Scientific Writing — What is a Master "Thesis"?

J.N. Oliveira

DIUM

HASLab/INESC TEC

QSE Group

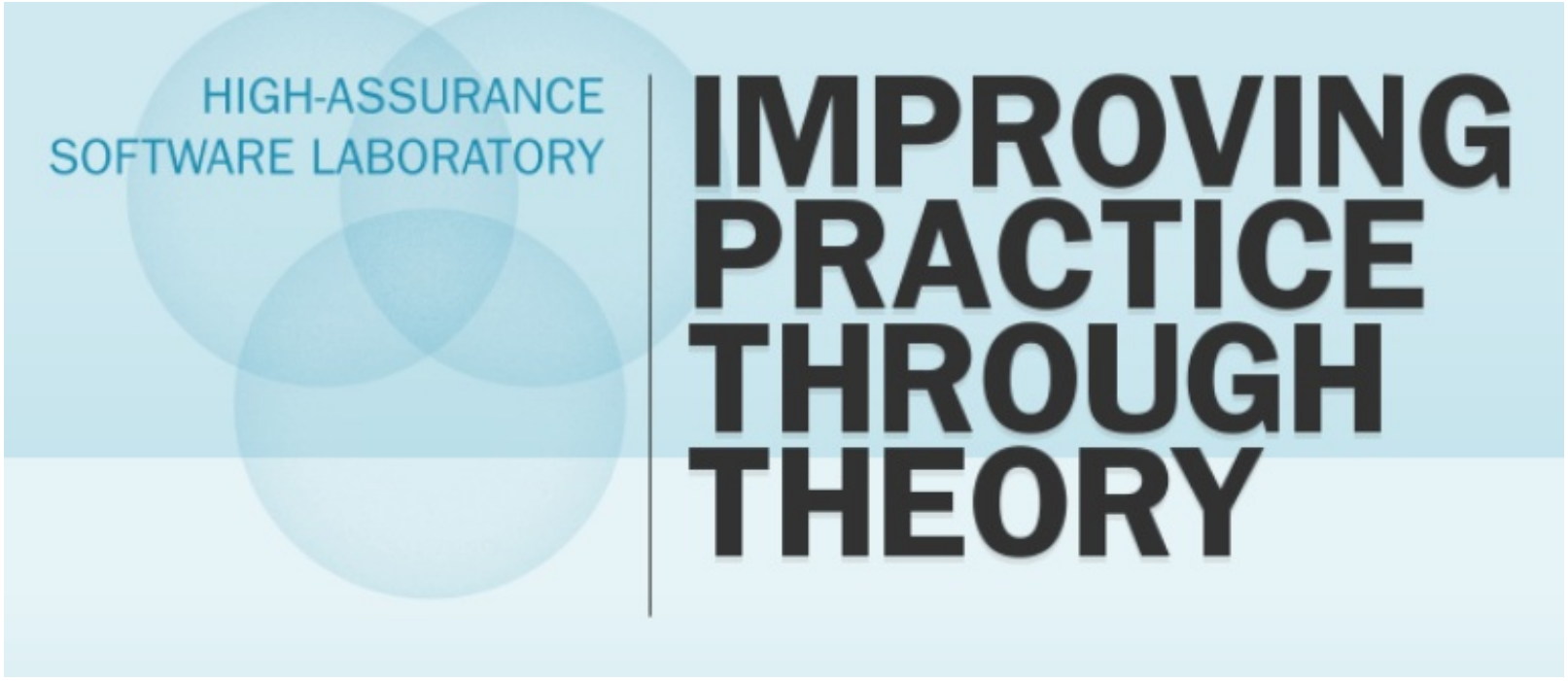
MEF Seminar (Course: EFA)

12th October 2023, Room E1-2.20
U.Minho, Braga, Portugal

Context

S2	Fundamentos de Física de Microssistemas	Fis	5
S2	Micro e Nanofabricação	MNT	5
S2	Sistemas Eletrónicos para Imagem Médica e Radioisótopos	EEIC	5
Ano 2			**60**
A	Dissertação	EFIS	45
S1	Engenharia Física Aplicada	EFIS	10
S1	Formação Empresarial e Empreendedorismo	EIS	5

HASLab motto



What has this to do with "scientific writing"?

Context: Learning cycles

UG Undergraduate study

PG Postgraduate study

Context: Learning cycles

UG

BSc — 1st cycle

PG

MSc — 2nd cycle

PhD — 3rd cycle

Context: Learning cycles

UG

BSc — 1st cycle

PG

MSc — 2nd cycle

PhD — 3rd cycle

Bologna Learning cycles

BSc — **1st cycle**: student expected to learn and apply **general**, well-established theories:

*The "**repeat**" phase*

MSc — **2nd cycle**: student expected to learn **specialized** theories and build solutions from them:

*The "**specialize**" phase*

PhD — **3rd cycle**: students (perhaps they can do better than their former teachers?) expected to pursue a new **conjecture** (thesis) and give scientific evidence of it:

*The "**create**" ("invent") phase*

Some terminology

PG academic **degrees**

MSc from the Latin

magister (teacher)

PhD from the Latin

philosophiæ doctor

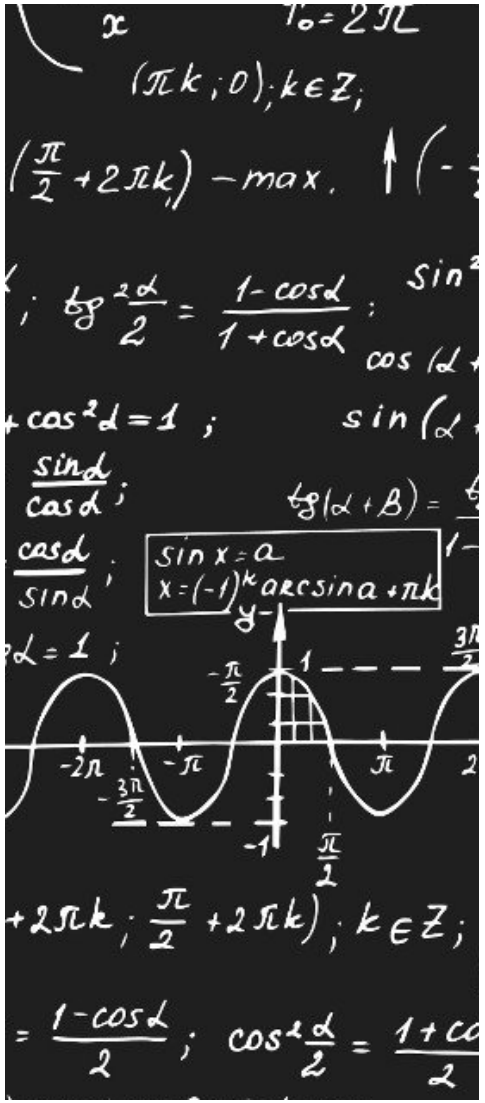
In Greek:
$\delta\iota\delta\alpha\kappa\tau\omega\rho\ \phi\iota\lambda\sigma\sigma\phi\iota\alpha\varsigma$
(*didaktos filosofias*)

