Can Ethics Be Taught?

Phillip Rogaway
Nomore University
Formerly: University of California, Davis, US
Chiang Mai University, TH
++
rogaway@pm.me
https://web.cs.ucdavis.edu/~rogaway/

Thanks to Tjerand Silde, Katrien De Moor, Emil André Røyrvik, Petter Grytten Almklov and everyone else involved in organizing this event. I appreciate the kind invitation to come give a talk here.

Thanks to the ~700 students who helped me learn about ethics and technology over the last 20 years. You are missed.

Cryptology and Social Life
NTNU — Trondheim, Norway
11 December 2025

Today
Part 1: Revisiting my 2016 talk here in Trondheim
Part 2: Why ethics can't be taught
Part 3: Twenty years of teaching ethics

Delighted to return here

The Moral Character of Cryptographic Work*

Phillip Rogaway

Department of Computer Science
University of California, Davis, USA
rogaway@cs.ucdavis.edu

December 2015
(minor revisions March 2016)

Abstract. Cryptography rearranges power: it configures who can do what, from what. This makes cryptography an inherently *political* tool, and it confers on the field an intrinsically *moral* dimension. The Snowden revelations motivate a reassessment of the political and moral positioning of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means to resist mass surveillance. I plead for a reinvention of our disciplinary culture to attend not only to puzzles and math, but, also, to the societal implications of our work.

Keywords: cryptography · ethics · mass surveillance · privacy · Snowden · social responsibility

Preamble. Most academic cryptographers seem to think that our field is a fun, deep, and politically neutral game—a set of puzzles involving communicating

Cryptography *vs.* Mass Surveillance

Phillip Rogaway
Department of Compute
University of California, D

Talk for
Crypto vs. Mass Surveillance:
The Uneasy Relationship workshop
14 November 2016
Trondheim, Norway

With thanks to
Stig Mjølsnes and
Britta Hale for
inviting me and
arranging my visit!

1 / 35

Did I believe an essay could change the values or work of a community?

I hoped so!

1. Encryption works, for concealing content,
2. and even has a natural democratizing *tendency*.
3. Cryptographers and developers are smart,
4. and their work *can* be relevant.
5. Metadata concealment is possible, and is already done (in Tor).
6. End-to-end and device encryption have become popular.
7. Open-source can improve assurance.
8. Some cryptographers *are* interested in privacy and societal issues
9. and in the political setting of our collective work.
10. We can rebalance to put greater emphasis on "crypto-for-privacy"
    (as opposed to crypto-for-security, crypto-for-crypto, or crypto-for-power).
11. And if our problems are "just" cultural, culture can and does change.

Did I believe an essay could change the values or work of a community?

I hoped so! But it didn't seem likely.

1. Culture doesn't change just because someone wills it to.
2. The vast majority of the cryptographic community cares about other things.
3. Architecture can make cryptography support the powerful *or* the powerless.
4. Endpoints are insecure. Code is buggy.
5. Security is a "weak-link" property, and cryptography is rarely that link.
6. Privacy measures can increase complexity and reduce usability.
7. Little moral or societal concern among computer scientists quite broadly.
8. Economic incentives all wrong: enormous value gained by mining information flows. That value flows to both corporations and governments.
9. Privacy instruments are weak, but surveillance instruments are strong.
10. Intelligence agencies have huge budgets & operate beyond reach of the law.
11. Open source is no panacea.
12. Monitoring physical spaces can obviate cryptography.
13. Plenty of information in the metadata, and concealing metadata is hard.
14. General-purpose computing is in retreat (appification, Apple Store, …).
15. Successful (mis-)framing by governments.
16. Corporatism and public-private "partnerships" have never been stronger.

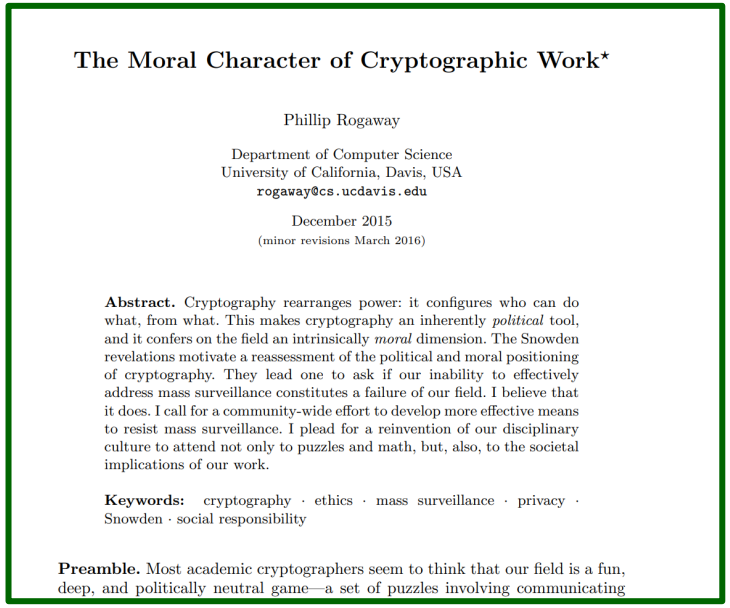
How did this work out?

- Plenty of invited talk, even media coverage.
- Lots of friendly emails, which continue to this day.
- Essay is pretty widely read in university classes.
- No open disagreement within the crypto community of the essay's central claims.
- Yet … *No recognizable change within the field …* and that really *was* the purpose.
- Grade: **B -**

Putting things in perspective,

The Moral Character of Cryptographic Work (2015)

was **one** attempt to nudge the values and work of **colleagues.**



The Moral Character of Cryptographic Work*

Phillip Rogaway

Department of Computer Science
University of California, Davis, USA
rogaway@cs.ucdavis.edu

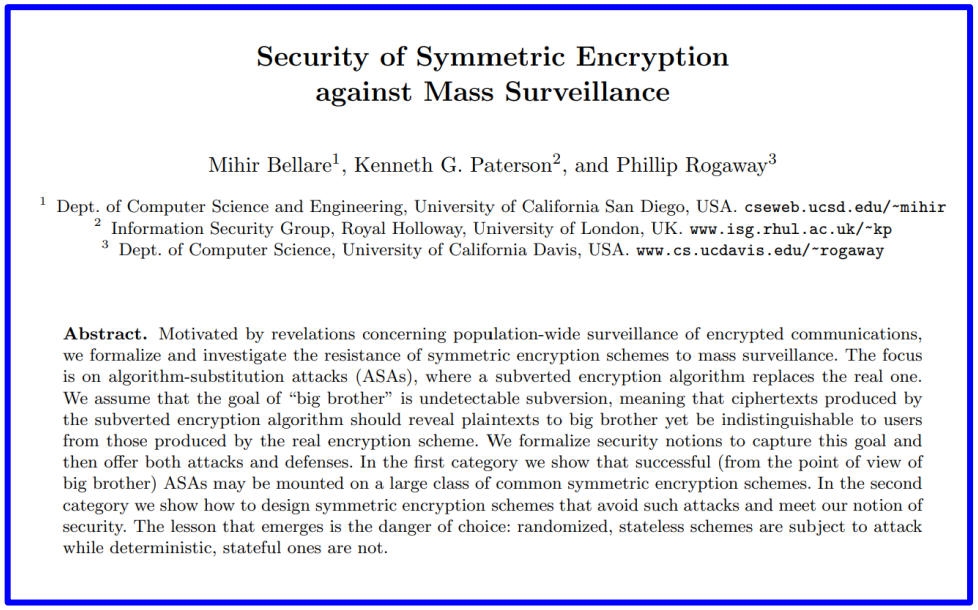
December 2015
(minor revisions March 2016)

Abstract. Cryptography rearranges power: it configures who can do what, from what. This makes cryptography an inherently *political* tool, and it confers on the field an intrinsically *moral* dimension. The Snowden revelations motivate a reassessment of the political and moral positioning of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means to resist mass surveillance. I plead for a reinvention of our disciplinary culture to attend not only to puzzles and math, but, also, to the societal implications of our work.

