Welcome to Philosophy of Science and Technology (GBI)

Welcome to Philosophy of Science (DMD)

Welcome to Reflections on IT (SDT)

Lecture 6: Values in Design

Judith Simon

IT UNIVERSITY OF COPENHAGEN

Course Overview

- L1: Introduction (Vasiliki & Christopher)
- L2: Conceptions of Technology (Vasiliki)
- L3: Post-Phenomenology & Technology (Finn Olesen)
- L4: Laboratories and non-human actors in STS (Christopher)
- L5: Computer Ethics (Judith)
- L6: Values in Design (Judith)
- L7: Workshop
- L8: Surveillance (Christopher)
- L9: The Relevance of Algorithms (Vasiliki)
- L10: Topic course-specific (Main lecturer)
- L11: Topic course-specific (Main lecturer)
- L12: Overview (Main lecturer)

Course Overview

- L1: Introduction (Vasiliki & Christopher)
- L2: Conceptions of Technology (Vasiliki)
- L3: Post-Phenomenology & Technology (Finn Olesen)
- L4: Laboratories and non-human actors in STS (Christopher)
- L5: Computer Ethics (Judith)
- L6: Values in Design (Judith)

Roots of Values in Design

- Philosophy of Technology
- Science and Technology Studies
- Computer Ethics
- CSCW
- Participatory Design
- Social Informatics

→ Values in Design

What is Computer Ethics?

Professional Ethics	Ethics of Usage	Ethics of Design
Designer & Developer	User & Usage	IT Artefacts

Values in Design

Values in Design

Helen Nissenbaum
Values in Design

Batya Friedman:
Value-sensitive
Design

Mary Flanagan:
Values at Play

Philip Brey:
Disclosive
Computer Ethics

Philip Brey: Departing from computer ethics

„Is it possible to do an ethical study of computer systems themselves independently of their use by human beings?“ (Brey 2010)

Typically computer ethics focuses on usage of technologies:

- Is it wrong for a system operator to disclose the content of employee email messages to employers or other third parties?
- Should individuals have the freedom to post discriminatory, degrading and defamatory messages on the internet?
- Is it wrong for companies to use data-mining techniques to generate consumer profiles based on purchasing behaviour, and should they be allowed to do so?
- Should governments design policies to overcome the digital divide between skilled and unskilled computer users?

Brey, P. (2010). Values in Technology and disclosive computer ethics. The Cambridge Handbook of Information and Computer Ethics. L. Floridi. Cambridge, Cambridge University Press: 41-58.

Philip Brey

Embedded Values Approach

„The embedded values approach holds that computer systems and software are not morally neutral and that it is possible to identify tendencies in them to promote or demote particular moral values and norms.“ (p. 42)

→„Computer ethics should not just study ethical issues in the use of computer technology, but also in the technology itself.“ (p. 42)

Philip Brey

- How technology embodies values

→ Science and Technology Studies (e.g. Latour & Winner)

Latour (1992):

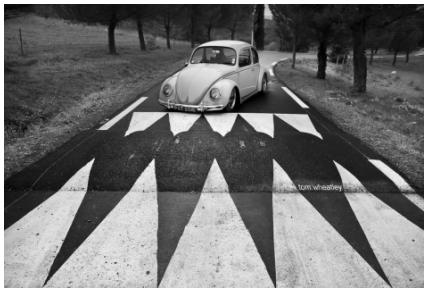
Red light, alarm enforcing driver to wear a seatbelt while driving.

→ Assembly of car & driver obeys the law.

Latour (1992):

- Delegation to humans and non-humans

Latour (1992):



Latour (1992):

"We have been able to delegate to nonhumans not only force (…) but also values, duties, and ethics. It is because of this morality that we humans behave so ethically, no matter how weak and wicked we feel we are."

- We delegate tasks, norms, duties, competences, values to things
- This delegation affects us in return, e.g. we have adopt our speed to the door, we reduce speed in case of the speed bump, we use the seatbelt, …
- The form of delegation/prescription may discriminate against certain humans, e.g. very small/old persons, furniture removers in case of the hydraulic door closer

Philip Brey

- How technology embodies values

→ Science and Technology Studies (e.g. Latour & Winner)

Do artefacts have politics?

Do artefacts have politics?

