Economics, Law and Ethics

Part IB CST

2025-26

Lecture 7: Philosophies of ethics

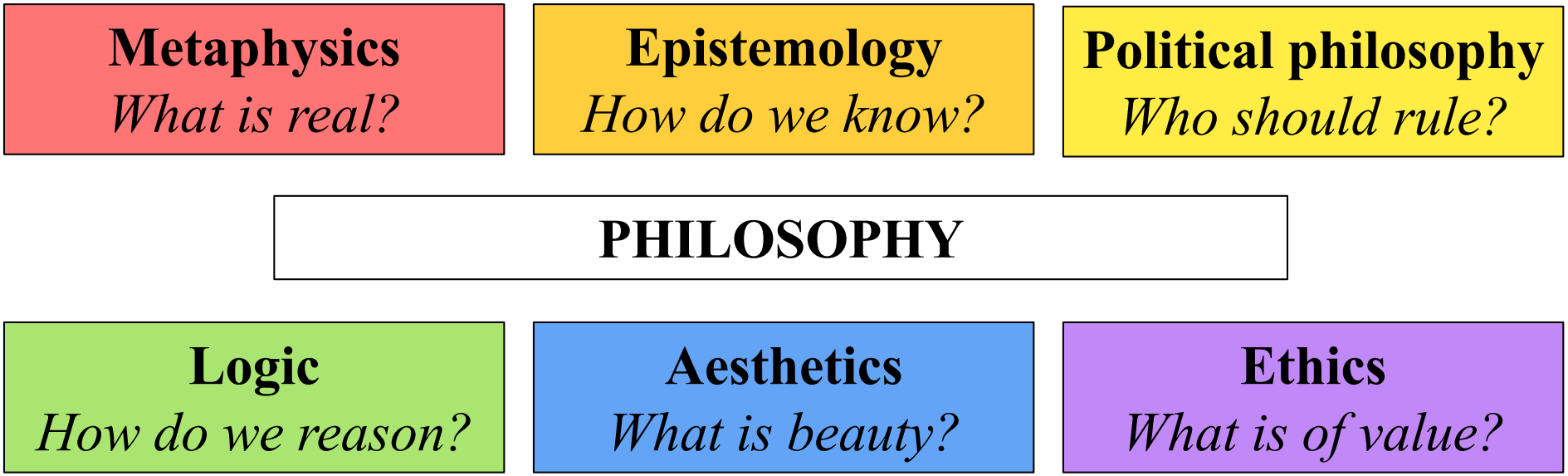
Alice Hutchings

Overview

- Philosophies of ethics:
  - An overview of philosophy
  - Ethical frameworks
  - Philosophical conflicts
  - Professional codes of ethics
  - Coordinated vulnerability disclosure
  - Ethics in research

Philosophy overview

- Ethics is one of the main branches of philosophy



Philosophy overview

- Each branch shapes our perspectives and decision-making processes
- Philosophy provides a tool to address questions of technology ethics; philosophy is ‘the software our minds work on’ (Hare, 2022)

Metaphysics

- What is reality?
  - The nature, structure, and origins of the universe
  - Misinformation; disinformation; algorithmic decision-making (does output of an algorithm/database query match our understanding of reality?); consciousness (Turing: can machines think?); what is reality when virtual or augmented?

Epistemology

- What does it mean to know?
  - What are our sources of knowledge? How do we acquire knowledge? What are its limitations?
  - How can we know if machines can think? Can results be reproduced? How do blackbox AI models reach a conclusion? Should mis/disinformation/hate speech be free or censored? Who creates knowledge and how is it established, tested, and verified?