Keywords: cryptography · ethics · mass surveillance · privacy · Snowden · social responsibility

Preamble. Most academic cryptographers seem to think that our field is a fun, deep, and politically neutral game—a set of puzzles involving communicating

So were most of my
post-Snowden technical papers (2013-2022)



Security of Symmetric Encryption
against Mass Surveillance

Mihir Bellare[1], Kenneth G. Paterson[2], and Phillip Rogaway[3]

[1] Dept. of Computer Science and Engineering, University of California San Diego, USA. cseweb.ucsd.edu/~mihir
[2] Information Security Group, Royal Holloway, University of London, UK. www.isg.rhul.ac.uk/~kp
[3] Dept. of Computer Science, University of California Davis, USA. www.cs.ucdavis.edu/~rogaway

Abstract. Motivated by revelations concerning population-wide surveillance of encrypted communications, we formalize and investigate the resistance of symmetric encryption schemes to mass surveillance. The focus is on algorithm-substitution attacks (ASAs), where a subverted encryption algorithm replaces the real one. We assume that the goal of "big brother" is undetectable subversion, meaning that ciphertexts produced by the subverted encryption algorithm should reveal plaintexts to big brother yet be indistinguishable to users from those produced by the real encryption scheme. We formalize security notions to capture this goal and then offer both attacks and defenses. In the first category we show that successful (from the point of view of big brother) ASAs may be mounted on a large class of common symmetric encryption schemes. In the second category we show how to design symmetric encryption schemes that avoid such attacks and meet our notion of security. The lesson that emerges is the danger of choice: randomized, stateless schemes are subject to attack while deterministic, stateful ones are not.

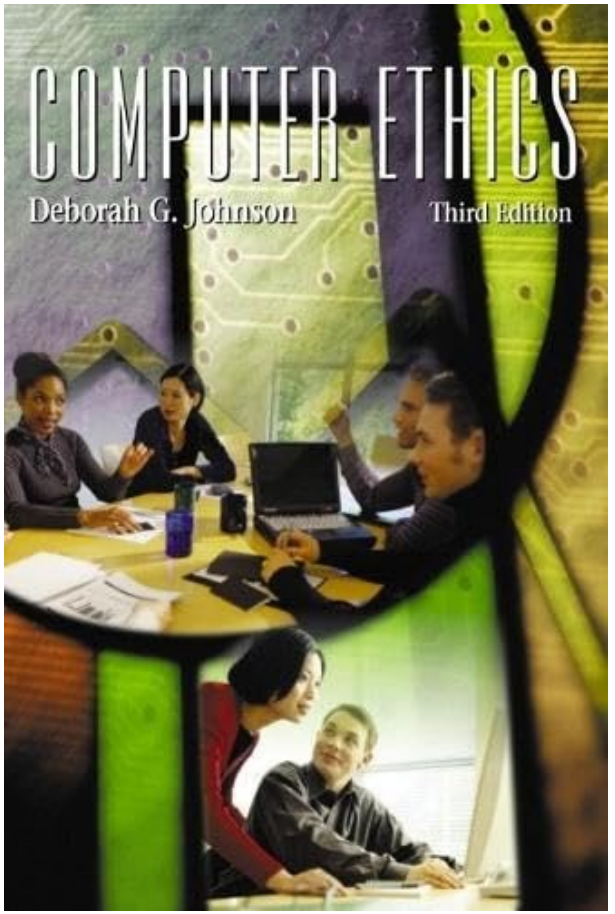
And I was *also* engaged in trying to nudge the values and future work of our **students —** mostly by teaching an **ethics class** (2004 – 2024)

I never publicly spoke about this before …

despite it being an obsession of mine for 20 years.

Our ABET-rooted answer to ethics education: a single course, ECS188, *Computer Ethics*

Before 2004: Boring, stupid, narrow, patronizing, conventional



1: Introduction
2: Philosophical Ethics
3: Professional Ethics
4: Ethics and the Internet I
5: Privacy
6: Property
7: Accountability
8: Ethics and the Internet II

The students thought the class was stupid

BLU-82 bomb used on Afghanistan

BLU-82 mushroom cloud

The US was busy trying to bomb Afghanistan back into the Stone Age …

CS students were happy to help out, filling jobs at Lockheed Martin, Raytheon, etc.

Yet their ethics class was talking about why they shouldn't "pirate" Microsoft Word?

Seriously?

My interest was beyond "neighbor ethics", to who gets what power, and the eschatological capacity of modern technology

"Modern technology has introduced actions of such novel scale… that the framework of former ethics can no longer contain them. … To be sure, the old prescriptions of the **"neighbor" ethics**—of justice, charity, honesty, and so on—still hold in their intimate immediacy for the nearest, day-by-day sphere of human interaction. But this sphere is overshadowed by a growing realm of **collective action** …

"An imperative responding to the new type of human action and addressed to the new type of agency … might run …: "**Act so that the effects of your action are compatible with the permanence of genuine human life**"

Hans Jonas
The Imperative of Responsibility
(1979/1984)

I also felt that the students had been lied to, indoctrinated with the "Standard Technological Narrative" (STN)

1. Technology is just a **tool**. It is **apolitical** and **ethically neutral**.
2. Due to technology, things are **great** and **getting better**.
3. Better technology can and will **fix** what inferior technology broke.
4. We will overcome the **climate/environmental** challenges.
5. Tech is driven by brilliant **individuals**, advanced by the **marketplace**.
6. We are on a journey to create a **technological utopia**.

My problem with the STN:

It's just a story

Problems with my teaching an ethics class

1. **No academically relevant background.** A science nerd. No academic training in philosophy, sociology, psychology, history, anthropology, political science, STS, …

2. **Felt that I needed an enormously wide background.** Didn't know where to even start to get it. Also felt I needed a background far *beyond* scholarship.

3. **No interest in ethics stripped from the imperative of acting well.**

> *The task of philosophical ethics is to explore what human beings ought to do, or more accurately, to evaluate the arguments, reasons, and theories that are proffered to justify accounts of morality.* D. Johnson, 2001

4. **Skeptical that ethics <u>can</u> be taught** (in an operationally meaningful way). My guess: taking a conventional class in ethics won't make anyone a better person. It might make them worse.

Skepticism about the teachability of ethics goes way back

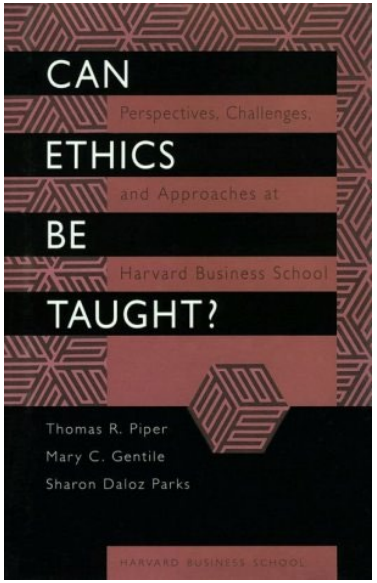
If virtue is knowledge, then it must be teachable. But we have found no teachers of virtue, and in fact those who are said to teach it do not agree among themselves… It appears that virtue arises in us not by teaching but by divine allotment

Plato, *Meno*, ~380 BCE

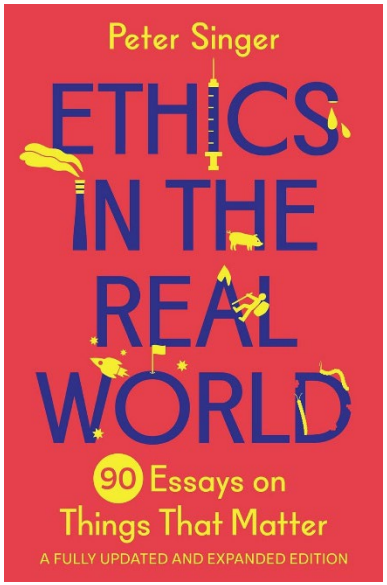
We become just by doing just actions, temperate by doing temperate actions, brave by doing brave actions… The virtues we get by first exercising them. … Words alone are powerless to make a man good.

Aristotle, *Nicomachean Ethics,* ~340 BCE

Modern scholarship on this is thin



1993



2017/ 2023

Where does ethics come from, anyway?

