscript in time for publication

Who is an author? University policy

2.1. Authorship is attributed only when a researcher has made a significant intellectual or scholarly contribution to a research output and is willing to take responsibility for the contribution. Researchers qualify as authors if they have made a significant intellectual or scholarly contribution through at least one, but often more than one, of the following:

- 2.1.1. Conception and design of the research described in the research output,
- 2.1.2. Acquisition of research data where the acquisition has required significant intellectual judgement or input,
- 2.1.3. Analysis and interpretation of research data,
- 2.1.4. Drafting of the research output or redrafting the research output so as to critically change or substantively advance the interpretation.

Source: http://policy.unimelb.edu.au/MPF1181

Authors must be listed

2.2. A person who qualifies as an author must not be included or excluded without their written agreement and a record of this agreement must be kept.

- 2.2.1. A written agreement for exclusion is not required where supervisors of student researchers are not typically included as authors on research outputs by student researchers in accordance with conventions of authorship for the discipline and as codified in the Faculty or Graduate School Guidelines for Authorship.

2.3. The record of authorship agreement must include a description of the contribution that each author made to the research output. The record of authorship agreement may be informal (e.g. email, letters, etc.).

Source: http://policy.unimelb.edu.au/MPF1181

What about non-authors?

2.4. Authorship must not be attributed when a researcher has not made a significant intellectual or scholarly contribution to a research output or is unwilling to take responsibility for their contribution. Contributions made solely through the provision of funding, the provision of technical support, technical advice or technical assistance, their position or as a gift, the provision of materials, infrastructure or access to equipment, are generally not considered to meet the criteria for authorship described in 2.1.

2.5. Contributions to the research output that do not meet the authorship criteria described in 2.1 must be properly recognised by acknowledgement where this is appropriate for the type of research output. This may include contributions made by researchers, funding bodies, research infrastructure facilities and organisations.

Source: http://policy.unimelb.edu.au/MPF1181

Challenges in authorship

- Author order: Alphabetical, primary author, turn about?
- Authorship or acknowledgement for supervisor/co-worker
- Conflict:
  - Discuss the issue before a paper is written
  - Resolution can be difficult
  - Everyone has an emotional investment
  - It is never wrong to argue your case
  - Don't give in for fear of causing offence
  - Consider the advice of wise people

Working well with others

Rights	Responsibilities
To be treated with respect	To treat others with respect
To be spoken to honestly	To be honest
To have a safe workplace	To address OHS issues
To have an available manager/supervisor	To use appropriate working practices
To have confidentiality respected	To respect confidentiality
To get credit where it is earned	To not take credit where it is unearned

These are universal:
Dislike, anger, and incompetence do not invalidate them

Research ethics

- Human subjects
- Fraud
- Conflict of interest

ARC guidelines

- A strong research culture will demonstrate
  - Honesty and integrity
  - Respect for human participants, animals and the environment
  - Good stewardship of the public resources used to conduct research
  - Appropriate acknowledgement of the role of others in research
  - Responsible communication of results

Source: https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf

ARC guidelines

- Misconduct in research includes:
  - Fabrication of results
  - Falsification or misrepresentation of results
  - Plagiarism
  - Misleading ascription of authorship
  - Failure to declare and manage conflicts of interest
  - Deviations from [the ARC] code that occur through gross or persistent negligence
  - Wilful concealment or facilitation of research
- It does not include
  - Honest errors
  - Honest differences of interpretation or data judgement

Human subjects

- Herb Green NZ National womens
- Delayed treatment of cervical cancer or pre-cancer over 20 years
- Up to 26 women died
- 1/3 developed invasive cancer

Human subjects: declaration of Helsinki

- The agreement from which all ethical principles on human research flows
  - Article 20: Participants must be voluntary and informed
  - Article 21: Participant data must be treated confidentially, and participants have the right to safeguard themselves
  - Article 22: Informed consent includes knowing the risks, knowing one doesn't have to participate, and knowing they can withdraw
  - Article 5: The health and wellbeing of participants is more important than the science

Source: http://www.who.int/bulletin/archives/79%284%29373.pdf

Human subjects here

- You must apply for ethics approval
  - This takes about a month *if* your proposal is straightforward
  - You need to work with your supervisor
  - You will nearly always need written consent
- Be clear on
  - Who you are asking
  - What you are asking people to do
  - What data you are collecting (be specific)
  - Why you are collecting that data
  - How you will protect the data once collected
- Your participants are people. Treat them like you would treat someone you care about