Yes, technologies embody „specific forms of power and authority"
(Winner 1980: 121)

Langdon Winner

Do artefacts have politics?

http://de.wikipedia.org/wiki/Robert_Moses

Langdon Winner

Winner argues that the design of those parkway bridges reflects "…Moses's social-class bias and racial prejudice" (Winner 1980: 123) and concludes: "Many of his monumental structures of concrete and steel embody a systematic social inequality, a way of engineering relationships among people that, after a time, becomes just another part of the landscape." (Winner 1980: 124)

http://de.wikipedia.org/wiki/Robert_Moses

Langdon Winner

Do politics have artefacts?

http://de.wikipedia.org/wiki/Robert_Moses

Langdon Winner

The refutation which did not matter - Joerges (1999) based on correspondences with US civil engineers:

"*Moses could hardly have let buses on his parkways, even if he had wanted differently"(Joerges 1999: 419).*

http://de.wikipedia.org/wiki/Robert_Moses

Langdon Winner

- Strong case for the *inscription of societal values into technology and the societal effects of such biased technologies*
- Political character of artifacts and the possibility of social engineering through technology
- Against neutrality of technologies
- Intentional inscription of values/biases into technology through a powerful agent
- Need to critically analyse artefacts

Philip Brey

Core: „The idea of embedded values is best understood as a claim that technological artefacts (and in particular computer systems and software) have built-in tendencies to promote or demote the realization of particular values. Defined in this way, a built-in value is a special sort of built-in consequence."

Philip Brey

- Consequences built into technology vs. neutrality thesis

„The neutrality thesis holds that there are no consequences that are inherent to technological artefacts, but rather that artefacts can always be used in a variety of different ways, and that each of these uses comes with its own consequences." p. 43
(e.g. hammer)

Bias: Friedman & Nissenbaum

Friedman & Nissenbaum (1997)

- Analyses of bias in computer systems
  - Differentiation of three types of bias
  - Claim that freedom from bias should be criterion for the evaluation of computer systems similar to criteria of usability, reliability, accuracy, etc.
  - Examples from the 1990ies, but the types of biases are still instructive for assessing contemporary systems

Friedman & Nissenbaum (1997)

- What is bias?
  - Bias means „slant" and can be applied to neutral or moral content
  - Biased systems are computer systems that systematically and unfairly discriminate against certain (groups of) individuals
  - Systematic: Random error is not bias
  - Unfair: There must be an unfair outcome
- Example 1) Flight booking system prefering on-line/single carrier flights – biased against international carrier
- Example 2): Automated credit advisor based upon poor payment records (justified difference) versus ethnic surnames (biased)

Friedman & Nissenbaum (1997)

- Types of Biases
  - Derived by examining actual computer systems for bias
  - 17 computers systems examined from diverse fields: banking, commerce, computer science, education, medicine and law
  - Instances of bias are identified and characterized according to their source
- Pre-existent bias
- Technical bias
- Emergent bias

Friedman & Nissenbaum (1997)

Table I. Categories of Bias in Computer System Design

These categories describe ways in which bias can arise in the design of computer systems. The illustrative examples portray plausible cases of bias.

1. Preexisting Bias
Preexisting bias has its roots in social institutions, practices, and attitudes. When computer systems embody biases that exist independently, and usually prior to the creation of the system, then the system exemplifies preexisting bias. Preexisting bias can enter a system either through the explicit and conscious efforts of individuals or institutions, or implicitly and unconsciously, even in spite of the best of intentions.

1.1. Individual
Bias that originates from individuals who have significant input into the design of the system, such as the client commissioning the design or the system designer (e.g., a client embeds personal racial biases into the specifications for loan approval software).

1.2 Societal
Bias that originates from society at large, such as from organizations (e.g., industry), institutions (e.g., legal systems), or culture at large (e.g., gender biases present in the larger society that lead to the development of educational software that overall appeals more to boys than girls).

Friedman & Nissenbaum (1997)

2. Technical Bias
Technical bias arises from technical constraints or technical considerations.

2.1 Computer Tools
Bias that originates from a limitation of the computer technology including hardware, software, and peripherals (e.g., in a database for matching organ donors with potential transplant recipients certain individuals retrieved and displayed on initial screens are favored systematically for a match over individuals displayed on later screens).

2.2 Decontextualized Algorithms
Bias that originates from the use of an algorithm that fails to treat all groups fairly under all significant conditions (e.g., a scheduling algorithm that schedules airplanes for take-off relies on the alphabetic listing of the airlines to rank order flights ready within a given period of time).