It's psychological basis / naturalistic origin

1. **Nature**: Morality is *built in.* It's in our genes. Cross-cultural differences are minor; fundamental principles are universal. *Nativist*

2. **Nurture**: Morality is a *cultural artifact*. Children are blank-slates that can go in almost direction. Morality varies in profound ways across time and place. *Empiricist*

3. **Rationalism**: Morality is rooted in reasoning. People *figure out* what is right. Careful reasoning is *how* you come to figure out what is / isn't right. *Rationalist*

Western moral philosophy has been dominated by rationalists

Utilitarianism: *It is the greatest happiness of the greatest number that is the measure of right and wrong.*

Jeremy Bentham, *A Fragment of Government* (1776)

Deontology: *Act only according to that maxim whereby you can at the same time will that it should become a universal law.*

Immanuel Kant, *Groundwork of the Metaphysics of Morals* (1785)

Utilitarian arithmetic: the felicific calculus in modern mathematical notation!

- Let $\mathcal{A}$ be the set of actions currently available to you.
- Each $a \in \mathcal{A}$ will result in a new world $W_a(0)$.
- It will evolve in time, progressing through worlds $W_a(t)$ for all $t \in [0, \infty)$.
- In each world $W_a(t)$ there will be a set of living people, $P(W_a(t))$.
- For each $p \in P(W_a(t))$, let $H_a(p, t)$ be how happy p is at time t in world $W_a(t)$, measured in *hedons*.
- We should probably regard $W_a(t)$, and therefore $P(W_a(t))$ and $H_a(p, t)$, as random variables, acknowledging the uncertainty about how the world changes, who will be alive in it, and how happy they will be.

The ethically correct action for you to take is the action $a \in \mathcal{A}$ that maximizes the *felicific utility* of action a, defined as

$$F(a) = \int_0^\infty \sum_{p \in P(W_a(t))} \mathbb{E}[H_a(p, t)]\, dt.$$

Utilitarian arithmetic: the felicific calculus in modern mathematical notation!

It assumes that happiness is the only intrinsic good. It isn't.

It assumes that people are the only thing that matters in the world. We aren't.

ns currently a

a new world

- It will evolve in time, progressing through worlds $W_a(t)$ for all $t \in [0, \infty)$.
- In each world $W_a(t)$ there will be a set of living people, $P(W_a(t))$.
- For each $p \in P(W_a(t))$, let $H_a(p,t)$ be how happy p is at time t in world $W_a(t)$, measured in *hedons*.
- We should probably regard $W_a(t)$, and therefore $P(W_a(t))$ and $H_a(p,t)$, as random variables, acknowledging the uncertainty about how the world changes, who will be alive in it, and how happy they will be.

This is ridiculous stupid

The ethically correct action for you to take is the action

It assumes more is better. Really?

It assumes we can anticipate the long-term consequences of our actions. We can't.

$$\mathbb{E}[H_a(p,t)]\,dt.$$

$$p \in P(W_a(t))$$

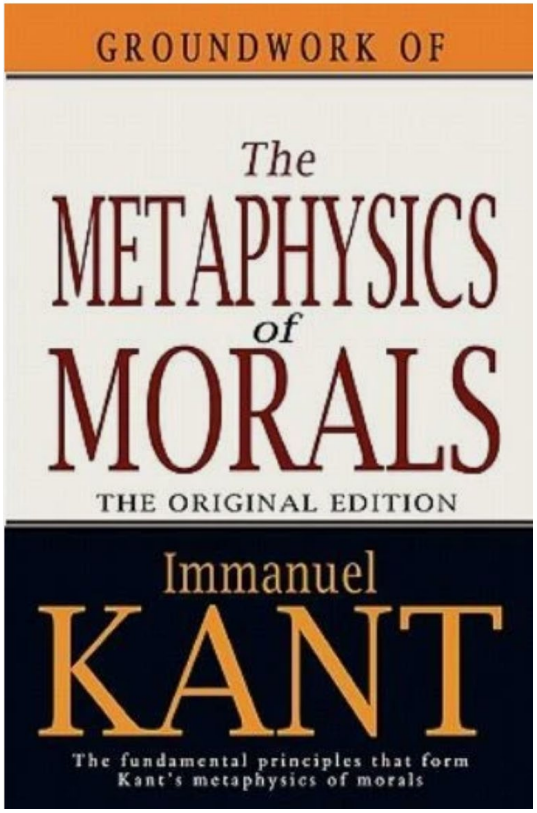
Worse than stupid: to Si-valley engineers, utilitarianism devolves into TESCREAL

Transhumanism
Extropianism
Singularitarianism
Cosmism
Rationalism [the Internet community]
Effective **A**ltruism
Longtermism

Timnit Gebru and Émile P. Torres, *The TESCREAL bundle: eugenics and the promise of utopia through artificial intelligence.*

> [I]f humanity colonizes the universe and creates planet-sized computers to run virtual-reality worlds populated by digital people, the future posthuman population could be enormous. … Bostrom estimates that there could be at least 10^58 such people within the accessible universe … . Why does this matter? Because, from the … perspective of "totalist utilitarianism" … our sole moral obligation is to maximize the total quantity of "value" in the universe. … [If] these 10^58 people in computer simulations were to have net-positive lives on average, the result would be literally "astronomical" amounts of "value" — which would be very "good." Since totalist utilitarianism bases what is morally right on what is good, this view entails that failing to bring these future digital people into existence would be profoundly wrong.
>
> Gebru and Torres, 2023

Kant favored logic over arithmetic, but, in my reading, can't put together a single logical argument



Cat-1: Act only according to that maxim whereby you can at the same time will that it should become a universal law. ⇔

Cat-2: Treat humanity, whether in your own person or in the person of any other, always at the same time as an end and never merely as a means. ⇔

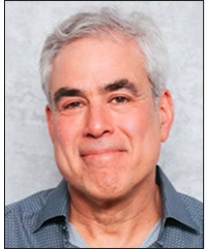
Cat-3: Act as if you were through your maxims always a lawmaking member in the universal kingdom of ends.

Ex-1: Suicide
Ex-2: False promising
Ex-3: Neglecting one's talents
Ex-4: Refusing to help others

By 1785, mathematicians had proven the law of large numbers, Taylor's theorem, Euler's polyhedron formula (v-e+f=2 for all convex polyhedra), hard identities like $\sum 1/n^2 = \pi^2/6$, developed the calculus of variations, … Suggest: **a category error** – treating ethical questions as though they were amenable to arithmetic or logic when they are not.

~~Rationalism: morality is rooted in reasoning~~

Jonathan Haidt
The Righteous Mind (2012)



Decades of evidence tell us that, descriptively, this is just **wrong.** The ***rationalist delusion***.

- Ethical decisions are made *very* quickly & are unaffected by cognitive load
- You can craft puzzles where people *can't* find a reason for their ethical beliefs
- The parts of the brain associated to *emotional* processing, *not* reasoning is what's activated by ethical dilemmas (fMRI studies staring with J. Greene)
- Psychopaths tend to be highly rational – it is feelings that they lack
- Ethical judgments are profoundly influenced by smells, tastes, and other rationally irrelevant aspects of our environment
- Non-human animals seem to behave on motives that I, at least, would regard as ethical motives
- Moral philosophers are *not* more ethical than others [Schwitzgebel and Rust, 2009]
  No more likely to: give to charity, call their mothers, donate blood, donate organs, respond to student emails, return their library books, clean up after themselves at a conference.
- Motivated reasoning can take you anywhere you want to go
  When you want to believe something: *can I believe it.*
  When you don't want to believe something: *must I believe it?*

"Intuition comes first; strategic reasoning second"

Some didn't drink the rationalist Kool-Aid

Reasoning is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them. …. It is not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. It is not contrary to reason for me to choose my total ruin, to avoid the least uneasiness of my little finger.
D. Hume (1739)

The emotional control centers in the hypothalamus and limbic system of the brain… flood our consciousness with all the emotions—hate, love, guilt, fear, and others—that are consulted by ethical philosophers who wish to intuit the standards of good and evil. What, we are then compelled to ask, made the hypothalamus and limbic system? They evolved by natural selection. That simple biological statement must be pursued to explain ethics and ethical philosophers … **E.O. Wilson: *Sociolbiology: the New Synthesis* (1975)**

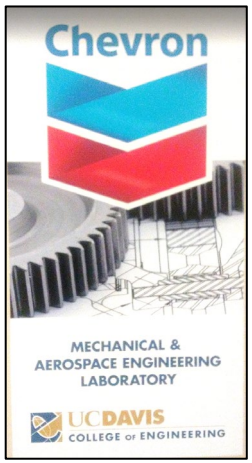
Deontology … is a kind of moral confabulation. We have strong feelings that tell us in clear and uncertain terms that some things simply cannot be done and that other things simply must be done. But it's not obvious how to make sense of these feelings, and so we, with the help of some especially creative philosophers, make up a rationally appealing story: There are these things called "rights" which people have, and when someone has a right you can't do anything that would take it away. **J. Greene: *The Secret Joke of Kant's Soul* (2008)**

Social-Intuitionist Model

A worked example

J. Haidt: *The emotional dog and its rational tail* (2001)

Triggering Event
Seeing a Chevron recruiting poster. Should I interview?