Fraud

From the ARC code:

- Fabrication of results
- Falsification or misrepresentation of results

In other words

- Don't make results up
- Don't "massage" data to make it match your theory
- Report negative results
- Don't imply anything that isn't true

Fraud types

- **Trimming:** smoothing irregularities to make data appear accurate and precise
- **Cooking:** retaining only those results that fit the theory and discarding others
- **Forging:** inventing some or all of the research data or reporting experiments that never happened
- **Misrepresentation:** Overstating results from an early experiment, making inaccurate broad sweeping statements, minimising or falsely describing others' work

Source: Honor in Science

Fraud: William McBride

News › World

'Thalidomide doctor' guilty of medical fraud: William McBride, who exposed the danger of one anti-nausea drug, has been disgraced by experiments with another, writes Robert Milliken in Sydney

From ROBERT MILLIKEN in Sydney | Saturday 20 February 1993 00:02 GMT | 0 comments

- Destroyed two scientific careers
- No approved morning sickness drugs (still), not financially worth it

Avoiding fraud

- ARC guidelines
  - Retain research data and primary materials.
  - …retain data…for sufficient time to allow reference to [it] by other researchers.
  - Research data should be made available for use by other researchers unless this is prevented by ethical, privacy, or confidentiality matters.
  - Keep clear and accurate records of the research methods and data sources.
  - Retain research data, including electronic data, in a durable, indexed and retrievable form.
  - Maintain confidentiality of research data and primary materials.

Record keeping

- Computer science notebooks
  - Experiment descriptions
  - Inputs
  - Outputs
  - Standard records of experiments including parameters
- Ethics consent forms
- Notes or transcriptions
- Recordings

Conflict of interest

- Sources of conflict
  - Reviewing a colleague's papers
  - Use of licensed data (e.g. commercial in confidence). What does the license cover?
  - Use of data gathered with ethical consent—what does the consent cover?
  - Use of e.g. loaned software written by others
  - Use of ideas from a paper you reviewed

There is no such thing as the appearance of a conflict of interest—someone has a conflict even if they do the right thing.

Plagiarism

- Intellectual property
- Copyright
- Stakeholders
- Plagiarism

Intellectual property

- The product of knowledge work
  - Ideas
  - Documents, images, video, audio…
  - Innovations, designs, prototypes…
- Might be owned by individual or their employer
- Copyright protects the *expression of ideas*, not the ideas themselves
- Patents etc protect the *ideas themselves*
- Fraud and misrepresentation laws protect *identity*: who said what first

Copyright in academic publications

- Copyright is usually assigned to publishers on publication
  - Diagrams text etc belong to the publisher
  - Author must seek permission to reproduce
- Publishers earn income by providing access to refereed papers
- Academics want papers widely accessible
- Consider licensing to publishers
- Services such as ACM authorizer

Stakeholders

- Publishers: Own the text, diagrams etc
- Institutions: Protect reputation by punishing plagiarism or covering it up
- Funding organizations, who want to see their money has been well spent
- Researchers
  - Build careers through publication
  - Rely on co-authors to do the right thing
- Students
  - Plagiarise through lack of knowledge
  - Plagiarise to cheat

Plagiarism

Monash head leaves over plagiarism charge

EducationGuardian.co.uk

The British vice-chancellor of Australia's biggest university is to stand down after admitting plagiarism.

Australian news websites reported last week that David Robinson would be leaving his post at Monash University in Melbourne.

Professor Robinson had admitted two incidents of plagiarism in the 1970s, but

- Plagiarism in this case was >30 years old!

Plagiarism

- Re-use without attribution of images or figures (even your own)
- Unattributed copying of short blocks of text
- Copying of a couple of paragraphs or more with some rearrangement and word substitution
- Use in detail of a previous paper without attribution
- Use of previous results without attribution
- Unattributed use of results from authors not on the paper

Legitimate re-use

- Re-using figures cited and with permission
- Extending a conference paper to a journal paper with full attribution and within the rules of the journal (usually 30% new content)
- Use of a sentence or definition in quotation marks and properly attributed

Example: Unattributed copying

Source: Information retrieval is a difficult problem because it requires describing information you do not yet have (Borgman 1996)

Wrong: There are many challenges with information retrieval, notably that it requires describing information you to not yet have. Because…

Right: There are many challenges with information retrieval, notably that it 'requires describing information you to not yet have' (Borgman 1996). Because…

or

There are many problems with information retrieval, including the requirement that users form a query that represents information they have never seen (Borgman 1996). Because…

Examples of plagiarism

- Academics A, B and C write a paper together, B writes the background section. C (a junior researcher) later extends the work In three more papers with A. C copies the background section, discarding the bits he doesn't need, in three new papers which are then submitted to a conference. B is not an author on these new papers
- Group Z writes a paper J that uses short, attributed sentences from W, X, Y, and Z in the literature review. They go on to write several more papers in the same area, re-using their literature review. The citations to W, X, Y and Z are progressively lost, and only J is cited.

