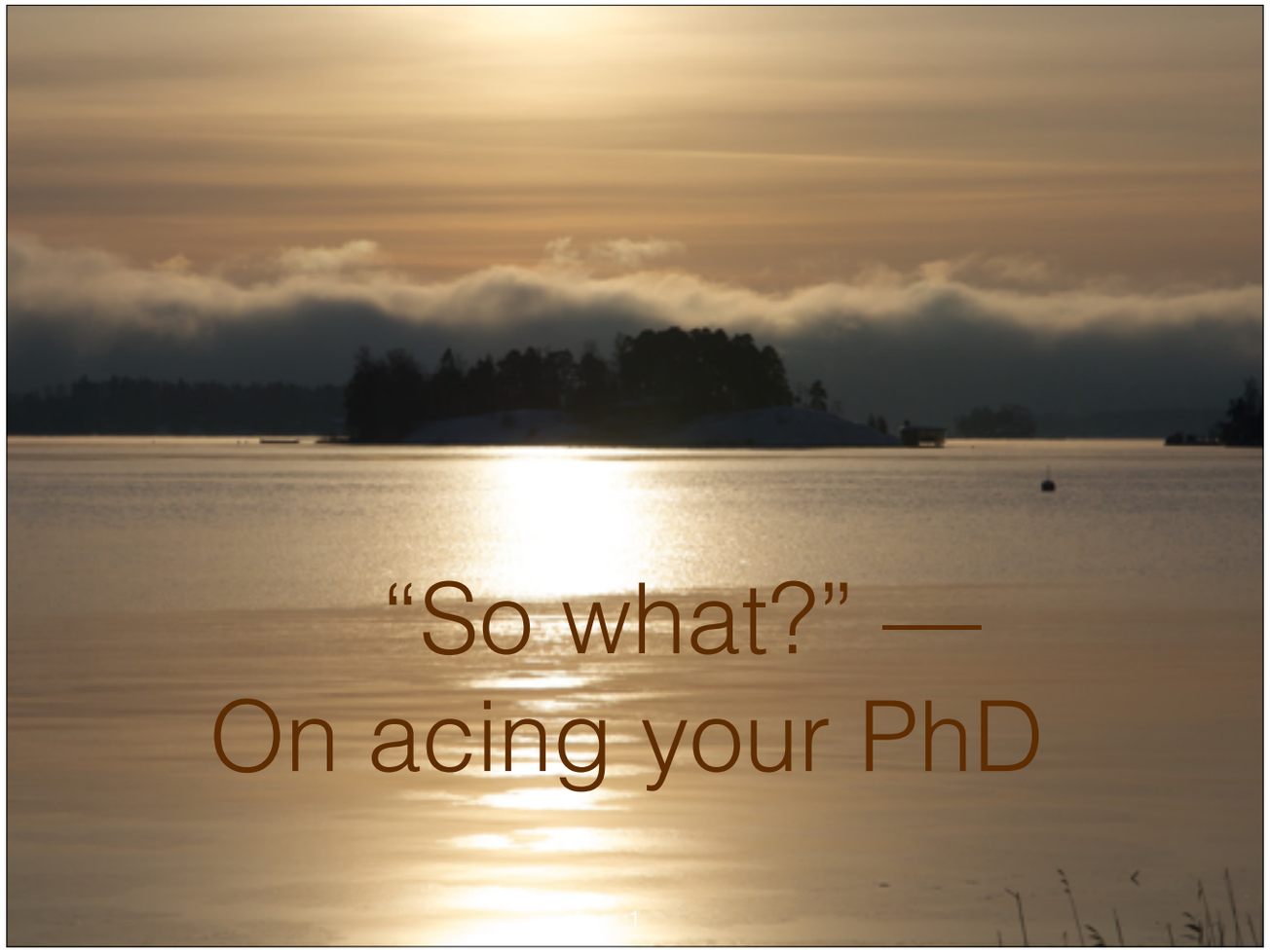

- PhD at Aalto, Finland
- Great place to do your PhD; remote, isolated, lots of space to roam around thinking, but also very highly educated nice people.
- Also good reasons to work on your PhD instead of other things, because although the country is stunningly beautiful, there is also this:

This was taken about 30 minutes later, and shows much better that it was actually -15 ºC outside.