Some terminology

PG thesis — a scientific **result** (from the Greek $\theta\varepsilon\sigma\iota\zeta$, *position*)

PG project — a **planned** undertaking

PG dissertation — a piece of **text**, originally a *discourse* (from the Latin *dissertatio* < *disserere*, to discuss).



Dissertation? A bit of history

Pedro Nunes (1502-1578) — perhaps the Portuguese **top scientist** ever.

His PG took place in Lisbon, on the 16th Feb, 1532.

The day before he went to the cathedral to know the topics (*"tirar os pontos"*) which he should address in his *dissertatio*.

According to the rules, his *"exame privado"* took place after *"sol posto"*.

Dissertation? A bit of history

From the minutes:

> *"(...) e emtraram em lugar pera isso aparelhado omde ficaram soos os mestres ou doctores da faculdade, camcellario Rector & scrivam (...)"* [His examiners were] *"mestre filipe e mestre francisco e Joam liam e Antonio Mendez (...) feita uma breve arenga"* [Pedro Nunes addressed the two topics] *"nos quaes pontos steve quasy duas horas pouco maes ou menos"* [Then he went out briefly] *"a tomar um pouco de folgo"* [and returned to answer the questions of his examiners] *"e asy em ler como em resumir e responder se houve tam suficientemente que foy aprovado por todos nemine discrepante (...)"*

(Auctarium III, MCLV, p.114)

Doing a PG — doing "science", ok?

PG **projects** are a standard way of undertaking research.

The root *philosophiæ* in PhD does not mean philosophy as a discipline — it means **depth of knowledge** or thought.

PG programmes range over the

- human (social) sciences
- natural sciences
- exact sciences.

Doing a **PG** — doing "science", ok?

However, what does **"science"** mean?

What tells **science** apart from other forms of human "knowledge"?

PG students should care about these questions!

Science? Pre-science?

In an excellent book on the history of scientific technology,

> *"How Science Was Born in 300BC and Why It Had to Be Reborn" (Springer, 2003),*

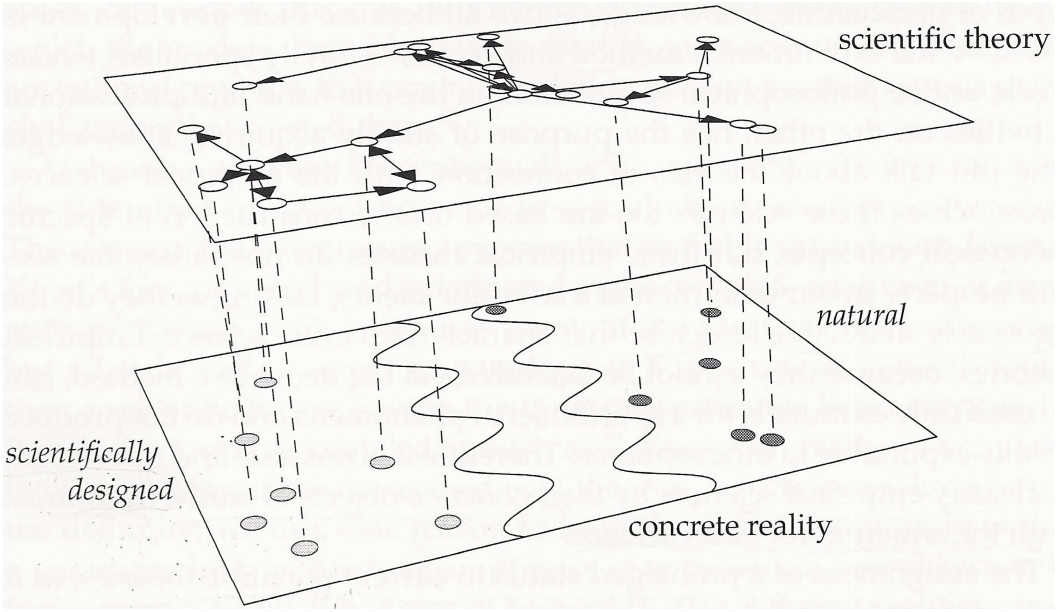
Lucio Russo writes:

> *The immense usefulness of* **exact** *science consists in providing* **models** *of the real world within which there is a guaranteed method for telling* **false** *statements from* **true**. *(...) Such models, of course, allow one to describe and* **predict** *natural phenomena, by translating them to the theoretical level via* **correspondence rules**, *then solving the "***exercises***" thus obtained and translating the solutions obtained back to the real world.*

Disciplines unable to build themselves around *"exercises"* are regarded as **pre-scientific**.

Scientific engineering ($e = m + c$)

Also from Russo's book :



Vertical lines mean **abstraction**, horizontal ones mean **calculation**:

engineering = model first, then calculate

($e = m + c$)

Example

- **Natural phenomena** — planetary motion, gravity, etc.
- **Correspondence rules** — Newton (1642-1727)'s laws of mechanics and gravitation stemming from **model**

$$F = G\frac{mM}{d^2}$$

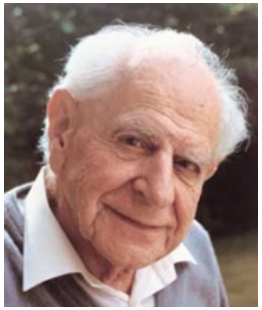
- **"Exercises"** — Earth's gravitational field,

$$g = \frac{GM}{R^2}$$

then $F = gm$, then $F = m\frac{dv}{dt}i = ma$, then... (you know the rest!)

- **Translation back to the real world** — ballistics, space missions, satellite technology, and so on.

Where does it all begin?