Political philosophy

- What is the nature of power and legitimacy?
  - What is the relationship between individuals and society? What sort of society do we want to live in? How should it be organised? Who should rule? What are people's rights and responsibilities?
  - Technology companies as political actors (power, wealth, market dominance); shaping political discourse (censorship v. freedom of expression); how big platforms are regulated and how power is centralised (privacy; civil liberties; human rights)

Logic

- How do we know what we know?
  - We use logic to determine if an argument is sound or if a hypothesis is supported
  - How do we know that you are you? (digital identity, facial recognition and other biometrics technologies); CAPTCHAs to differentiate between human and bots

Aesthetics

- What is experience?
  - Relates to senses and perception
  - User interface (UI) and user experience (UX) leading to compulsive use; accessibility and inclusivity (design needs of users: with low vision; who use screen readers; who are deaf or hard of hearing; with physical or motor disabilities; dyslexic or autistic people); value-sensitive design; personalisation; data handover between platforms; friction; data visualisation

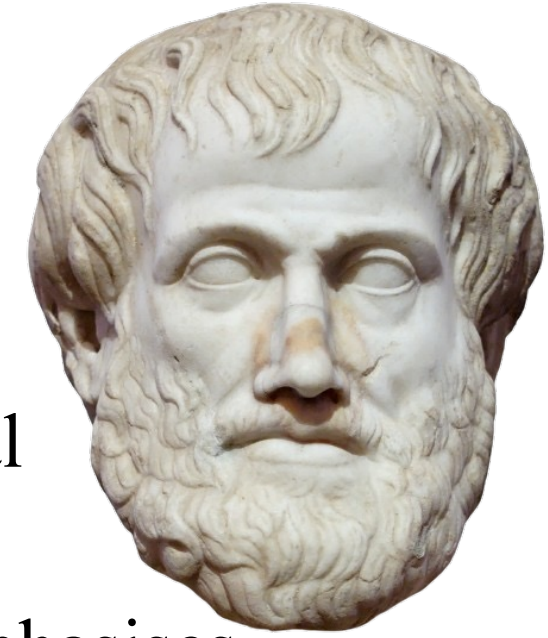
Ethics

- How should we live?
  - What is right and wrong? What constitutes a good life?
  - Technology ethics: applied ethics (ethics put into practice)
  - A continued conversation, not just a checklist, as dynamic as the technology it is applied to
  - Not just the responsibility of a person or a legal/policy/public relations team; concerns everyone involved in a product/service throughout its lifecycle

Ethical frameworks

- Practical ethics: In what circumstances should we restrain our actions more than the law requires?
- Collingridge dilemma: early technology is easy to regulate, but risks are unclear; later, harms are clear, but regulation is difficult
- Self-regulation in academia, through ethics committees, to reduce harm to participants
- Businesses: Front page test
- Hippocratic oath for computer scientists?

Virtue ethics



- Hippocratic oath: emphasises the moral character and virtues of the physician
- What kind of person should I be?: Emphasises the decision-maker's agency and character
- Influences: Aristotle, Confucian ethics, Aquinas
- Virtues:
  - stable character traits that enable people to act well
  - developed through habit and practice, not just intention

Virtue ethics

- Virtues lie between extremes (vices):

Rashness	**Courage**	Cowardice
Wastefulness	**Generosity**	Stinginess
Deceitfulness	**Honesty**	Tactlessness
Indifference	**Compassion**	Enabling
Insensibility	**Self-Control**	Gluttony

- Well-tested, peer-reviewed code pushed to production is **courage**; shipping untested code to live servers is **rashness**; refusing to deploy safe fixes is **cowardice**

Virtue ethics

- Difficulties:
  - Identifying/agreeing on virtues
    - May differ across cultures/communities, no clear consensus on what traits should guide action
  - Emphasis on individual moral reasoning
    - Is it always ethical to be honest when that may result in harm to others?
  - No clear method for resolving clashes between virtues
    - Honesty and kindness

Do you know where the Ubuntu Linux distribution gets its name from?

The philosophy behind the name

- Ubuntu is a fusion of normative ideas that largely inform beliefs, attitudes, and practices in Sub-Saharan Africa.
- Supports collectivism over individualism.
- Central values include reciprocity, common good, peaceful relations, human dignity, the value of human life, consensus, tolerance, and mutual respect.
- Maxim: *umuntu ngumuntu ngabantu.*
  "I am because we are"; or "a person is a person through other persons."