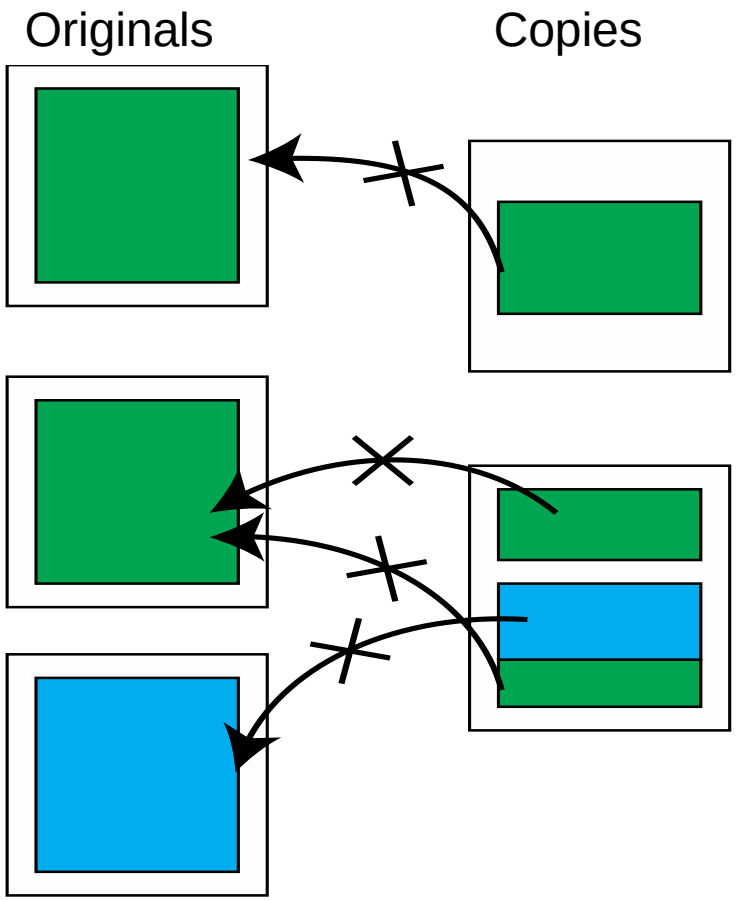
Self-plagiarism

- Re-using your own text without attribution
- ARC guidelines:

  > It is not acceptable to include the same research findings in several publications, except in particular and clearly explained circumstances, such as review articles, anthologies, collections, or translations into another language. An author who submits substantially similar work to more than one publisher, or who submits work similar to work already published, must disclose this at the time of submission'

- Banned by ACM and IEEE
- Often not punished, but not great to be known by senior academics as ethically marginal

An example of self-plagiarism



This document had only 3% new content!

Illustration taken from Buchanan and McKay (2017): The Lowest Form of Flattery: Characterising Text Re-Use and Plagiarism patterns in a Digital Library Corpus

Constantinos V. Papadopoulos

- Reported by Chris Jesshope (see http://infolab.stanford.edu/~shiva/SCAM/end.html)
  - Numerous papers containing plagiarism
  - In some cases source material could not be identified
  - Some innocent bystanders included as 'authors'
  - Plagiarist had already been expelled from one graduate school for plagiarism

Published papers

'On the Heterogeneity of Distributed Databases – Integrating Commit Protocols', C.V. Papadopoulos, IEEE Proceedings of the 14th International Conference on Distributed Computing Systems, Poland, June 1994. (Origin: 'Integration of Commit Protocols in Heterogeneous Databases', A. Tal, R. Alonso, Princeton technical report TR-375-92.)