K. Popper (1902-94)

Following the eminent philosopher of science of the 20c Karl Popper (1902-94), science does not arise from **observation** or **inductive** perception of reality only.

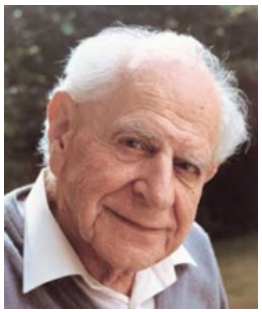
Scientific theories, and human knowledge in general, are conjectural or hypothetical, and are generated by **creative imagination**.

This links **science** with **art**.

It means that æsthetic attributes such as **beautiful**, **elegant**, **horrible**, **ugly**, etc. also apply to science.

Beware: this applies to your **PG** work as well!

Where does it all begin?



K. Popper (1902-94)

Following the eminent philosopher of science of the 20c Karl Popper (1902-94), science does not arise from **observation** or **inductive** perception of reality only.

Scientific theories, and human knowledge in general, are conjectural or hypothetical, and are generated by **creative imagination**.

This links **science** with **art**.

It means that æsthetic attributes such as **beautiful**, **elegant**, **horrible**, **ugly**, etc. also apply to science.

Beware: this applies to your **PG** work as well!

"Beauty is our business"

E.W. Dijkstra (1920-2002)

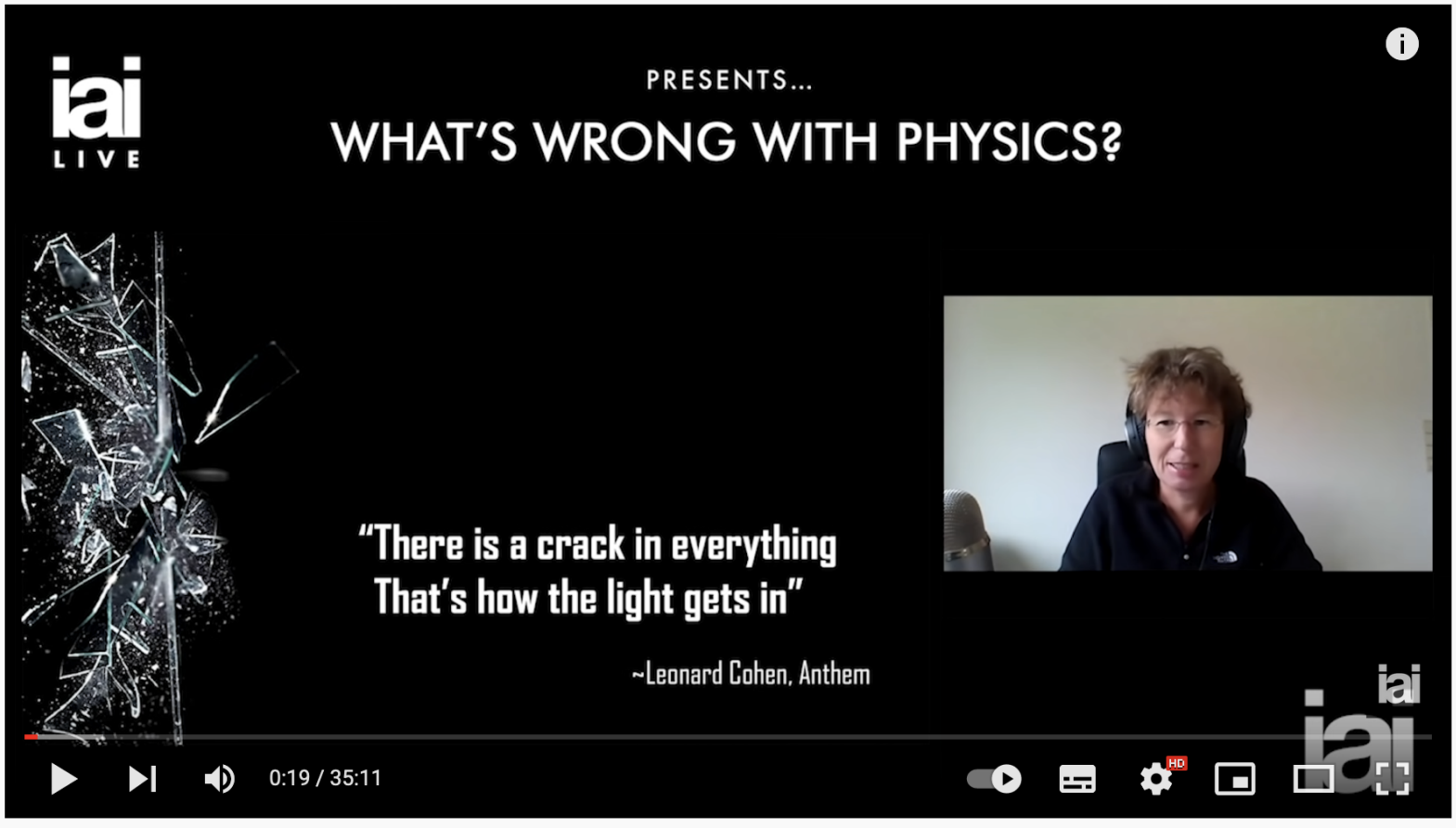
"...when we recognize the battle against **chaos**, **mess**, *and unmastered complexity as one of computing science's major callings, we must admit that '***Beauty Is Our Business***'".*

(E.W. Dijkstra, EWD697)

Still, the questions remain:

- How many follow Dijkstra's advice?
- Are we doing research in the right way?
- Are we using the right notation, language?
- Does more **technology** mean better **science**?

Not everyone seems happy



#SabineHossenfelder #quantummechanics #scientificmethod

What's wrong with physics? | Sabine Hossenfelder

(https://www.youtube.com/watch?v=8aUk6oi_AmM)

Not everyone seems happy

Vinton Cerf (1943-)

ACM President

DOI:10.1145/2347736.2347737

Where is the Science in Computer Science?

"... we have a responsibility to pursue the science in computer science. We must develop better tools and much deeper understanding of the systems we invent and a far greater ability to make predictions about the behavior of these complex, connected, and interacting systems.".

(Vinton G. Cerf, Letter from the ACM President, CACM 55(10), Oct. 2012)

Experimental sciences

Where is the place of research **experiments** in the $e = m + c$ equation?