2.3 Random Number Generation
Bias that originates from imperfections in pseudorandom number generation or in the misuse of pseudorandom numbers (e.g., an imperfection in a random-number generator used to select recipients for a scarce drug leads systematically to favoring individuals toward the end of the database).

2.4 Formalization of Human Constructs
Bias that originates from attempts to make human constructs such as discourse, judgments, or intuitions amenable to computers: when we quantify the qualitative, discretize the continuous, or formalize the nonformal (e.g., a legal expert system advises defendants on whether or not to plea bargain by assuming that law can be spelled out in an unambiguous manner that is not subject to human and humane interpretations in context).

Friedman & Nissenbaum (1997)

3. Emergent Bias
Emergent bias arises in a context of use with real users. This bias typically emerges some time after a design is completed, as a result of changing societal knowledge, population, or cultural values. User interfaces are likely to be particularly prone to emergent bias because interfaces by design seek to reflect the capacities, character, and habits of prospective users. Thus, a shift in context of use may well create difficulties for a new set of users.

3.1 New Societal Knowledge
Bias that originates from the emergence of new knowledge in society that cannot be or is not incorporated into the system design (e.g., a medical expert system for AIDS patients has no mechanism for incorporating cutting-edge medical discoveries that affect how individuals with certain symptoms should be treated).

3.2 Mismatch between Users and System Design
Bias that originates when the population using the system differs on some significant dimension from the population assumed as users in the design.

3.2.1 Different Expertise
Bias that originates when the system is used by a population with a different knowledge base from that assumed in the design (e.g., an ATM with an interface that makes extensive use of written instructions—"place the card, magnetic tape side down, in the slot to your left"—is installed in a neighborhood with primarily a nonliterate population).

3.2.2 Different Values
Bias that originates when the system is used by a population with different values than those assumed in the design (e.g., educational software to teach mathematics concepts is embedded in a game situation that rewards individualistic and competitive strategies, but is used by students with a cultural background that largely eschews competition and instead promotes cooperative endeavors).

Friedman & Nissenbaum (1997)

- Examples
  1. The National Resident Match Program (NRMP)
     - Matches med school graduates with jobs
     - Systematically prefers hospital preference over graduate preference in cases of conflict (pre-existent bias)
     - Additional problem of married graduates (emergent bias)
     - Due to unavoidability of system, bias is particularly troublesome

Friedman & Nissenbaum (1997)

2. A Multilevel Scheduling Algorithm (MLSA)
   - Timeshare computer systems – sharing processing power
   - Discriminating against users with long-running processes (technical bias)

Friedman & Nissenbaum (1997)

3. The British Nationality Act Program (BNAP)
   - 1981: British Citizenship, British Dependent Territories, British Overseas Citizenship
   - Biased against children of unmarried British men
   - Act was transferred into a computer program → pre-existent bias + emergent bias due to usage of system in non-expert environments
   - General: any legal expert system faces danger of emergent bias when legal system depends upon case law

Friedman & Nissenbaum (1997)

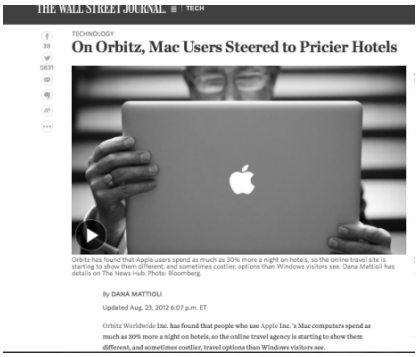
- Goals
  - identify bias in any given system
  - develop methods of avoiding bias in systems and correcting it when it is identified
- Minimizing bias
  - **Pre-existent**: good understanding of relevant societal biases
  - **Technical:** envision it in a context of use, envision the design etc.
  - **Emergent**: envision not only a system's intended situations of use, but to account for increasingly diverse social contexts of use; anticipate probable contexts of use and design for these; where it is not possible to design for extended contexts of use, designers should attempt to articulate constraints on the appropriate contexts of a system's use.

Friedman & Nissenbaum (1997)

„[…] In practice we must approach actively the task of minimizing bias in our designs. Furthermore, as a community we must hold our designs accountable to a reasonable degree of freedom from bias against which negligence can be judged."