A's intuition
(a positive affective flash)

A's judgment
(Yes, let's interview!)

A's reasoning
(reasons for)
(+reasons against)

B's intuition
(a negative affective flash)

B's judgment
(no way)

B's reasoning
(reasons against)

If ethical behavior isn't rooted in rationalism, then ethics can't be taught in a class.

Pure sophistry. We teach all kinds of things in classrooms that aren't rooted in rationalism.

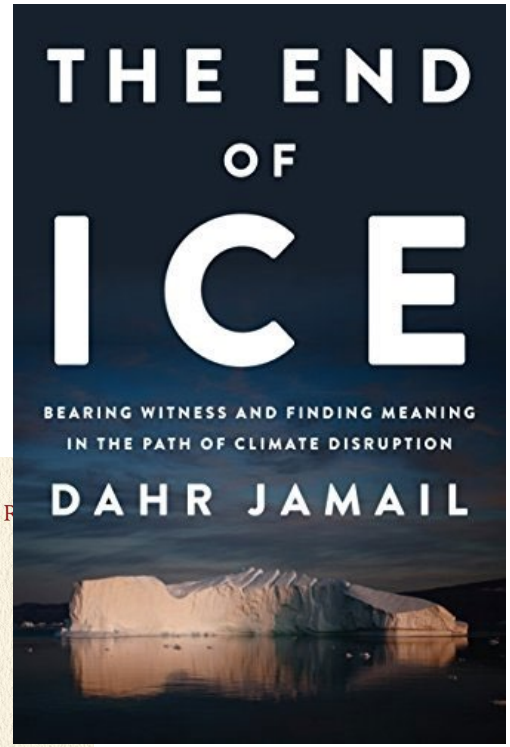
> if our goal is to produce good behavior, not just good thinking, then it's … important to reject rationalism and embrace intuitionism. ***Nobody is ever going to invent an ethics class that makes people behave ethically after they step out of the classroom*.** Classes are for riders [deliberative thinking], and riders are just going to use their new knowledge to serve their elephants [pos-hoc rationalization of whatever the hell they wish to do] more effectively.
>
> Jonathan Haidt: *The Righteous Mind*, 2012

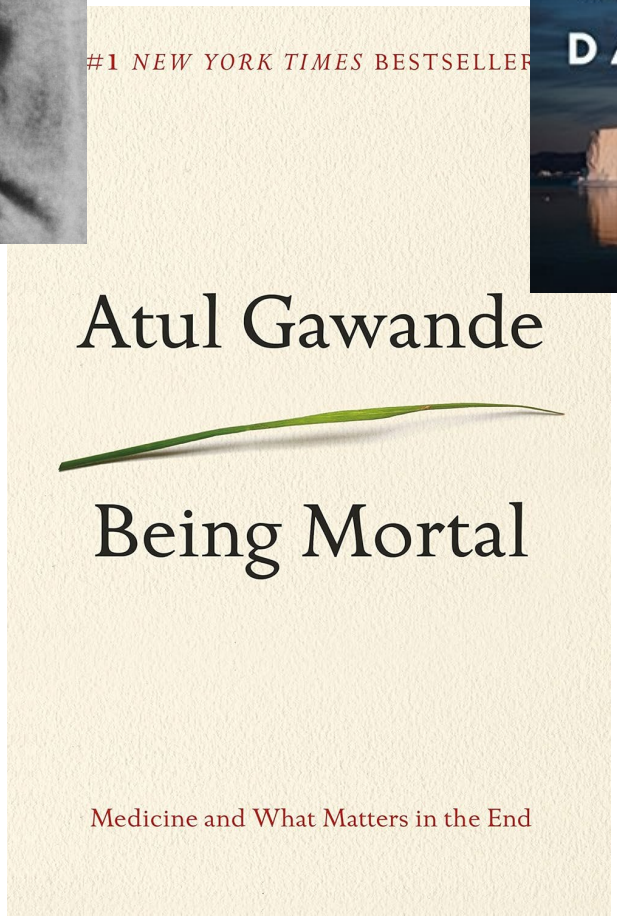
How I've run class

1. Get the student to *feel*
2. Use films. Open with one
3. Careful classroom mechanics (don't lecture, sit in a circle, etc.)
4. No tech in the classroom
5. Safe space, rebooted
6. Chatham House Rule
7. Make it personal (excursions, baked goods, etc.)
8. Select diverse and difficult material that challenges the STN
9. Avoid almost all moral philosophy
10. Avoid unrealistic ethical puzzles
11. Avoid anything that gamifies ethics
12. Encourage artistic expression
13. Practice "radical honesty"
14. Evolve

Get students to <u>feel</u>

Testament (1983)