Ubuntu the OS

- Name was chosen to reflect the African ethical philosophy
- Developed and maintained by a global open-source community that aims to make computing accessible and empowering
- Free and open source, enabling users worldwide to use, modify and share Ubuntu regardless of wealth or location
- Code of conduct: ‘be respectful’, ‘take responsibility’, ‘be collaborative’, etc.

Ubuntu

- Difficulties:
  - Risks subordinating individual autonomy, dissent or minority rights to collective interests
  - Potential for patriarchal or hierarchical reinforcement under the banner of preserving harmony

Consequentialism

- Principles of consequentialist approaches:
  - Whether an act is right or wrong depends only on the results of that act
  - The more good consequences an act produces, the better or more right that act

Consequentialism

- Consequentialist theories include Hume, Bentham and Mill's utilitarianism: maximise $W = \sum U_i$ (or, 'greatest happiness of the greatest number')
- Difficulties:
  - Predicting consequences: uncertainty, long-term effects, unintended consequences
  - Controversial applications: e.g.. appeals to catastrophic consequences to defend torture in anti-terrorism

Deontology

- Deontological approaches:
  - What is right and wrong depends on duties, principles, or motives – not solely on consequences
- Some actions are morally required/forbidden regardless of outcomes

Deontology

- Kantian ethics:
  - Categorical Imperative: universal principle of morality (it should make sense for everyone to act that way)
  - Never treat people as means to an end, but as an end in themselves
- John Rawls 'Theory of Justice':
  - Make moral decisions from behind a "veil of ignorance" – not knowing class, wealth, abilities, or social position
  - Maximise the welfare of the worst-off: $W = \min U_i$

Deontology

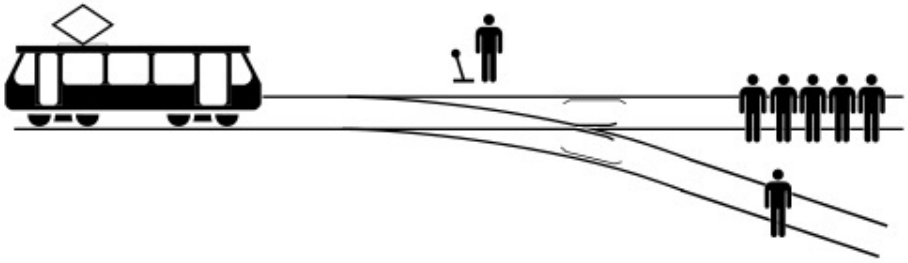
- Difficulties:
  - Can require following a rule even when doing so has obvious harmful consequences (e.g. telling the truth to someone wishing to harm another)
  - Concepts like ‘categorical imperative’ are not always intuitive for guiding everyday decisions

Questions to ask

Virtue ethics	• Does this align with my values and character? • Will it contribute to moral growth or flourishing?
Ubuntu ethics	• Will this uphold dignity and strengthen relationships? • Does it promote communal harmony and collective wellbeing?
Consequentialism	• Which option produces the greatest overall benefit? • What are the likely harms and gains of each outcome?
Deontology	• Am I respecting others as ends, not means? • Would I accept this as a universal rule?

Trolley problems

- Philippa Foot and Judith Thompson
- Thought experiments that reveal conflicts between ethical principles.
  - A runaway trolley is heading down a track towards five people. They will be killed if it continues. You can pull a lever to divert the trolley onto another track, but one person will die on that track.
  - Do you pull the lever?



Trolley problems (cont.)

- Consequentialism:
  - Actions are right if they produce the greatest good for the greatest number
  - Pull the lever – saving five lives is better than one
- Deontological ethics:
  - Actions are right or wrong based on moral rights or duties, regardless of outcomes
  - Don’t pull the lever – it’s wrong to intentionally kill an innocent person

Philosophical conflicts

- Trolley problems are simplified thought experiments
- Other conflicts between ethical frameworks frequently occur, e.g.:
  - Ethical implications of using facial recognition technology for applications that include surveillance and law enforcement or unlocking phones and devices

Philosophical conflicts (cont.)