'Provably Optimal Algorithms for Signal Routing', C.V. Papadopoulos, Journal of Computer Systems Science and Engineering, November 1994, vol.9, no 4. (Origin: 'Provably Good Performance-Driven Global Routing', J. Cong, A. Kahng, G. Robins, UCLA technical report CSD-910013, April 91, and IEEE CAD, vol 11, no 6, 1992.) '

A New Hashing Algorithm for Parallel Processors', C.V. Papadopoulos, Parallel Algorithms and Application Journal, Vol 4, November 1994. (Origin: 'Work Efficient Hashing on Parallel and Vector Computers', T.J. Sheffler and R.E. Bryant, CMU report MCU-CS-92-172.)

'On the Parallel Execution of Combinatorial Heuristics', C.V. Papadopoulos, IEEE Proc. First International Conference on Massively Parallel Computing Systems (MPCS), Los Alamitos, CA, May 1994 . (Origin: 'A Massively Distributed Parallel Genetic Algorithm (mdpGA)', S. Baluja, CMU technical report CMU-CS-92-196R.)

On the Parallelism of Data', C.V. Papadopoulos, Published at: Proc. Parallel Languages and Architectures Europe, PARLE'94, Athens, 1994. (Origin: 'Relations + Reductions = Data-Parallelism', V. Austel, R. Bagrodia, M. Chandy, M. Dhagat, UCLA report CSD-930009, later in the J. of Parallel and Distributed Computing.)

'Optimal Algorithms for Online Scheduling of Parallel Process', C.V. Papadopoulos, Proc. Parallel Languages and Architectures Europe, PARLE '94, Athens, 1994.) (Origin: 'Optimal Online Scheduling of Parallel Jobs with Dependencies', A. Feldman, M.-Y. Kao, J. Sgall, S.-H. Teng, CMU report CMU-CS-92-189, Proc. STOC'93, 25th Annual ACM Symposium on Theory of Computing.)

'A Formal Study on the Fault Tolerance of Parallel and Distributed Systems', C.V. Papadopoulos, IEEE Proc. First International Conference on Architectures and Algorithms for Parallel Processing (ICA3PP 95), Australia, April 1995. (Origin: 'On the Fault Tolerance of Some Popular Bounded-Degree Networks', T. Leighton, B. Maggs, R. Sitaraman, Princeton report CS-TR-385-92.)

Submitted but not accepted

'Maintaining Cache Coherency for Shared Memory Multiprocessors', C.V. Papadopoulos and E.G. Karagianni, Submitted to EURO-PAR'95. (E.G. Karagianni may not be a real person.) (Origin: 'Cache Coherency for Shared Memory Multiprocessors Based on Virtual Memory Support', K. Petersen and K. Li, Princeton report 1992, TR-400.)

'Implementation of Dynamic Data Structures on Distributed Memory Multiprocessors', C.V. Papadopoulos, Submitted to EURO-PAR'95. (For this one he claimed to work at IBM Zurich.) (Origin: 'Supporting Dynamic Data Structures on Distributed Memory Machines', A. Rogers, M. Carlisle, J. Reppy, L. Hendren, Princeton report TR-447-94, also to appear in ACM Transactions on Programming Languages and Systems.)

'A Multiprocessor Architecture for Concurrent Data Structures', C.V. Papadopoulos, Submitted to EURO-PAR'95. (IBM Zurich again.) (Origin: 'Transactional Memory: Architectural Support for Lock-Free Data Structures', M. Herlihy, J. Elliot, B. Moss, DEC technical report CRL-92-07.)

'Verification of Concurrent Objects', C.V. Papadopoulos and E.G. Karagianni, Submitted to 5th Hellenic Informatics Conf, 1995, Greece. (Origin: 'A Library of Concurrent Objects and Their Proof of Correctness', J.M. Wing and C. Gong, CMU report CMU-CS-90-151.)

More submitted but not accepted

'Improving Applications' performance on distributed mshared memory multiprocessors with cache coherence protocols', C.V. Papadopoulos and {innocent bystander}, Submitted to EUROMICRO-95. (Origin: 'Application Specific Protocols for User-Lever Shared Memory', B. Falsafi, A. Lebeck, S. Reinhardt, J. Schoinas, M. Hill, J. Larus, A. Rogers, and D. Wood, Proc. Supercomputing'94, Nov 1994, Linkoping, Sweden.)

'Concurrency Control of Atomic Accesses in Distributed File Systems', C.V. Papadopoulos, Submitted to: EUROMICRO'95. (Origin: 'An optimal algorithm for concurrent accesses to a single replicated file', X. Jia, Y. Zhang, IFIP Trans. A, 1992, vol A-12, and Proc. Algorithms, Software, Architecture, Information Processing, Madrid, Spain.)