Okay, enough silliness; this talk will be surprisingly serious. I hope you will all find at least 1 useful piece of advice in this talk.

About me

- ’13–now: Research Associate, University of Bristol
  - Formalising subjective interestingness in exploratory data mining (FORSIED, ERC Tijl De Bie)
- ’09–’13: Dr.Sc.(Tech), Aalto University, Espoo
- ’08–’09: Consultant, Crystalloids, Amsterdam
- ’01–’08: MSc, Utrecht University
  - Thesis at Philips Research Eindhoven

Of course you are wondering; who the hell is this guy, so I feel it necessary to present some credentials.

As you can see, I did not really do my studies very efficiently. Especially my BSc degree was bumpy; I took a year off, there were some unpleasant events in my life, I also get bored very quickly, etc.

During my studies, I never expected to be interested in doing a phd, but I learned doing research really fits me well when doing a research project during my masters, and then I also became an intern at Philips Research where I wrote my MSc thesis.

While looking for a nice place to do my phd, I worked as a consultant in data mining. This was really useful actually, because it made me even more determined that the freedom and independence that you have as a phd student would fit me well.

Finally, in 2013 I graduated, about 4.5 years after starting and my degree was graded *with distinction* and I received an award for the best thesis of the school that year.

Currently doing maximum entropy modelling for subjective interestingness in exploratory data analysis.

Contents of this talk

1. Why do a PhD?
2. How to gain interesting knowledge & skills
3. How to become independent
4. How to do good research
5. How to keep going
6. [How to finish your PhD]

Doing a phd, even though it was somewhat successful, does not make me an expert to give advice to other students. So I thought about this for a while and read lots of material on the internet and some papers. I want to touch upon the following subjects because they are relevant and I have something to say about them.

Some will be longer than others, and the last I found actually difficult to give general advice about, but I include it anyway because for many students it is the biggest hurdle.

Contents of this talk

1. **Why do a PhD?**
2. How to gain interesting knowledge & skills
3. How to become independent
4. How to do good research
5. How to keep going
6. [How to finish your PhD]

Then, the first topic. Of course almost all of you have already decided to do a PhD, otherwise you would not be here. However, I will refer back to this frequently later in this talk.

Why I did a PhD

- Learn new things
- Possibility to do something that matters
- *Independence*

Extremely curious person, I want to everything about everything

- Your studies, especially if you organise them well, are a unique opportunity in your life to learn new things. In no other job will you have so much time for educating yourself.
- The second, well… I haven't really achieved that to the extent that I would like. However, at least there will be minimal interference from obligations and other people to making this happen. In any case, science is typically additive and especially in our field, the results are very general and have a potential to impact very many people around the world.
- Finally, you can decide or at least influence to a great extent what you spend your time on, and you will learn to become independent and steer yourself. For me, this is the primary motivation why I am in academia, and I see independence also as heavily related to the other reasons why academia is a great place to work in.

Why to do a PhD (or not)

- Gain deep subject expertise
- Advance the world
- Enjoy freedom, opportunity
- Develop many skills
- Travel, meet many people

- Excessive working hours
- Stress
- Slow pace of achievements
- Everything is *your* problem
- No business experience
- Low income

This is a very brief summary why you could like or dislike PhD studies.

- Explicitly not listed “opening doors”, because other experiences also open the same or different doors.