Friedman & Nissenbaum (1997)

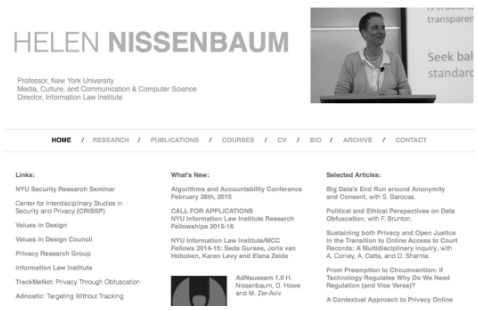
Friedman & Nissenbaum (1997)



Values in Design: The Practical Turn

Values in Design: The Practical Turn

Values in Design: The Practical Turn

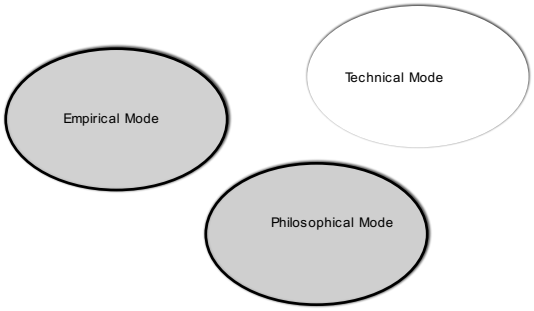


Flanagan, Nissenbaum & Howe (2008)

- Main premise: values are embodied in technical systems
- If that's the case, can we intentionally embody desired values into technology? If so, how?

„If an ideal world is one in which technologies promote not only instrumental values such as functional efficiency, safety, reliability and ease of use, but also substantive social, moral and political values to which societies and their people subscribe, then those who design systems have a responsibility to take this latter values as well as the former into consideration as they work." (322)

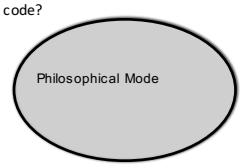
Flanagan, Nissenbaum & Howe (2008)



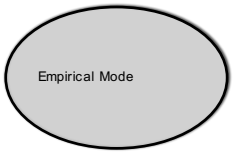
Flanagan, Nissenbaum & Howe (2008)

B. Philosophical Mode - questions about…

- i) origin and source of values
- ii) conceptual investigation of particular values
- iii) how to translate these values into design decisions, into software specifications, into code?



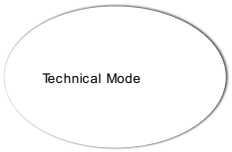
Flanagan, Nissenbaum & Howe (2008)



Empirical Mode

- Which values are desired by which relevant groups?
- Analysis if intended values are embodied in technology

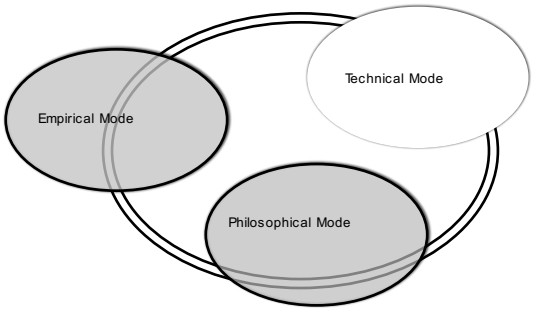
Flanagan, Nissenbaum & Howe (2008)



Technical Mode

- „state-of-the-art knowledge on design specifications" (325)
- invention and innovation
- how to translate these values into design decisions, into software specifications, into code?

Flanagan, Nissenbaum & Howe (2008)



Flanagan, Nissenbaum & Howe (2008)

- Aim: providing one methodological framework for incorporating values during design processes
- Social values are evaluation criteria for software in addition to classical values such as efficiency, reliability, robustness, safety, etc.
- Interdisciplinary efforts needed: social scientists, philosophers & computer scientists working together

Flanagan, Nissenbaum & Howe (2008)

rapunsel
Real Time Programming for Underrepresented Student's Early Literacy

- Values at Play → apply approach to game design

Flanagan, Nissenbaum & Howe (2008)

rapunsel
Real Time Programming for Underrepresented Student's Early Literacy

- Goal: promote interest and competence in computer programming among girls at middle-school age, incl. girls from disadvantaged homes
- Hypothesis: marked absence of women in the field of technology development, particularly in computer programming, is due, at least in part, the style in which basic scientific subjects are taught.