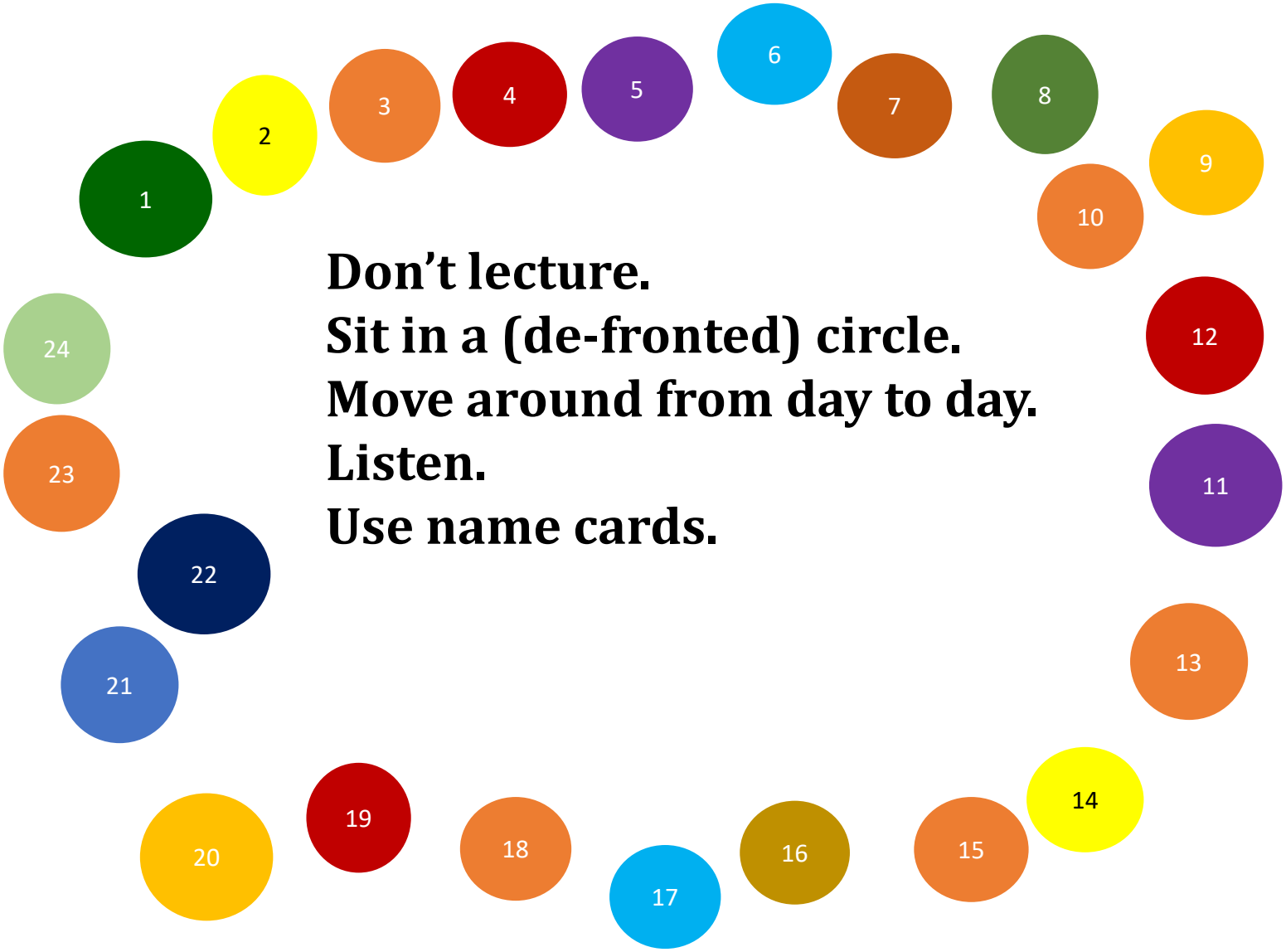




Use films.
Open with one.

Dekalog 1 (1988), Krzysztof Kieślowski

Classroom mechanics



Don't lecture.
Sit in a (de-fronted) circle.
Move around from day to day.
Listen.
Use name cards.

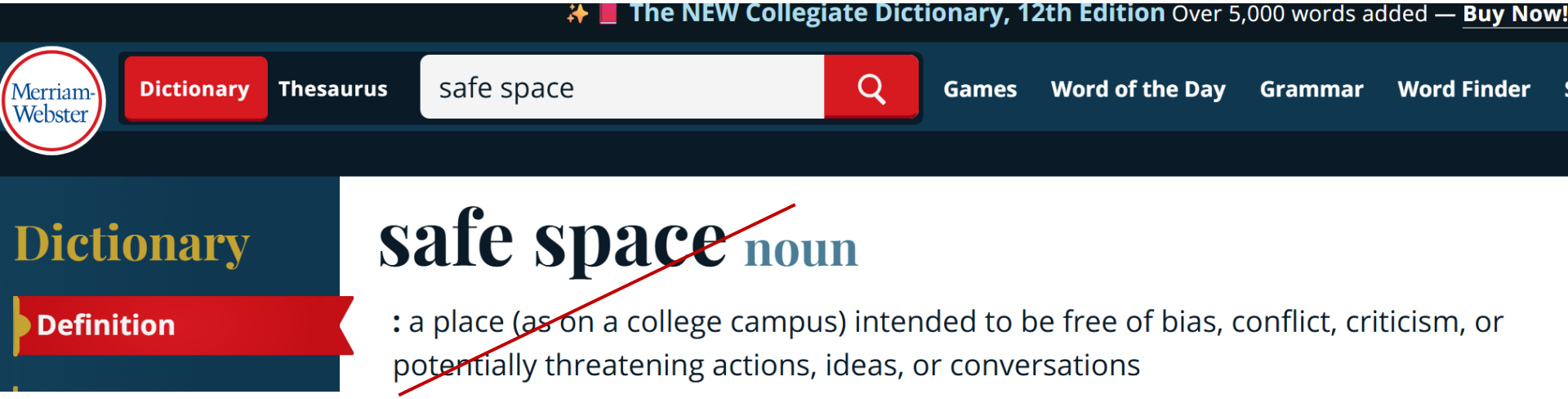
No tech in the classroom

Dear ECS 188 Students:

Mobile phones must be turned off or silenced and placed in your bag before entering this classroom

Violators will be fed to the Komodo dragons in the basement of Kemper. Undigested bits will be fed to OSSJA.

Safe space, rebooted



Not what is intended.

I intend a space where students feel free to speak their minds, unfiltered. To tell their stories. Potentially a space **full** of "bias, conflict, criticism" and uncomfortable ideas and conversations.

Chatham House Rule

When a meeting, or part thereof, is held under the **Chatham House Rule**, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed. rev. 2001 (1927, 1992)

Leaving Las Vegas (1995)

Narrower than the Las Vegas slogan (2003)

"What happens here, stays here."

Make it Personal

e.g.,

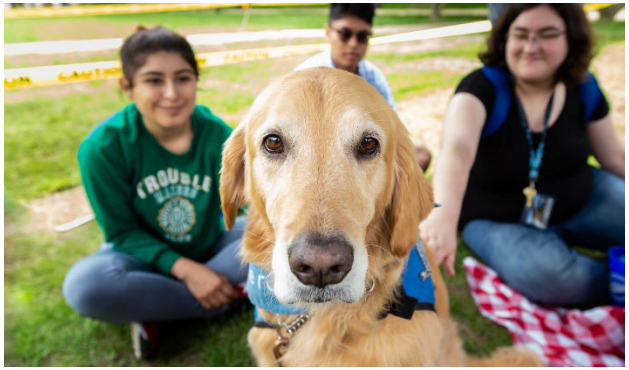
Bake vegan treats

Invite students to the mountains,

a climbing gym,

or your home

Dog day



Start with some apropos music

Select diverse and difficult material that challenge the STN

- Whole books, book chapters, articles, podcasts, newspapers, magazines, films, videos, ...
- Old & new
- Fiction & nonfiction
- Scholarly & non-scholarly
- Stuff you disagree with
- Manifesto of a murderer
- Political writing
- Religious writing
- Things that make you cry
- Things that are hard to look at
- No attempt at "balance"

Will it make the students *feel* anything?

Forward

A Brief Note to the Student (Phillip Rogaway, 2018)
Basic attitudes towards technology (Survey)

Reader,
p. 1 of 6

Technology

The Machine Stops (E. M. Forster, 1909)
Dekalog I (1989) (Krzysztof Kieslowski) (51 mins)
Industrial Society and Its Future (Ted Kaczynski, 1995)
Why America Failed: The Roots of Imperial Decline (Morris Berman, 2011)
Sapiens (Yuval Noah Harari, 2014)
Marshal McLuhan Interview (1969)
Views of Technology (Ian Barbour, 1993).
Do Artifacts have Politics? (Langdon Winner, 1984)
Do Politics have Artefacts? (Bernward Joerges, 1999)
Do Machines Make History? (Robert L Heilbroner, 1967)
Five Things We Need to Know About Technological Change (Neil Postman, 1998)

Distraction

The Murderer (Ray Bradbury, 1953)
Have Smartphones Destroyed a Generation? (H Twenge, 2017)
How Technology is Hijacking Your Mind (Tristan Harris, 2016)
'Irresistible' By Design (Fresh Air, 2017, 30 mins)
The Value of Deep Work (Shankar Vedantam, 2017, 37 mins)
Center for Humane Technology (website, Tristan Harris)
Empowering Design (Joe Edelman, 2015, 19 mins)

Capitalism

The Corporation (2003) (Mark Archbar, Jennifer Abbott, Joel Bakan; 145 mins)
Manufacturing Consent (Noam Chomsky, 1992)
The Lexus and the Olive Tree (Thomas Friedman, 1999)
The Lexus and the Olive Tree Revisited (Ha-Joon Chang, 2008)
An Economic Hit Man Confesses and Calls to Action (John Perkins, 2016, 19 mins)
A Road Map for Natural Capitalism (Lovins, Lovins, Hawken, 1999)
Chap-1, Chap-4, Chap-6, and Chap-7 from *Making Globalization Work* (Joseph Stiglitz, 2007)
Bhopal Lives (Suketu Mehta, 1996). CNN photos

War

War (original URL) (Brian Orend, 2005)
Farewell Address to the Nation (Dwight D. Eisenhower, 1961)
Why We Fight (2005) (Eugene Jarecki, 98 mins)
Angry Eyes: The "God Trick" and the geography of militarized vision (D. Gregory, 2016, 48 mins)