- Potential benefits:
  - increased security
  - convenience
- Potential harms:
  - privacy
  - accuracy

Philosophical conflicts (cont.)

- Consequentialist approach: Weigh the costs and benefits
  - Facial recognition technology can be used if the social benefits outweigh the potential harms
- Deontological approach: Focus on the principles
  - Facial recognition technology can be unethical even if it has positive outcomes if it violates duties like consent or privacy

Professional codes of ethics

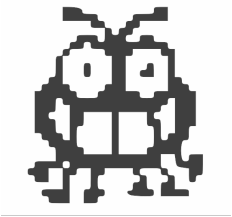
- Typically rules-based
- ACM's code of ethics https://ethics.acm.org/code-of-ethics/using-the-code/
- A computing professional should…
  - Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing
  - Avoid harm
  - Be honest and trustworthy
  - Be fair and take action not to discriminate
  - Respect the work required to produce new ideas, inventions …
  - Respect privacy
  - Honour confidentiality

Coordinated vulnerability disclosure

- If vulnerabilities are found, a range of responses from not disclosing to immediately making public
- Coordinated disclosure: Confidential disclosure to those who can remedy or mitigate the impact
- Public disclosure then occurs after a period of time has elapsed

Coordinated vulnerability disclosure (cont.)

- Those who have the ability to fix the vulnerability may not have the incentive to do so
- Public disclosure after a set period of time changes those incentives (also informs the security and practitioner community, encourages patches to be installed, etc)
- How to encourage people to report the vulnerability in the first place? Why report it when you can exploit it or sell it?



Ethics in research

- 1940s: Nazi human experimentation
- 1930s-1970s: Tuskegee syphilis experiment
- 1960s: The Milgram experiment
- 1970s: Stanford prison experiment

But it’s all fine now, right?

Ethics in research

- 2010s: Facebook emotional manipulation study
  - 700,000 users' news feeds were manipulated to display positive/negative posts
  - Emotional contagion: after viewing negative posts, users post more negatively
  - No informed consent
  - Deliberately induced negative emotions in unwitting participants

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b], and Jeffrey T. Hancock[c,d]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; [b]Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143; and Departments of [c]Communication and [d]Information Science, Cornell University, Ithaca, NY 14853

Ethics in research

- 2010s: Predicting sexual orientation using facial analysis algorithms
  - Data obtained from an online dating site without consent
  - Inferring sexual orientation is sensitive
  - Questions around privacy, bias, and potential misuse

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

By Wang, Yilun,Kosinski, Michal

Journal of Personality and Social Psychology, Vol 114(2), Feb 2018, 246-257

Ethics in research

- 2020s: Predicting criminality using facial analysis algorithms
  - Reproduces biases in CJS
  - Pseudoscientific theories of biological determinism relating to criminality are completely discredited
  - Paper was widely condemned and later withdrawn

Ethics in research

- 2020s: Reddit AI persuasion study
  - Used AI-authored posts on r/changemyview to measure if they were more successful in changing opinions than human-authored posts
  - Purported to be from a range of human identities
  - Widely condemned for experimenting on users without their knowledge or consent

(and many others!)

Ethics in research

- Research Ethics Boards:
  - Ethics Committees in the UK, Institutional Review Boards (IRBs) in the US
- Research funding bodies
- Program committees and journal editors
- Professional Ethical Guidelines or Codes of Practice
- For computer science: The Menlo Report
  - Core principles: respect for persons, beneficence, justice, and respect for law and public interest.

Ethics in research

- Your Part II project may involve human participants
- Independent review by uninvolved scientists greatly reduces risks of both civil litigation and criminal prosecution if things go wrong
- Pay attention to the department's ethics policy

https://www.cst.cam.ac.uk/local/policy/ethics