Contents of this talk

1. Why do a PhD?
2. **How to gain interesting knowledge & skills**
3. How to become independent
4. How to do good research
5. How to keep going
6. [How to finish your PhD]

Now making this more practical:

How to gain interesting knowledge & skills

- Consider what you want to do *after* your PhD
- Consider what skills & knowledge to develop
- Choose research accordingly
- A PhD is an education, for *you* the learning is as important as the research output

The fact that I actually spent time on this early in my PhD is due to Kai Puolamäki, who was taking care of our research group after Heikki Mannila (my advisor) became vice-president of Aalto University.

I think it really is possible to choose research according to the skills you want to develop, and if you are more motivated because you can do things that you would like to do, you are probably also more productive.

How to gain interesting knowledge & skills *in practice*

- Talk about this with your advisor before you start
- Regularly review this, e.g., yearly
  - What you want to do *after* your PhD
  - What skills & knowledge to develop
  - Choose research accordingly

This is also a good test to see how much your advisor is interested in your development. However, I would expect virtually all advisors to be positive and also have a clear opinion on knowledge and skills that would be helpful to you.

I think many students have no clear plan what they want to do after their PhD. This is perhaps fine at the start of your PhD. However, when you are one or two years in your PhD, you should have a clear view on this, because it should be determining how you develop yourself during your PhD.

How to gain interesting knowledge & skills *in practice*

- Read, read, read
  - Research articles
  - Texts about skills (e.g., writing)
- Take courses on topics and to develop skills
- Attend local talks
- Participate/organise reading groups

In practice, knowledge is gained primarily by reading. You should read a lot, maybe some textbooks about topics that are not the focus of your research, but mainly original scientific articles. Also, read texts about writing style, how to write productively, and other skills that you would like to develop. Secondly, take courses, for example at your department, or online.

Apart from practice, courses are the easiest way to develop soft skills such as networking and communication skills.

Then, to broaden your view, attend local talks. I attended almost all 30+ talks at HIIT during my first year.

Finally, if you want to learn about a new topic that no-one in your university is teaching, you can find likeminded colleagues and do a reading group. They can be great fun and are typically a very efficient manner of gaining new knowledge.

Contents of this talk

1. Why do a PhD?
2. How to gain interesting knowledge & skills
3. **How to become independent**
4. How to do good research
5. How to keep going
6. [How to finish your PhD]

How to become independent

- Independence is a great asset
- Independence is often a necessity (to graduate)
  - Your advisor is not available 24/7
- Your PhD is your problem
  - This is true anyway
  - You will achieve more if you take responsibility

Independence is perhaps the most important distinguishing factor why employer's hire PhDs; the degree tells them that you are an independent problem-solver. In part this is a selection process, because independence is often a necessity to do successful research and graduate. However, this can definitely also be learned to some extent.

I think it is very useful to realise that your PhD is primarily your problem. Nobody will hold your hand and in the end you will have to make sure all problems that you encounter are solved. If you rely heavily on your advisor, then it will take a very long time, or you will never finish at all, because he is not available all day and night.

How to become independent

- You should advance your research and educate yourself
- However, independent ≠ solitary
- If you are stuck or need help, find it
  - There is no shame in this, no one is perfect
  - In fact, scientists are typically not very knowledgeable about science outside their focus area
- Your advisor should be available, if not, demand it

Take control and make sure yourself that your research progresses and also make sure that you educate yourself with relevant knowledge and skills.

With independence, I do not mean at all that you should be doing everything alone. On the contrary, I mean that you should be leveraging your environment as much as possible. One particular point that I would like to stress is that advisors can be busy and if you are doing well they may think that you need no help. However, if you think you need their attention, ask it or even demand it.

Contents of this talk

1. Why do a PhD?
2. How to gain interesting knowledge & skills
3. How to become independent
4. **How to do good research**
5. How to keep going
6. [How to finish your PhD]

How to do good research

- Develop a large knowledge base and diverse set of skills
- Engage with others
- Take control
- Leverage your advisor's skills and knowledge
  - If anything, he is probably great at research

In the Netherlands education is more and more focused on soft skills; debating, writing essays, etc. I see very often that people think they have great ideas and can contribute to society and research and that they have the skills to do so. However, you cannot be creative without knowing a lot as well. In technical fields, it is also very important to have a diverse set of technical skills, in ML those skills are things like algebra, proving techniques, programming, etc.