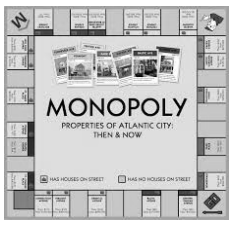
Flanagan, Nissenbaum & Howe (2008)

Embodying the values - Three activities to be done:
a) Discovery: compile list of relevant values
b) Translation: operationalize and implement values into material design features
c) Verification: test implementations to verify whether design intentions have been met

Flanagan, Nissenbaum & Howe (2008)

Two types of values:
a) Expressed values: related to game content, e.g. character representation, game plot, …
b) Materially embodied values: constraints and affordances through specific design decisions, e.g. reward system,
→ Moor's emphasis on invisibility

Flanagan, Nissenbaum & Howe (2008)



New Value:
Cooperation

Flanagan, Nissenbaum & Howe (2008)

New Value:
Trust

Value-Sensitive Design: Friedman et al.

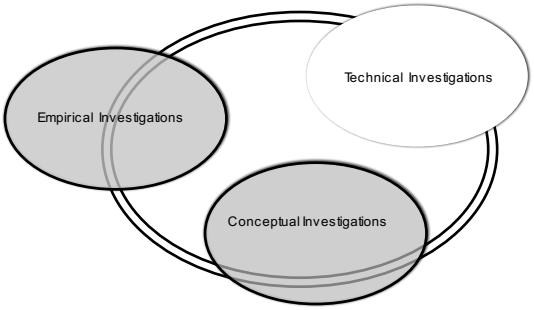
Friedman et al. (2006)

- Value: "the principles or standards of a person or society, the personal or societal judgement of what is valuable and important in life."
  (Oxford English Dictionary)
- Basis of VSD: Computer Ethics, Social Informatics, CSCW, Participatory Design

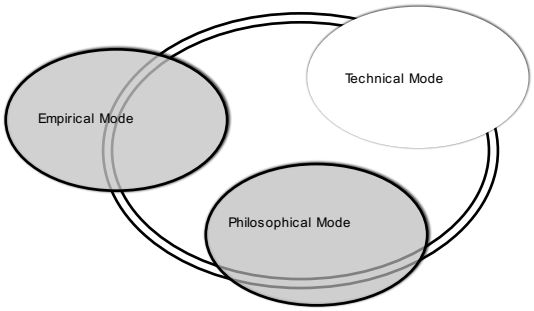
Friedman et al. (2006)

"Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. It employs an integrative and iterative tripartite methodology, consisting of conceptual, empirical, and technical investigations"

Friedman et al. (2006)

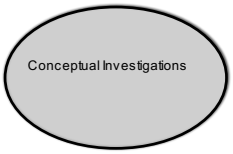


Flanagan, Nissenbaum & Howe (2008)

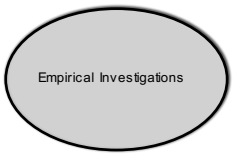


Friedman et al. (2006)

Who are the direct and indirect stakeholders affected by the design at hand? How are they affected? What values are implicated? How should we engage in trade-offs among competing values in the design, implementation, and use of information systems (e.g., autonomy vs. security, or anonymity vs. trust)? Should moral values (e.g., a right to privacy) have greater weight than, or even trump, non-moral values (e.g., aesthetic preferences)?



Friedman et al. (2006)

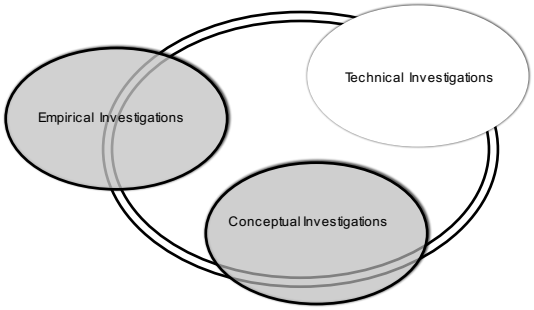


1) How do stakeholders apprehend individual values in the interactive context? How do they prioritize competing values in design trade-offs? How do they prioritize individual values and usability considerations? Are there differences between espoused practice (what people say) compared with actual practice (what people do)?
2) Evaluation of systems

Friedman et al. (2006)

1) focus on how existing technological properties and underlying mechanisms support or hinder human values
2) proactive design of systems to support values identified in the conceptual investigation

Friedman et al. (2006)



Friedman et al. (2006)