From the Wikipedia:

> *"In the scientific method, an* **experiment** *is an empirical procedure that arbitrates competing* **models** *or* **hypotheses**.

In Popper's philosophical terms, the fact that you select *some* **experiments** to carry out (instead of others) already is evidence of your **creative** mind ...

Complexity, Complication, Obfuscation

Mind the differences:

- **Complexity** — property of being intricate but with formalizable structure
- **Complication** — messy, lacking structure
- **Obfuscation** — formalization intended for bewilderment rather than enlightening (worst of all).

By definition, a PG project is close to some **frontier of knowledge**. Therefore:

- Do not expect an easy task
- It will be **complex** — so, do not **complicate** it further
- Never dare going into obfuscation!

Be open minded

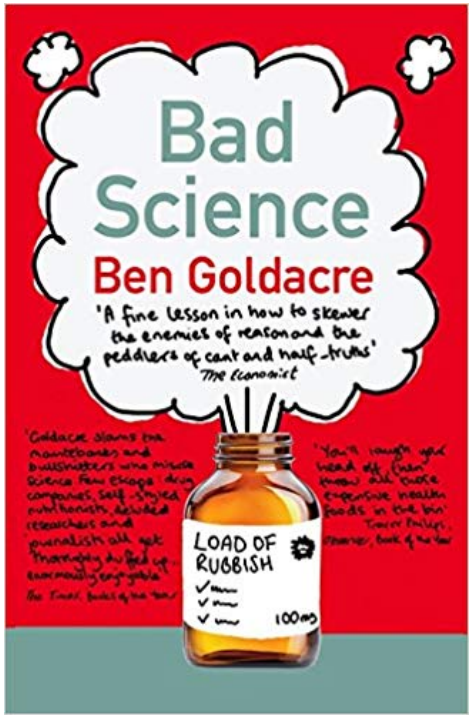
Share your **data** and **results**.

This will add to your own credibility.

Open Data — let others check your results (like the maths people do whenever they publish proofs) – use e.g. GITHUB or anything similar.

Open Data Trend — see e.g. the EU Open Data Portal.

Bad science — be aware of the risks of not doing so by reading the book with the same title by Dr Ben Goldacre.

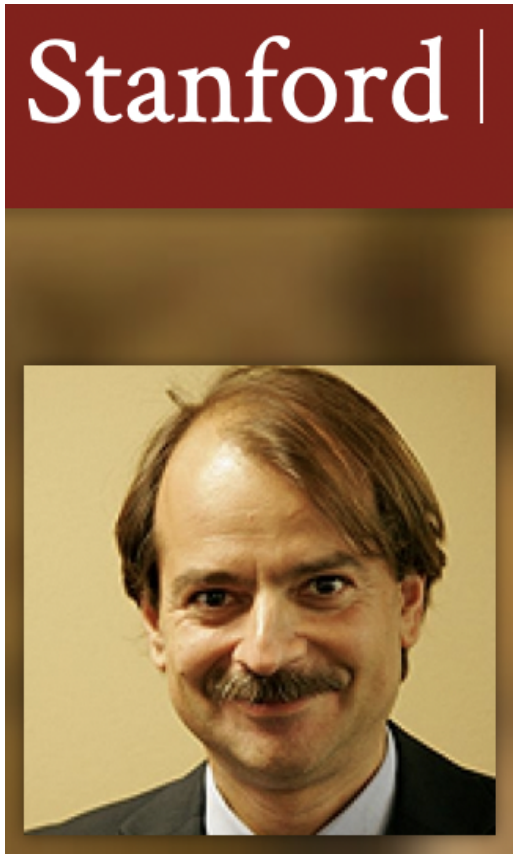


NB: A solid **negative** result is worth more than a shaky **positive** one.

Beware

"Why Most Published Research Findings Are False" (Ioannidis, 2005):

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. (...)



Recent scandals

Stanford president to resign over concerns about integrity of his research

Marc Tessier-Lavigne said he will step down because he expects continued debate about his ability to lead the university

Harvard withdraws papers in dishonesty expert scandal

Research on 'how dishonesty can lead to greater creativity' retracted as other behavioural science findings queried

Beware the side effects

Doing a **PG** research project will change your life forever. (For the **good**, I think!)

Beware of the side effects, e.g.

> **Dunning-Kruger effect** — *(...) competent students tend to underestimate their own competence, and erroneously presume that tasks easy for them are also easy for other people.*

Beware the side effects

This often leads to the so-called *fraud syndrome* among researchers:

> **Fraud syndrome** — *psychological pattern in which individuals doubt their accomplishments and have a persistent internalized fear of being exposed as a "fraud".*

If you feel any of these, don't worry — many **highly successful** people feel the same!

Planning a PG dissertation

Questions

- **How** should I structure it?
- **When** should I start?
- **What** should I write?

Natural questions, aligned with the so-called **Aristotelian categories** ("natural dimensions of things"):

What	*the thing is about*
What for	*the purpose of the thing*
Why	*bother with the thing*
When	*does/did the thing happen?*
Where	*is/was the thing taking place?*
How	*is/was the thing carried out?*

Questions

- **How** should I structure it?
- **When** should I start?
- **What** should I write?

Natural questions, aligned with the so-called **Aristotelian categories** ("natural dimensions of things"):

What	*the thing is about*
What for	*the purpose of the thing*
Why	*bother with the thing*
When	*does/did the thing happen?*
Where	*is/was the thing taking place?*
How	*is/was the thing carried out?*

What is it?

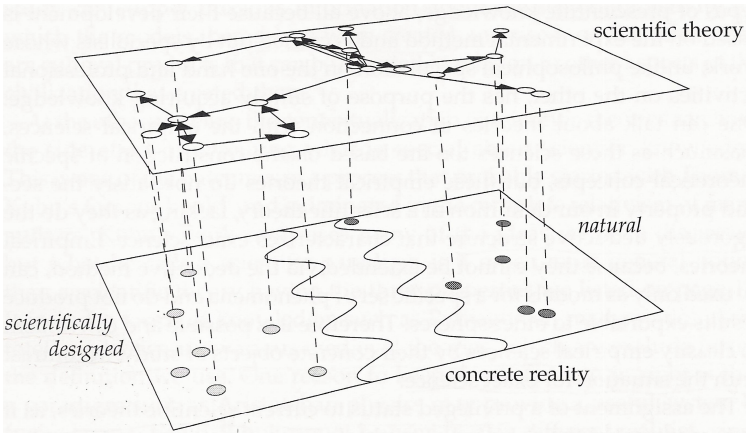
Recall that:

- A **PG** dissertation is a **document** which should provide scientific **evidence** of some **novel** result(s) in some area of knowledge.
- Following the **scientific method**, the concepts involved in such results should be **formalized** first (vertical arrows in Russo's diagram) and then **reasoned** about (horizontal arrows in the same diagram).