Nuclear Threats

Existential Risk Prevention as Global Priority (Nick Bostrom, 2013)
Arkhipov Day: Celebrate the Man Who Saved the World (Milan Rai, 2014)
Testament (1983) (Lynne Littman, 90 mins)
Threads (1984) (Mick Jackson, 112 mins)

Environmental Threats

Ishamel (Daniel Quinn, 1992)
This Civilisation is Finished (Rupert Read and Samuel Alexander, 2019)
The Uninhabitable Earth (David Wallace-Wells, 2017)
The End of Ice (Dahr Jamail, 2019)
The Land Ethic (Aldo Leopold, 1949).
The Conceptual Foundations of the Land Ethic (J. Baird Callicott, 1989)
Oldest Living Tree Tells All (Michael P. Cohen, 2004)
The Tragedy of the Commons (Garrett Hardin, 1968)
What Lies Beneath: The Underestimate of Existential Climate Risk (D. Spratt and I. Dunlop, 2018)
Deep Adaptation: A Map for Navigating Climate Tragedy (Jem Bendell, 2018)
Trajectories of the Earth System in the Anthropocene (Will Steffen *et al.*, 2018)
Chasing Coral (2017) (Jeff Orlowski, 87 mins)
Time to Choose (2015) (Charles Ferguson, 97 mins)
Green Illusions (Ozzie Zehner, 2012, 63 min)

More Ways to Die

Why the Future Doesn't Need Us (Bill Joy, 2000)
Promise and Peril (Ray Kurzweil, 2001)

Philosophical Ethics

Philosophical Ethics (Deborah Johnson, 2001)
The Question Concerning Technology (Martin Heidegger, 1954/1977)
The Altered Nature of Human Action (Hans Jonas, 1985)
Technological Subversion (David Strong, 1995)
The Superior Human? (2014) (Jenia Meng, 73 mins)

Food

Moving Beyond Fast Food Nation (Peter Singer and Eric Schlosser, 2006, 89 mins)
Interview with Michael Pollan (Marc Eisen, 2008)
Earthlings (2005) (Shaun Monson, 95 mins)
Vegucated (2011) (Marisa Miller Wolfson, 76 mins)

Reader, p. 4 of 6

Personal Well-Being

Technology and Happiness (James Surowiecki, 2005).
Happiness: has social science a clue? (Richard Layard, 2003)
World Happiness Report 2019 (U.N.; editors J. Helliwell, R. Layard, and J. Sachs)
Could a Green New Deal Make Us Happier People? (Kate Arnoff, 2019)
Higher Social Class Predicts Increased Unethical Behavior (Stancato et al., 2012)
How We See Ourselves and How We See Others (Emily Pronin, 2008)
Can Psychology be Taught? (Daniel Kahneman, 2011)

Dying and Death

Being Mortal (Atul Gawande, 2014)
Why I hope to Die at 75 (2014) (E. Emanuel)

Discrimination and Bias

The Student as Nigger (Jerry Farber, 1967)
13th (2016) (Ava DuVernay)
Welcome to Your Authentic Indian Experience™ (Rebecca Roanhorse, 2017)
The 'Thumbprint of the Culture': Implicit Bias and Police Shootings (Hidden Brain, S Vedantam, 2019)
Project Implicit Bias
Google's Ideological Echo Chamber (James Damore, 2017)
Why is Silicon Valley so Awful to Women? (Lizy Mundy, 2017)
Unlocking the Clubhouse: Women and Computing (scan3, scan6) (Margolis and Fisher, 2002)

Surveillance, Machine Learning, and AI

Reader, p. 5 of 6

Snowden revelations:
Verizon (2013.06.2013), PRISM (2013.07.07),
The NSA Revelations All in One Chart (2014.06.30),
How spy agencies defeat security mechanisms (2013.09.06). Videos: Poitras 1 (2013, 13 mins)
Poitras 2 (2013, 7 mins), How we take back the internet (Edward Snowden, 2014, 80 mins),
Administration response on telephone metadata collection (2013.08.09), PCLOB Report on PRISM/Upstream surveillance, TED2014 talk by Richard Ledgett.
The Lives of Others (2006) (Florian Henckel von Donnersmarck, 137 mins)
The Fate of Internet (John Naughton, 2013)
Taking back the Internet (Bruce Schneier, 2013)
The Moral Character of Cryptographic Work (Phillip Rogaway, 2015)
Information Consumerism (Evgeny Morozov, 2013)
Big other: surveillance capitalism and the prospects of an information civilization (S Zuboff, 2015)
Automated Inference on Criminality using Face Images (Xiaolin Wu and Xi Zhang, 2016)
The Paranoid State (Sarah Kendzior, 2013)
The Transparent Society (David Brin, 1996)
Against Transparency (Lawrence Lessig, 2013)
Cambridge Analytica - The Power of Big Data and Psychographics (Alexander Nix, 2016, 11 mins)
How China is Using Artificial Intelligence in Classrooms (Wang, Hong, and Tai, 2019/2021)
Haunted by Data (Maciej Ceglowski, 2015)
The Data That Turned the World Upside Down (Hannes Grassegger and Mikael Krogerus, 2017)
Facebook's Role in Brexit—and the threat to democracy (Carol Cadwalladr, 2019)
Bodies in Seats (Casey Newton, 2019) and The Trauma Floor (Casey Newton, 2019)
The Social Dilemma (94 mins) (Jeff Orlowski, 2020)
Haunted by Data (Maciej Ceglowski, 2015, 20 mins)

Intellectual Property

The Public Domain: Enclosing the Commons of the Mind (James Boyle, 2008)
The GNU Manifesto (original URL) (Richard Stallman, 1985)
Microsoft Research DRM Talk (Cory Doctorow, 2004)

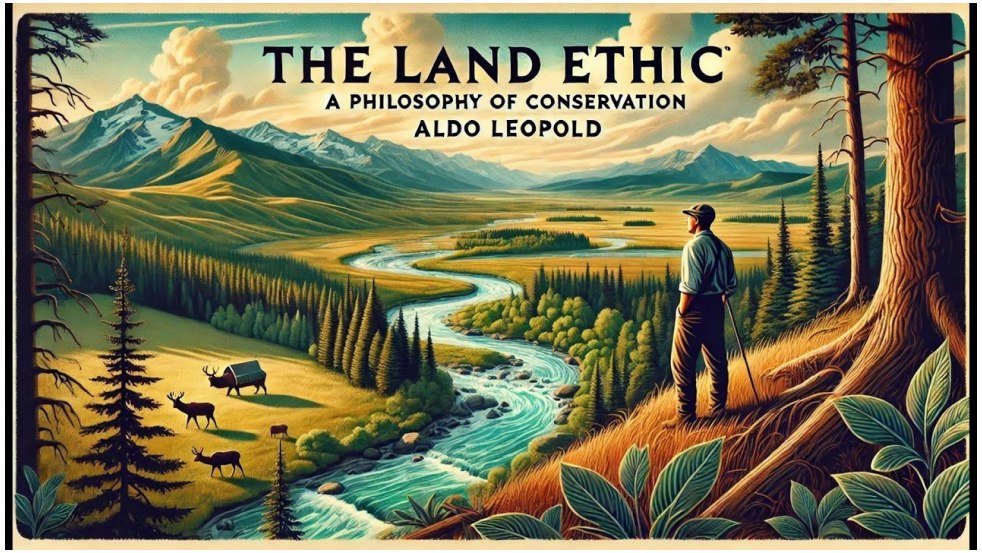
The Imperative to Act

Disciplined Minds (Jeff Schmidt, 2001)
The Disadvantages of an Elite Education (William Deresiewicz, 2008)
Computers, Ethics, and Collective Violence (scan) (Craig Summers, Eric Markusen, 1992)
The Fifty-Nine-Story Crisis (Joe Morgenstern, 1995)
Google Employees Resign in Protest Against Pentagon Contract (Kate Conger, 2018)
An open letter (Norbert Wiener, 1947)
neveragain.tech pledge
Russell-Einstein Manifesto (1955) and Bethe to Clinton (1985)
Fatal Dose: Radiation Deaths linked to AECL Computer Errors (Barbara Wade Rose, 1994)
Dirty Money S1:E1: Hard NOx (Alex Gibney, 2018) (75 mins)
Pinto Fires and Personal Ethics (Dennis Gioia, 1992)
ACM Code of Ethics and Professional Conduct (1992) , IEEE Code of Ethics (2006) , and Software Engineering Code of Ethics and Professional Practices (1999)
Scenarios (drawn from Sara Baase, 2004)
The Two Cultures (C. P. Snow, 1959)
An Interview with Ralph Nader (Pete Davis, 2019)
'You Did Not Act in Time': Greta Thunberg's Full Speech to MPs (Greta Thunberg, 2019)
Rachel Carson (2017) (113 mins)
This is Silicon Valley: I feel myself becoming part of the machine (Glori Liou, 2019)