Also, engage with others to discuss your research as much as possible. The next few slides will discuss this in more detail.

The other points I have mentioned before; take control of your PhD, and leverage your advisor's opinion for example to prioritise your research. You will encounter many opportunities and have far too many (good) ideas. However, which to execute is often not an easy choice. Your advisor was probably very successful in the past to make such choices wisely, so use his experience.

About my dissertation

I worked on an interdisciplinary project. Heikki Mannila, in the middle, was my advisor, Panagiotis Papapetrou the postdoc in our group, and Kai Puoalamaki on the far right was leading our research group after Heikki became vice-president of Aalto University (one month after I started my PhD).

I worked a lot with Terttu Nevalainen, Tanja Säily, and Turo Vartiainen who are doing research in historical socio-linguistics using corpora, mainly English.

How I did good research

- Read, read, read, talk, talk, talk
- Spend a lot of time with end-users
  - Finding problems interesting to both of us
    - Actually very difficult, “perhaps more suitable for tenured staff” [Heikki Mannila, in response to a question at the graduate school meeting]
- Massive effort to explain methods to lay-people

Essentially, I spent a lot of time reading other people’s work and talking to others. Sitting at your desk all day programming is not a way to ace your PhD.

Particularly, my research was motivated by problems that the linguistics encountered in their research. However, many of their problems could be solved by things that would not be publishable in data mining, because they were largely solved problems. On the other hand, what would be perhaps theoretically interesting in DM did not have any real application in linguistics. So we spent a lot of time finding problems that were interesting to the both of us. I probably spent about 20% of time at their department.

In any case, explaining the methods that you develop to non-experts is a great way of synthesising your thoughts around it.

How I did good research

- Present and ask for feedback

Conference Presentations and Posters

- Tanja Säily, Terttu Nevalainen, Jefrey Lijffijt. **Tracing significant change in 17th-century English lexis: the civil war effect.** HPSCG 15, 23 - 25 August, Helsinki, Finland, 2012. (Poster, Handouts)
- Jefrey Lijffijt, Tanja Säily, Terttu Nevalainen. **Chi-square test considered harmful: Better methods for testing the significance of word frequencies.** ICAME 33, 30 May - 3 June, Leuven, Belgium, 2012. (Presentation, Implementation)
- Panagiotis Papapetrou, Jefrey Lijffijt, Tanja Säily, Kai Puolamäki, Terttu Nevalainen, Heikki Mannila. **Are you talking Bernoulli to me? Comparing methods of assessing word frequencies.** Helsinki Corpus Festival, 28 Sep - 2 Oct, Helsinki, Finland, 2011. (Presentation)
- Turo Vartiainen, Jefrey Lijffijt. **Can articles predict the word class of the premodifier? A study of the -ing participle.** ICAME 32, 1 - 5 June, Oslo, Norway, 2011.
- Turo Vartiainen, Jefrey Lijffijt. **Premodifying -ing participles in the parsed BNC.** ICAME 31, 26 - 30 May, Giessen, Germany, 2010. (Presentation)
- Jefrey Lijffijt, Harri Siirtola, Tanja Säily, Turo Vartiainen, Terttu Nevalainen, Heikki Mannila. **Towards interactive visual analysis of corpora.** ICAME 31, 26 - 30 May, Giessen, Germany, 2010. (Poster)
- Jefrey Lijffijt. **Local and global lexicon: a novel approach to quantifying persistence.** XXXVII Kielitieteen päivät Helsingin Yliopistossa, 20 - 22 May, Helsinki, Finland, 2010. (Presentation)

Other Presentations and Posters

- Jefrey Lijffijt. **Analysis of linguistic variation.** Poster: Spring Workshop on Mining and Learning (SML), Bad Neuenahr, Germany, 2012. (Poster)
- Jefrey Lijffijt. **Data mining tools for analysis of linguistic variation.** Poster: Lorentz Workshop on Mining Patterns and Subgroups, Leiden, The Netherlands, 2010. (Poster)

- These are just the external presentations not directly related to papers

Besides talking to potential end-users, I presented preliminary results lots and lots of times. On this slide are presentations and posters at conferences where they have some form of selection procedure, but no proceedings like commonly seen in CS.