What is it?

This entails some structure in the text:

- **Definitions** for each correspondence rule (in Russo's sense)
- **Theorems** for each "exercise" (in Russo's sense).



What about the overall text?

How should I structure it?

Recall the typical structure of a mathematical argument, leading to results in the form of **theorems**, each involving:

1. **Thesis** (T)
2. **Hypothesis** (H)
3. **Proof** ($H \Rightarrow T$)
4. **Corollaries**
5. **Lemmas**
6. Others' theorems

How should I structure it?

Since the purpose of a **PG** dissertation is that of providing scientific evidences, its **overall structure** mirrors the shape of a **mathematical argument**. Here it goes:

Maths	**PG** project	**Dissertation**
Thesis (T)	Main result	Contribution chapter
Hypothesis (H)	Context	State of the art [1]
Proof ($H \Rightarrow T$)	Evidence	Core chapters
Corollaries	Application	Case studies
Lemmas	Support results	Appendices
Others' theorems	Previous evidence	Bibliography

So, in a sense, writing up your **dissertation** means *proving "your theorem"* — your **thesis**, recall the terminology.

[1] Inc. previous work.

How should I structure it?

Therefore, it is no wonder that a **PG** **dissertation** should be structured something like [2]:

- Introductory material:
  - 1st Chapter — Context, motivation, main aims
  - 2nd Chapter — State of the art review; related work
  - 3rd Chapter — The problem and its challenges

- Core of the dissertation:
  - 4th Chapter — Main result(s) and their scientific evidence
  - 5th Chapter — Application of main result (examples and case studies)
  - 6th Chapter — Conclusions and future work

[2] Number of chapters not strict: it may vary according to the needs.

How should I structure it?

- Auxiliary material:

  Bibliography — List of works referred to in the main text

  Appendix A — Support work (auxiliary results which are not main-stream)

  Appendix B — Proofs of some results (lengthy, technical proofs arguments which would compromise readability of main text)

  Appendix C — Listings (should this be the case)

  Appendix D — Tools (should this be the case)

This should be complemented by some extra matter, as in the following slide.

How should I structure it?

1. Front matter:

   Title page — institutional, as a rule
   Abstract page — summary of the work
   Acknowledgements — thanks to tutors, colleagues, institutions (funding), etc
   Glossary — list of **acronyms** and their meaning
   Lists — of tables, of figures etc (automatically generated if using a proper **authoring system**)

2. Rear matter:

   Index of terms — index of mentioned **entities**, with references to where (page numbers) they are mentioned in the text.

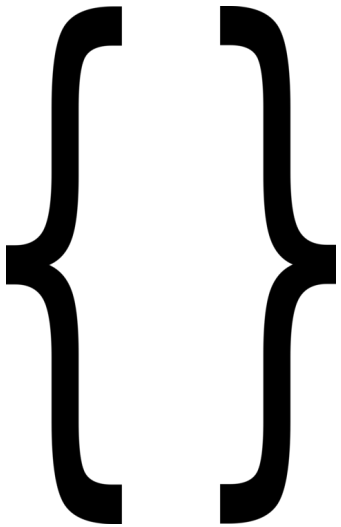
Last but not least

- Do not **nest** your dissertation too much (Dewey Decimal Classification may work against you if you do so).
- A chapter is not a section (**length**!)
- Each chapter can be regarded as a *mini-dissertation* (thus it shares, in a sense, the same structure — introduction, summary at the end [3], etc).
- Do not forget to **spell check** the whole document!

[3] Introduction chapter excluded, whose summary should be an overview of the **overall structure** of the dissertation.

Last but not least — style

- **Symmetry** — introduction and conclusions should be "*matching* parentheses" (check at the end).
- **Aesthetics** — style, elegance and design alone are not enough, but help a lot.



Last but not least — style

STYLE
ORGANIC GROWTH

Text fluidity, no gaps, no abrupt steps...

Last but not least — style

STYLE
ORGANIC GROWTH

(See this video YouTube for more, if you are interested.)

Writing up

When should I write it?

- How about starting writing-up your dissertation on the **very first day** you start your project?
- Of course, this assumes you've understood your **project theme** sufficiently well 
- By then only the **skeleton** of the dissertation can be written — but already following the standard chapter structure.
- Use this sketch as a **road map** and **diary** — you can always keep auxiliary information in the form of **comments**.
- Comments may even include **time stamps** — these will tell how fast you've done your work (useful in measuring effort and **productivity**).

Whom should I write it for?

To **everybody** (!) — ... I mean:

- Introductory and conclusive matter should be written in a style easy to understand by **non-specialists**.
- Core chapters will inevitably be technical, so they are bound to be written for the **specialist**.

Final check up — the question is:

> *Do I master my domain of knowledge upon completion of my project?*

Well...

- ...you should be able to **explain** what you did to **anyone** you may meet on the street (train your **abstraction** skills!)

Abstraction

Quoting Jeff Kramer [4]:

Abstraction *is widely used in other disciplines such as* **art** *and* **music**. *For instance (...) Henri Matisse manages to clearly represent the* **essence** *of his subject, a naked woman, using only simple lines or cutouts. His representation* **removes** *all detail yet* **conveys** *much.*

[4] Commun. ACM, 50:4, pages 37–42, April 2007.