Concluding Remarks

Some pledges
Video clips from The Ascent of Man (Jacob Bronowski, 1973)]

Avoid almost all moral philosophy



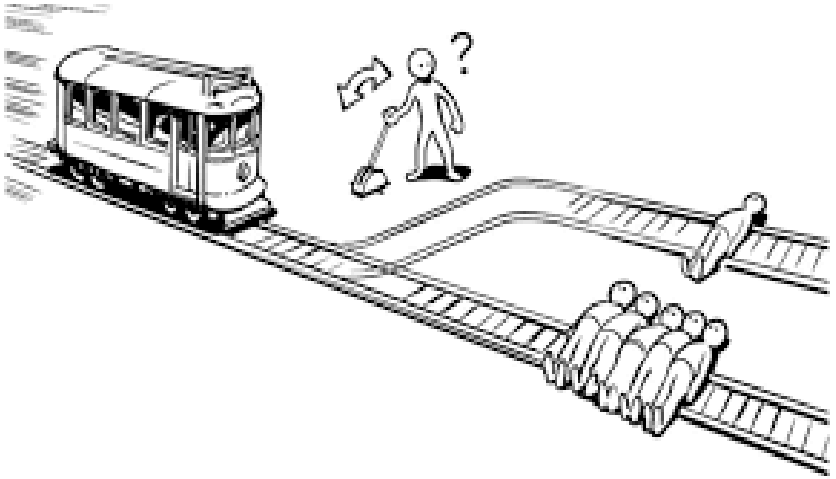
"An ethic, ecologically, is a limitation on freedom of action in the struggle for existence. An ethic, philosophically is a differentiation of social from anti-social conduct. These are two definitions of one thing. The thing has its origin in the tendency of interdependent individuals or groups to evolve modes of co-operation."

But not all

"A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise."

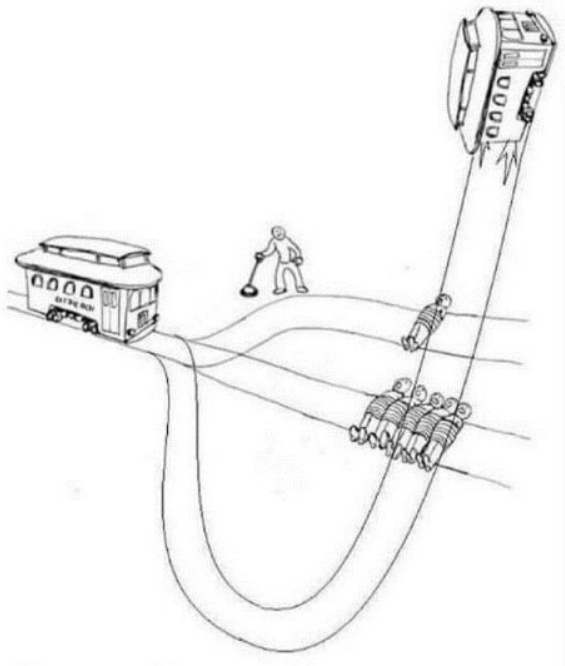
Avoid unrealistic ethical puzzles

The classical Trolly Problem



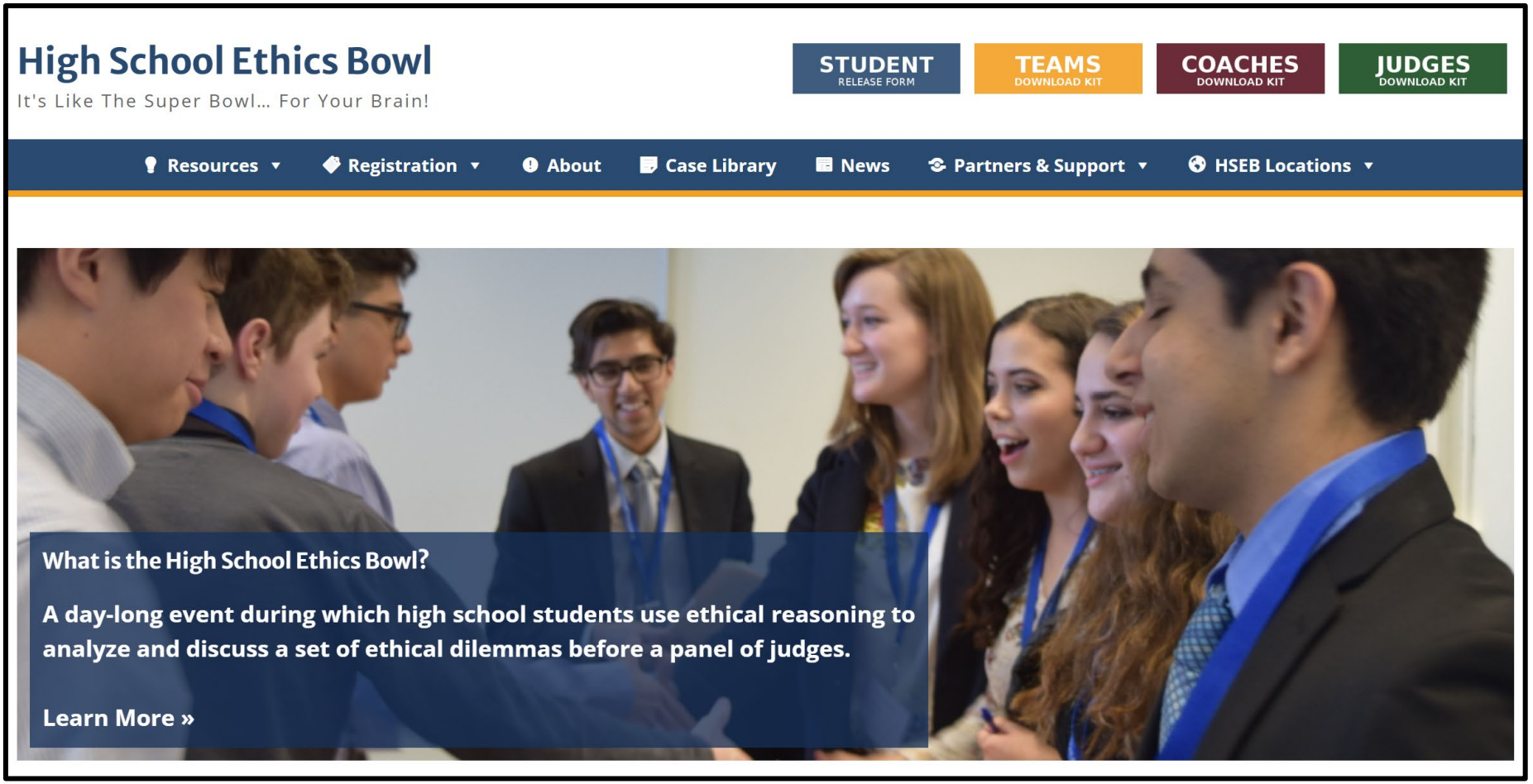
- Do nothing and <u>five</u> die, or
- Pull a lever and <u>one</u> dies.