'A formal study of networks reliability' C.V. Papadopoulos and {innocent bystander}. Submitted to: Discrete Applied Mathematics. (Origin: 'On the Fault Tolerance of Some Popular Bounded-Degree Networks', T. Leighton, B. Maggs, and R. Sitaraman, Princeton report CS-TR-385-92.)

Accepted but suspicious

'Communication Complexity of Specialized Paralllel Algorithms', C.V. Papadopoulos, B.D. Clarke, and S. Carderinis, Submitted to: 5th Hellenic Informatics Conference, 1995, Greece.

C.V. Papadopoulos and {innocent bystander}, 'Modelling the complexity of parallel and VLSI computations with Boolean circuits', Microprocessors and Microsystems, Feb 1995, vol 19, (no 1):43-50.

C.V. Papadopoulos, 'Stochastic modeling of multiprocessor reliability', Proc. First International Conference on Massively Parallel Computing Systems (MPCS), Ischia, Italy, 2-6 May 1994, pp 620-30.

C.V. Papadopoulos and A.H. el Zahni, 'Protection and routine algorithms for network management: the case of transmission networks', Proc. Nineteenth EUROMICRO Symposium on Microprocessing and Microprogramming (EUROMICRO 93), Barcelona, Spain, 6-9 Sept. 1993). Also Microprocessing & Microprogramming, Sep 1993, vol 38, (no 1-5):163-70.

C.V. Papadopoulos, 'A Protocol Architecture for Eliminating Latency in ATM Networks', Microprocessing and Microprogramming Journal, January 1996.

C.V. Papadopoulos, 'Integrating Formal Methods in Software Engineering', Software-Concepts and Tools Journal, August 1995

Accepted but suspicious

C.V. Papadopoulos, 'Exploitation of IP Datagrams for Internetworking Mobile Computers', Computer Communications Journal, October 1995.

C.V. Papadopoulos, 'Programming Language Support for Distributed Systems with Multicast Communication', Computer Communications Journal, December 1995.

C.V. Papadopoulos, 'Optimizing the Response Time of Multiserver Networks', European Transactions on Telecommunications Journal, April 1995.

C.V. Papadopoulos, 'Statistical Measurements on IP Networks as a Basis for ATM Switch Design', Computer Communications Journal, February 1996.

C.V. Papadopoulos, 'Evaluation of Multigrid Algorithms on Message-passing Multiprocessors', Parallel Algorithms and Applications Journal, September 1995.

C.V. Papadopoulos and R.C. Hofmann, 'Performance Modelling of a Distributed ISDN Protocol Test System', IEEE Proc. First Int. Conf. on Architectures and Algorithms for Parallel Processing (ICA3PP '95), Australia, April 1995.

Another case

- I Researchers at institution A published a technical report in mid-1994.
- A student at institution B heard about the work, and obtained the technical report and a copy of the code through a colleague of the researchers at A.
- The student used the paper as the basis of a thesis, taking headings, table layout, choice of experiments to run, and much of the text.
- The student and his supervisors at B edited the thesis into a paper submission. I The work from A was published in a conference in January 1995, the work from B was submitted in late 1994 and appeared in July 1995.
- The researchers at A later expanded their work and submitted it to a journal. They were accused of plagiarism.
- Following legal advice, institution B refused to agree that there had been wrongdoing, but did concede that the work from A appeared first.