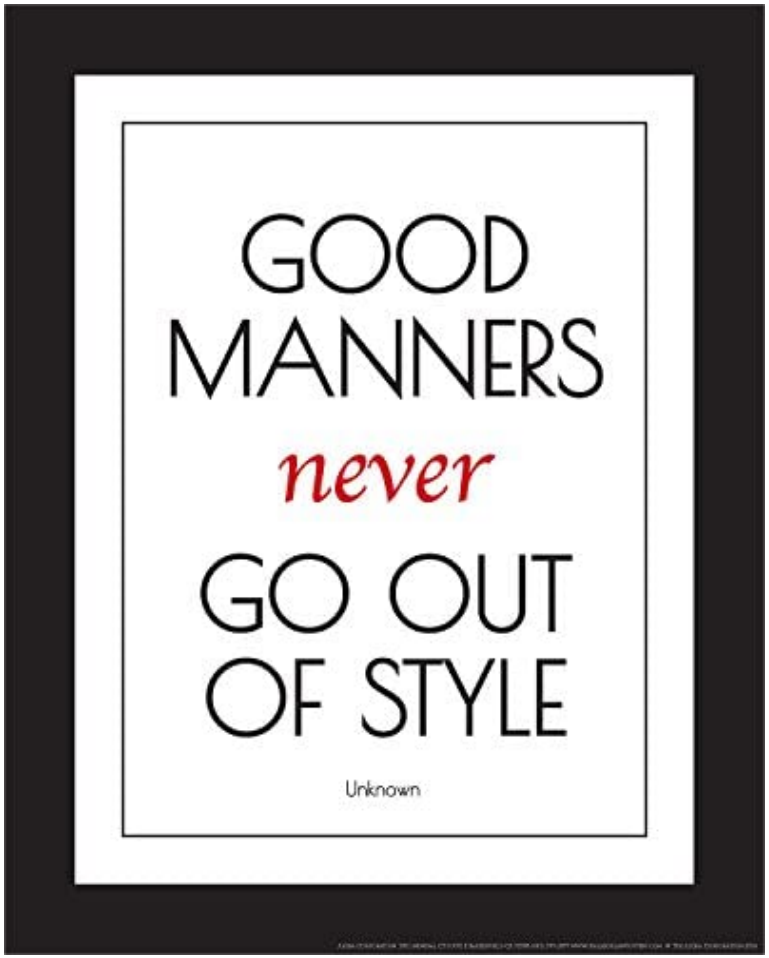
Abstraction

The famous "abstract map" of London's Underground (1939):

How should I write it?

The two sides of this question:

- **Style** (text quality, etc)
- **Production** (editing and publishing).



How should I write it?

Style:

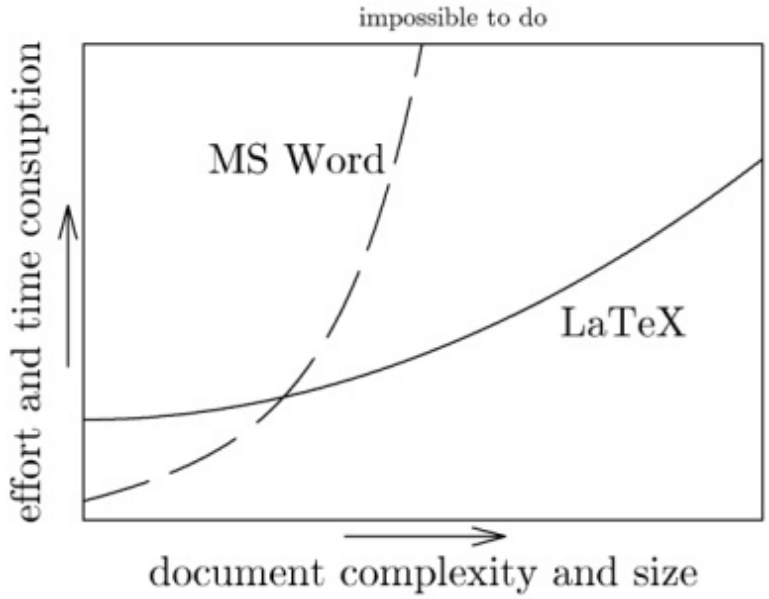
- Never use **colloquialisms**, e.g. contractions: *it's*, *let's*...
- Avoid any form of **majestic** style (*"we"*, *"our"*, ...) — be **modest**.
- Avoid **past tenses** (scientific writing is *not* **story telling**!).
- Text often "comes in pairs":
  - Backward integrity — **declaration** always before **use** (e.g. definition before application).
  - Forward integrity — make sure you **fulfill** whatever you **promise**.

  Cf. offer / demand , production / consumption, etc.

How should I write it

Production — use a proper **text authoring system**. By **proper** I mean one that:

- Handles **references** and maintains **referential integrity**.
- Automates **routine tasks** such as numbering, bibliography, generation of lists and indices.
- Integrates well with **other tools**.



One such system is the **Knuth-Lamport's LaTeX**'s text preparation system (Goossens et al., 1997).

(Maybe you know of others).

How do I write it?

Handling **references** (name spaces):

- Concepts, entities etc have a **name** (reference) and often a type.
- Textual information (implicitly) contains a set of **name spaces**.
- A name in each name space identifies a unique object — it is a **reference**.
- Name spaces call for **referential integrity**.
- Most of these are ensured by the text authoring system itself — e.g. names (numbers) of **figures**, **tables**, **sections**, **theorems**, etc.
- One should be very careful about handling any other references (names).

How do I write it?

For those not handled, here is how I have dealt with them (for LaTeX users only — sorry!): for each **entity**, e.g.

- Entity: University of Minho
- Acronym: UM

define (under package `hyperref`) its reference name:

```
\newcommand{\uminho}[1]{
    \href{http://www.uminho.pt}{#1}
    \index{UM!University of Minho}}
```

Every time you write e.g. `\uminho{the university}`,

- a link to the **website** of the mentioned entity is included
- an entry is added to the **index of terms**, meaning that the occurrence of term `uminho` in the **current page** is recorded.

How do I write it?

If needed, an **acronym** (short-cut) can be defined:

```
\newcommand{\UM}{\uminho{\textsc{u.m.}}}
```

So, every time you use acronym \UM, LaTeX typesets U.M. and does the same as above concerning **hyperlinking** and **index** management.

This will save you from referring to entities that are not in the list of terms.

How do I write it?

Last but not least:

- Keep your dissertation in a document version control system like e.g. **GitHub**, **SVN**, etc

In case of using LaTeX, **Overleaf** is a particular good choice.

Dissertation templates for  are available from `https://web.di.uminho.pt/sitedi/latex/`.

Interfacing with others' work

Last but not least, we need to be concerned with **bibliography** management:

- Nobody doing relevant research is alone.
- Research is actually a **social** activity, with continued interaction in the form of meetings, conferences, and so on.
- Giving **credit** to the others' **contributions** is a very important rule of the game.
- With the information resources of today, managing this may be hard (**too much data**!) without a proper infra-structure.
- This should take the form of a **bibliography database**.