- Ex. 1: Cookies & Informed Consent
- Conceptualize "informed consent" as core value
- Use this investigation to analyze existing technologies
- Redesign Mozilla browser based upon tripartite methodology
- Emergent design criterion of "minimal distraction"

Friedman et al. (2006)

- Ex. 2: Room with a View
- Premise: psychological benefits of watching nature
- Importance of both direct & indirect stakeholders
- Conflicting values of different stakeholders

Friedman et al. (2006)

- Ex. 3: UrbanSim: Simulation to plan urban development

"a large simulation package for predicting patterns of urban development for periods of twenty years or more, under different possible scenarios [..]. Its primary purpose is to provide urban planners and other stakeholders with tools to aid in more informed decision-making, with a secondary goal to support further democratization of the planning process."

Friedman et al. (2006)

- Ex. 3: UrbanSim: Simulation to plan urban development
- Variety of values at stake
  - Explicitly supported values of system (fairness, freedom from bias, accountability, democracy)
  - Versus (divergent & potentially conflicting) stakeholder values
  - Distillation of economic, environmental, social indicators
- Technological implications of supported values (e.g. fairness, accountability, …)

Friedman et al. (2006)

- VSD – an unique constellation of **eight features**:

1. VSD seeks to be proactive; e.g. UrbanSim → influence design early on
2. VSD covers the issues beyond workplace; e.g. UrbanSim
3. VSD introduces a unique tripartite methodology; e.g. cookies project
4. VSD enlarges the scope of human values beyond cooperation

Friedman et al. (2006)

5. VSD distinguishes between usability and human values with ethical import; e.g. cookies project

6. VSD identifies (i) direct stakeholders and (ii) indirect stakeholders; e.g. room with a view work

7. VSD is an *interactional theory*, values are neither inscribed into technology nor simply transmitted by social forces → technologies hinder/support certain action and are societally influenced themselves

8. VSD accepts certain values to be universal on an abstract level

Friedman et al. (2006)

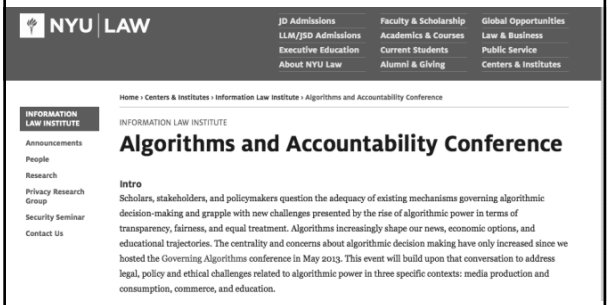
- Practical suggestions for using VSD
  1. start with a value, technology, or context of use
  2. identify direct and indirect stakeholders
  3. identify benefits and harms for each stakeholder group
  4. map benefits and harms onto corresponding values
  5. conduct a conceptual investigation of key values
  6. identify potential value conflicts
  7. integrate value considerations into organizational structure
  8. check their list of human values (with ethical import) often implicated in system design
  9. empirical heuristics: semi-structures interviews for stakeholder interviews + open questions
  10. technical heuristics: map value-conflicts with design trade-offs for different stakeholders; design for flexibility to account for emergent bias; allow for control of information flow.

Some more examples…

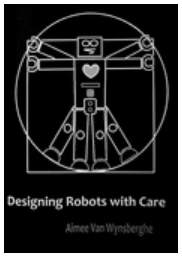
Privacy by Design

Privacy & TrackMeNot

Algorithmic Accountability



VID & Care Robots



„What will the impact of the robot be on the provision of good care, on the manifestation of care values; will it change the standards of care and ultimately lower them?
Will care robots displace and/or de-skill care workers?
What will the existential impact be on the care-giver and the care-receiver; is the use of robots a devaluing of either or both?…"

http://www.aimeevanwynsberghe.com/phd-dissertation.html

Course Overview

L1: Introduction (Vasiliki & Christopher)	L7: Workshop
L2: Conceptions of Technology (Vasiliki)	L8: Surveillance (Christopher)
L3: Post-Phenomenology & Technology (Finn Olesen)	L9: The Relevance of Algorithms (Vasiliki)
L4: Laboratories and non-human actors in STS (Christopher)	L10: Topic course-specific (Main lecturer)
L5: Computer Ethics (Judith)	L11: Topic course-specific (Main lecturer)
L6: Values in Design (Judith)	L12: Overview (Main lecturer)

Thank you all very much for your attention!