Outcomes of Plagiarism

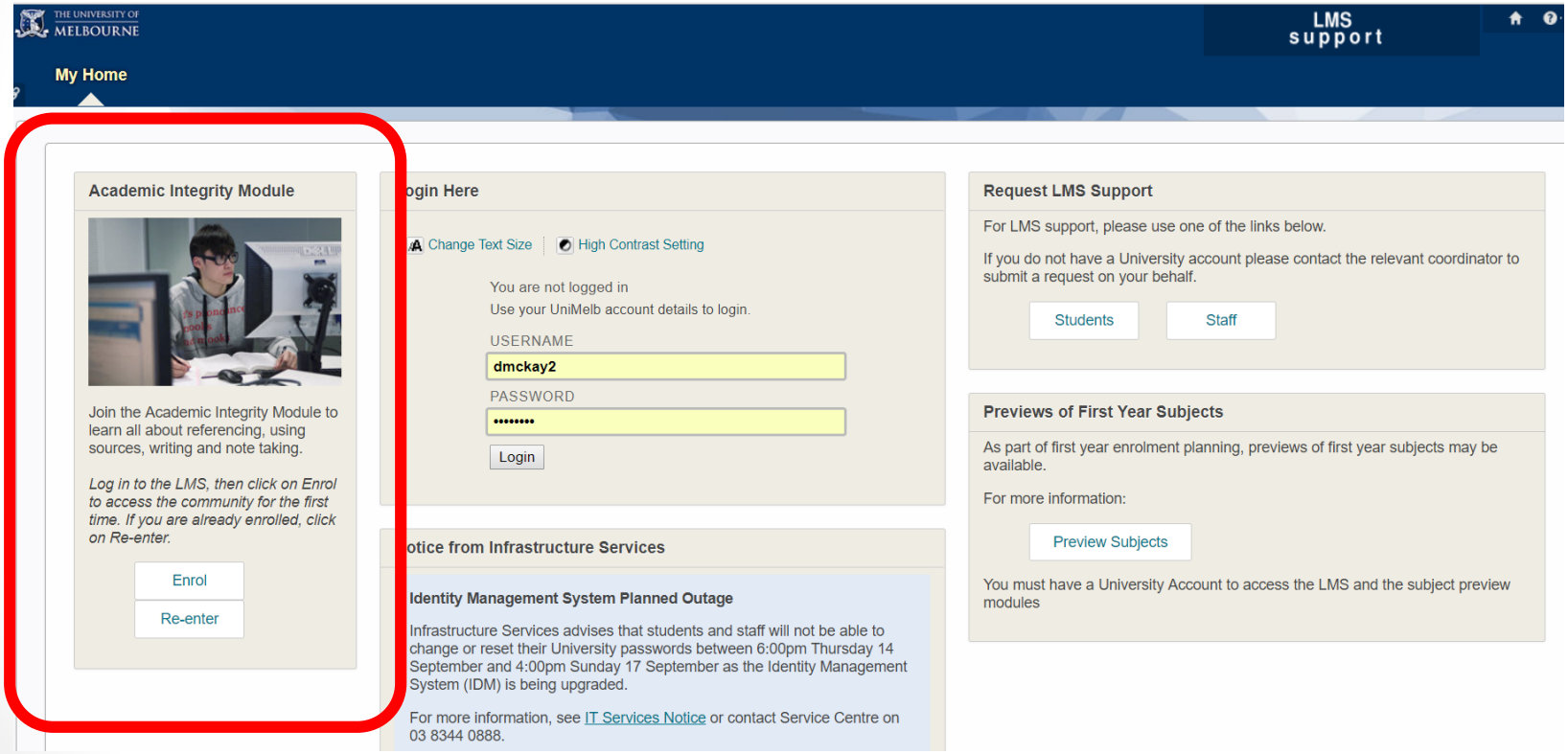
- Failing an assignment or degree
- Loss of collegial relationships
- Damage to reputation
- Job loss (see the Monash case!)
- Being banned from publishing venues
- Being prosecuted for copyright infringement

- You may well get away with it, but it is ethically, morally and often legally wrong

How to avoid plagiarism

- Write fresh every time
- Cite only things you have actually read
- State when ideas or results are prior work
- Put quotes in quotation marks
- Create your own figures
- Attribute fully—make it clear who did or said what
- Use citations liberally and generously

How to avoid plagiarism



Dealing with ethical problems

- Avoid them—if you are unsure, be conservative
- Resolve problems early
- In the first instance speak to your supervisor
- If your supervisor is the problem speak to another trusted academic in confidence
- Justin Zobel cares and will take you seriously
- Do not make public accusations until
  - You are very sure of your facts
  - All other avenues have been exhausted (supervisor, head of department, dean, VC, publishers etc)

Summing up

Disgraced South Korean scientist Hwang Woo-suk speaks to reporters after his trial at the Seoul Central District Court in Seoul on Monday. Hwang, who falsely claimed to have achieved major breakthroughs in stem cell research, was convicted of charges connected to his research, but avoided jail as the court suspended a two-year prison term for him. | Photo Credit: Ahn Young-joon

- Unethical behaviour is bad because it is *wrong*
- Codes of conduct exist to protect accusers and accused
- Proven accusations can have severe consequences
- Unproven allegations can also damage reputations

Image source: http://www.thehindu.com/sci-tech/science/Disgraced-cloning-expert-convicted-in-South-Korea/article16888704.ece