Interfacing with others' work

Last but not least, we need to be concerned with **bibliography** management:

- Nobody doing relevant research is alone.
- Research is actually a **social** activity, with continued interaction in the form of meetings, conferences, and so on.
- Giving **credit** to the others' **contributions** is a very important rule of the game.
- With the information resources of today, managing this may be hard (**too much data**!) without a proper infra-structure.
- This should take the form of a **bibliography database**.

Interfacing with others' work

- Systems around BibTEX provide very easy management of bibliography data
- A bibtex record is like a database record, eg:

```
@book{GRM97
   , title     = {The LaTeX Graphics Companion}
   , author    = {Michel Goossens and
                  Sebastian Rahtz and Frank Mittelbach}
   , publisher = {Addison-Wesley}
   , year      = {1997}
   , note      = {ISBN 0-201-85469-4}
}
```

- You may add your own attributes (which do not get printed) like the ID of this book in your own library, bibliometric stuff, and so on.

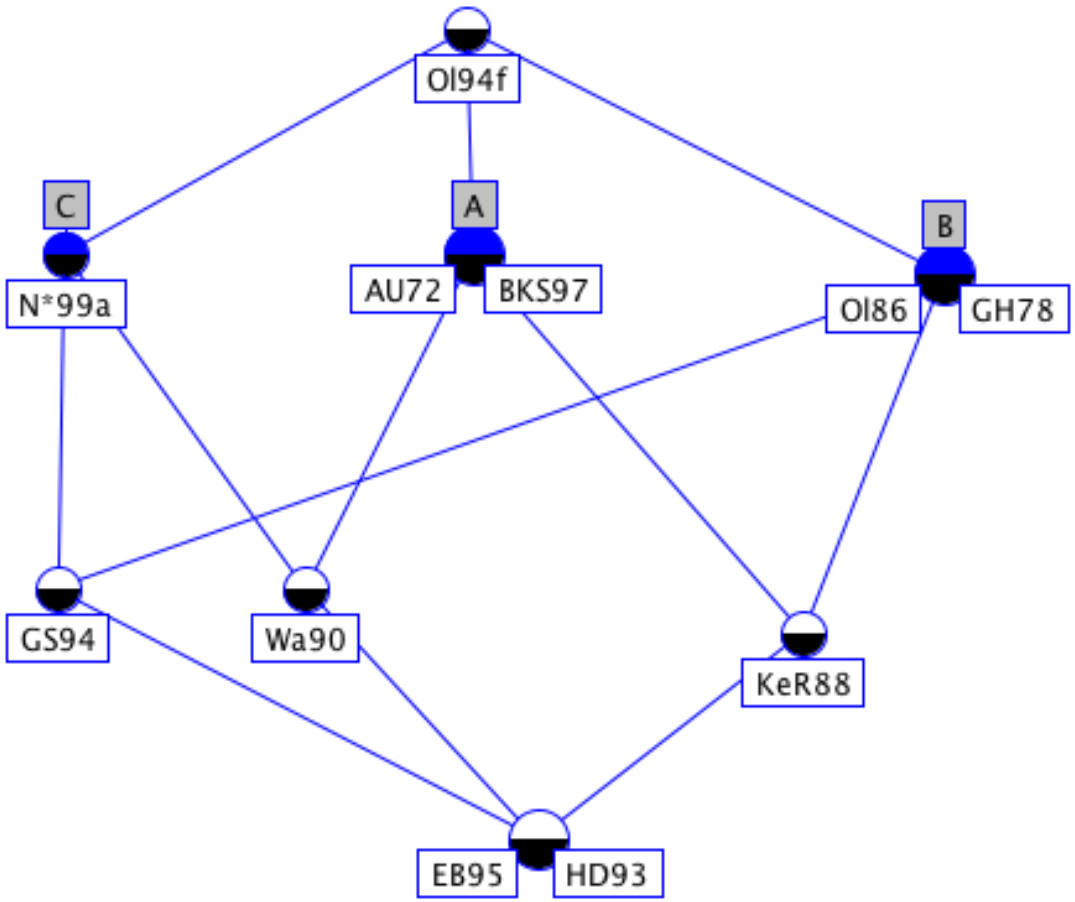
Interfacing with others' work

Classifying your bibliography:

- In particular, you may add a BibTEX attribute named `keywords` to each record of interest
- This will **classify** your records according to keywords relevant to your research
- You many even use the technique of **formal concept analysis** (FCA) developed by Ganter and Wille (1999) to structure your data in a **lattice** of concepts
- Some FCA systems (such as ConExp) offer you a user interface to manage and display your concept lattice (next slide).

Interfacing with others' work

Example concept lattice (11 records, three attributes A, B abd C):



Defense

Your "**day-D**" will eventually happen! How to prepare for it?

- Again, start as early as possible
- **Rehearse** and try to **explain** your work to others
- **Present** your research in workshops, conferences (organize your own locally)
- Get used to **defend** your ideas and beliefs
- Do some **teaching** to train your **speech** skills
- If you are used to **public performance** (e.g. music, theater), don't worry too much...