Other feedback used

- Harassed the local post-doc with all my thoughts
- Two mentors from graduate school
- After (poster) presentations at
  - graduate school,
  - workshops and conferences,
  - HIIT,
  - ALGODAN Centre of Excellence.

How to do good research

- Ask yourself some questions
- If I solve this problem, then
  - So what?
  - Who should care about this?
  - Why should they care?
- Write this down *before* you start

When you have a ‘great’ idea, I would suggest that you ask yourself these questions. Answering them may make you realised that this idea is perhaps not so great, or maybe that you still have to think more about certain aspects. If you can answer these questions clearly, you have the motivation part of your paper.

Some final advice: get published ≠ do good research

- Learn to write scientific articles
  - Read Eamonn Keogh's tutorial slides
  - Read guides on writing style
  - Practice, practice, practice
- "To get published, writing is more important than contents." (personal opinion)

In my experience, there is a large difference between getting published and doing great research. CS is an engineering discipline, and I often see that students have solved some problem and then think that they are done. You are not, because science requires that you actually inform the world about your achievement.

With respect to publishing, I think you can write so poorly that even the best scientific results can be rejected because they go unrecognised, and vice versa, it is possible to publish very incremental results if you write them up very well.

Contents of this talk

1. Why do a PhD?
2. How to gain interesting knowledge & skills
3. How to become independent
4. How to do good research
5. **How to keep going**
6. [How to finish your PhD]

How to do keep going

- “[…] drop-out rates suggest that many students become dispirited. In America only 57% of doctoral students will have a PhD ten years after their first date of enrolment.” [Economist, 16 Dec 2010]
- Most students have periods with low motivation
  - Uncertain future, lack of achievements, loneliness (in work, life, or both)

Reading other sources regarding this number, in truth it appears to be that between one third and half of students quit their studies. Of course some quit because they encounter a great opportunity, take it, and no longer care about the degree. However, for most quitters that is probably not the case.

I think it is good to realise in advance that a PhD is tough, and you will have periods that are unhappy about it, have low motivation, become unproductive, or worse; depression etc.

How to do keep going

- Be wary for demoralisation
- Focus on the fun part or why you are here
- Have a life (at the department)
  - Meet people, talk about your ambitions and your research
- Attend as many departmental events as possible

I think these periods can be controlled and minimised by being wary for them and doing the following things.

Contents of this talk

1. Why do a PhD?
2. How to gain interesting knowledge & skills
3. How to become independent
4. How to do good research
5. How to keep going
6. **[How to finish your PhD]**

How to finish your PhD

- “There are two kinds of theses: perfect theses and finished theses.” [Heikki Mannila, personal correspondence]
- In the end, just finish, your PhD is the start, not the end
- Do everything you can to finish while you are funded
- Start by writing a very blunt full draft, as fast as you can
  - Don’t look back or optimise any text
- Then ask feedback, and write it all again
- You are all smart, if you work hard, you *will* make it

People may have different ways to make themselves productive writers, however in general I think it often works for your thesis to take this advice; do a brain dump and optimise nothing of your text. Then ask feedback about the structure, the contents, whether parts are missing, etc. Then write it again. This feedback should of course come from your advisor, but you could also ask another colleague or some friend.

Summary

- Focus on educating yourself
- Engage with others as much as possible
- With respect to research problems, ask yourself
  - So what?
  - Who should care about this?
  - Why should they care?


Thank you!