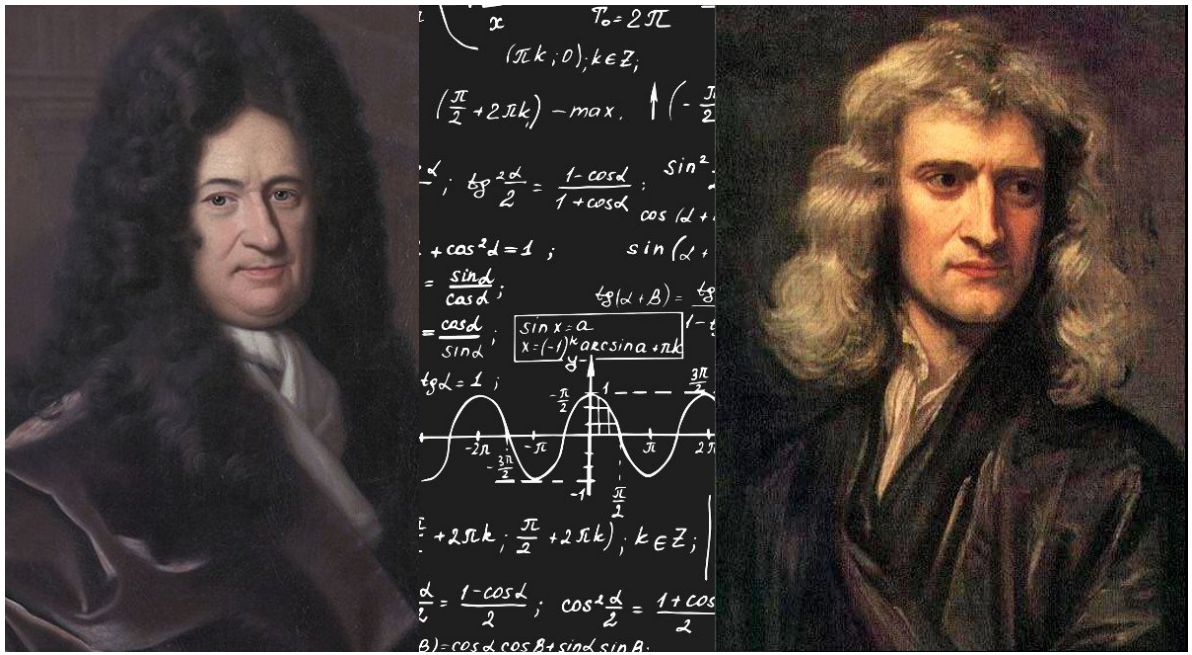
Closing

Final suggestions:

- **Interact** with other researchers in your field.
- Once you have something to show, build a **research blog**.
- Try and **publish** your work in good conferences — the best way to validate your contributions.
- Good **papers** convert to good chapters in the dissertation.
- Offer your **services** in **conferences** in your area (e.g. student volunteering).

Closing

- Be concerned with the **history** of your research field

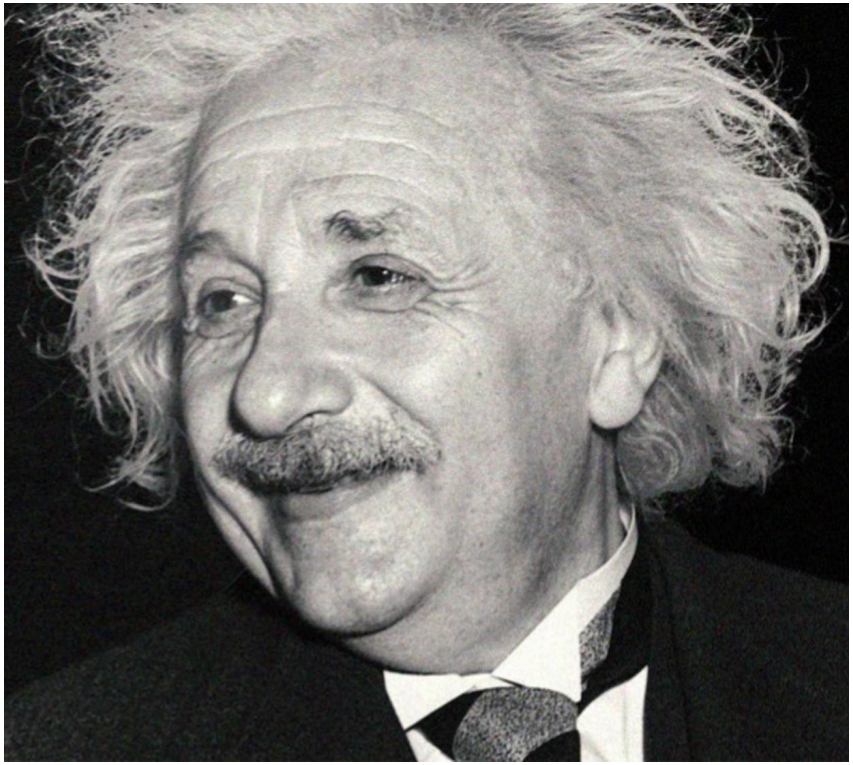


- Makes you more confident to learn from your errors
- Lets you better evaluate “modernity”.

Closing

And do not forget

- to be **creative** (recall K. Popper)
- to have **fun**: if you don't get excited with your project... who will?



Closing

Finally, beware — **bibliometrics** is not everything...

The "publish or perish" syndrome is doing much harm to true science nowadays...

(Check the **DORA** Declaration.)

Peter Higgs: I wouldn't be productive enough for today's academic system

Physicist doubts work like Higgs boson identification achievable now as academics are expected to 'keep churning out papers'

Some links

- *BibSonomy* (a system for sharing bookmarks and lists of literature) — www.bibsonomy.org
- *Mendeley* (a system for organizing research and bibliography, collaborating in project teams etc) — www.mendeley.com
- *DBLP Computer Science Bibliography* (comprehensive account of BibTEXrecords) — www.informatik.uni-trier.de/~ ley/db/index.html

Some more links

The number of sites, blogs, articles about *"doing a PhD"* has increased a lot in recent years.

Some suggestions:

- *How to get a PhD* — http://www.amazon.com/How-Get-PhD-Estelle-Phillips/dp/033520550X — a 236 page (!) book on the subject by Estelle Phillips and Derek Pugh.
- *Writing and Presenting Your Thesis or Dissertation* — www.learnerassociates.net/dissthes/
- *How to Write a PhD Thesis* — www.phys.unsw.edu.au/˜ jw/thesis.html

Some more links

- *How to write a great research paper* — in http://research.microsoft.com/en-us/um/people/simonpj/papers — excellent guidelines by Simon Peyton Jones, Microsoft Research: just replace 'research paper' by 'dissertation' while reading
- *Small guide to making nice tables* — www.inf.ethz.ch/personal/markusp/teaching/guides/guide-tables.pdf

References

Bernhard Ganter and Rudolph Wille. *Formal concept analysis: Mathematical foundations*. Springer-Verlag, Berlin-Heidelberg, 1999.

Michel Goossens, Sebastian Rahtz, and Frank Mittelbach. *The LaTeX Graphics Companion*. Addison-Wesley, 1997. ISBN 0-201-85469-4.

J.P.A. Ioannidis. Why most published research findings are false, 2005. PLoS Med 2(8): e124. `https://doi.org/10.1371/journal.pmed.0020124`.

L. Russo. *The Forgotten Revolution: How Science Was Born in 300BC and Why It Had to Be Reborn*. Springer-Verlag, September 2003. URL `http://www.springer.com/978-3-540-20396-4`.