The Billionaire's Trolly Problem

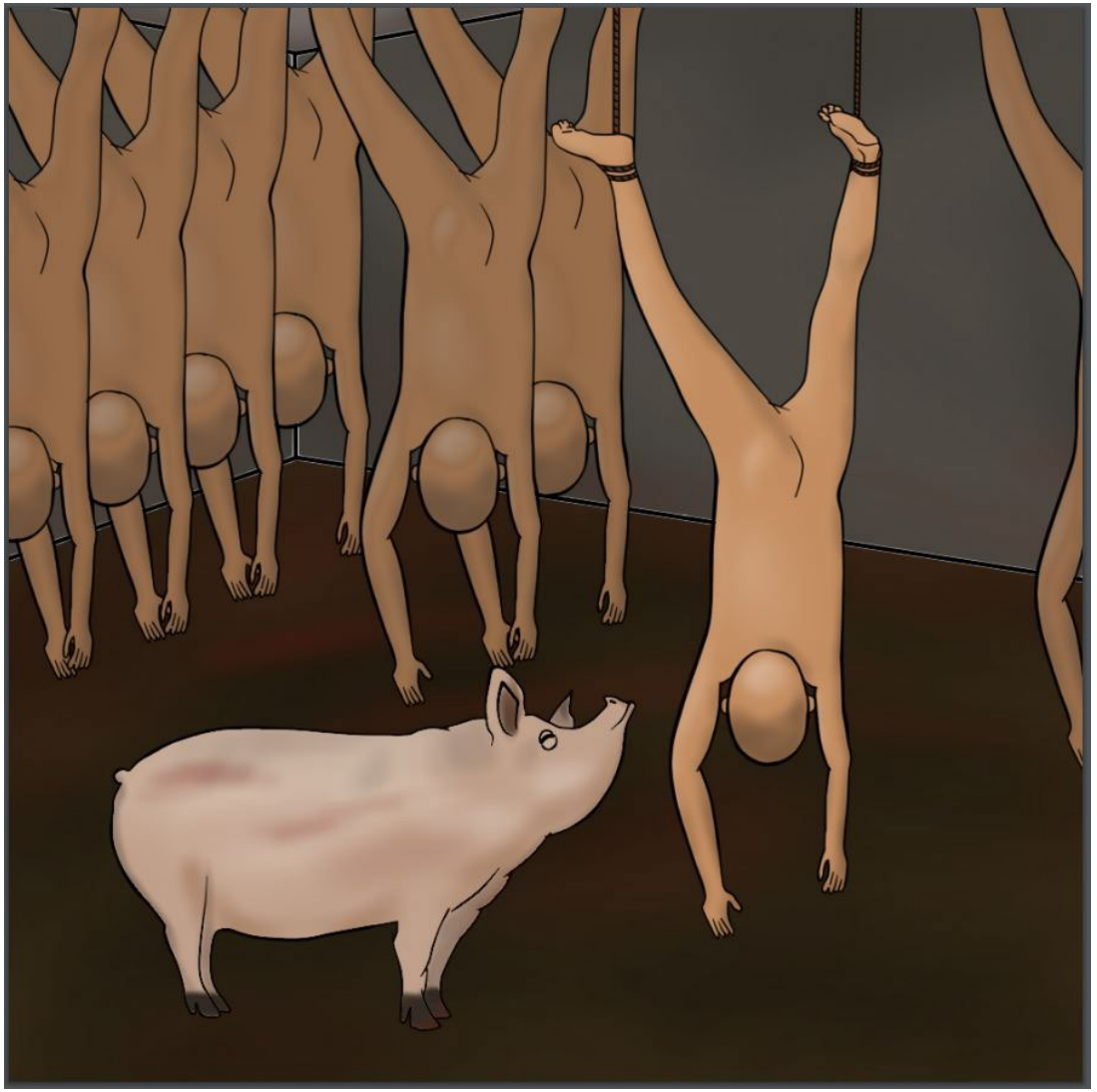


- Do nothing and <u>five</u> die, or
- Pull a lever and <u>one</u> dies, or
- Run over them all and become the first Trolley in Space!

Avoid anything that gamifies ethics, like holding debates



Encourage artistic expression



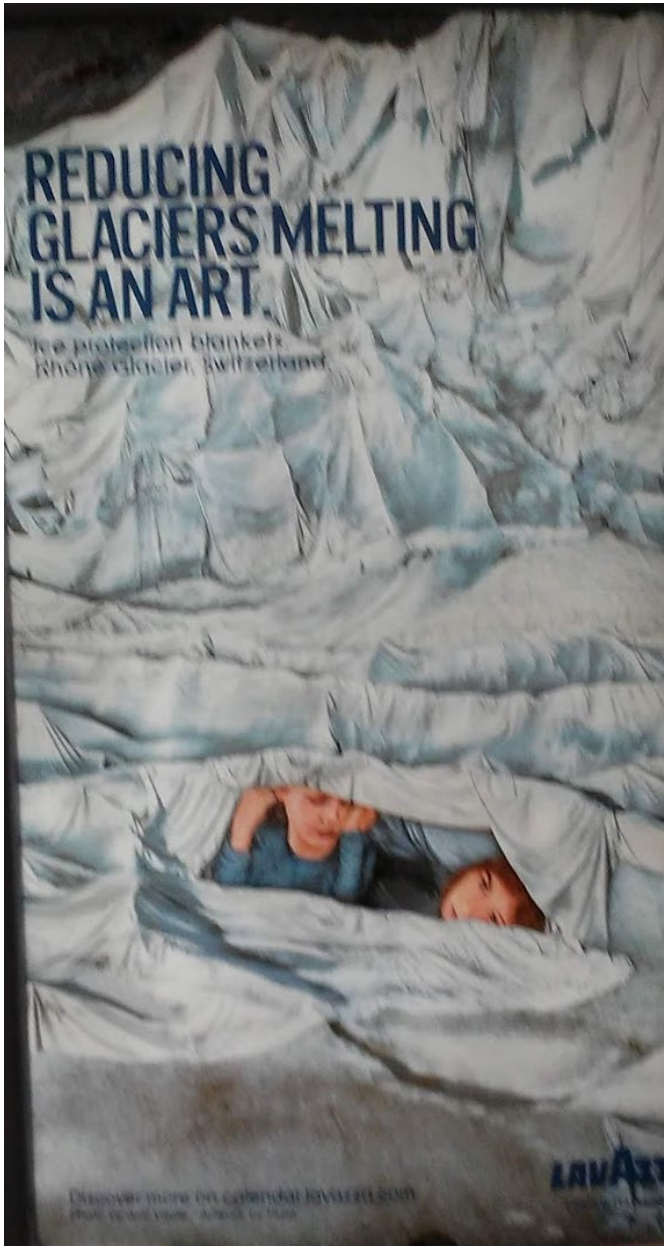
Student artwork, anonymous, 2019

Encourage artistic expression

Student artwork, anonymous, 2019

Practice “radical honesty”

End-of-term goodbye note



“I spent quite a bit of this term thinking about climate change. As though the world knew this, when I got off the plane in Frankfurt a few weeks ago the very first thing I saw was a drawing of a glacier (the Rhône glacier of Switzerland) covered in white blankets. Two children are peering out from beneath the blankets. They are safe, cozy, and secure.

“The poster is a visual lie. It beautifully illustrates how to lie without needing words. We are not safe, not remotely, and children least of all.

“I am frustrated that the greatest wisdom I encountered this term concerning climate change was Dahr Jamail’s conclusion that we should say our goodbyes, grieve, and live our remaining days with dignity. Jamail urges us to connect to the planet as we would a dying friend. He is not wrong.”

Practice “radical honesty”

End-of-term goodbye note

“Besides climate change, the thing most on my mind has been my son. … [He] moves from obsession to obsession, and is currently obsessed with seeds. All over the house are seeds germinating in water glasses, in pots, and between damp paper towels. Banlu wants to know how a plant emerges from a seed. I haven’t the slightest idea. I assume that any honest botanist would say that same. I have to remind myself that it is magical. As we age, [we tend to] lose our sense of awe.”

Practice “radical honesty”

End-of-term goodbye note

“I was moved by some of the student responses on the depression assignment**.** One student told me that his mom committed suicide last year. How can one deal with such a thing? Other students described their own experiences with depression. I think it is OK. Honestly, how can you look upon our world and not, at least sometimes, feel overwhelmed by sadness and fear?

“Here is part of one student’s response to that assignment.

> “The Dictionary of Obscure Sorrows is an internet project created by John Koenig. It seeks to create and add new words to the English language that … can define emotions we humans experience but never termed. In that Dictionary is … *Sonder*. It is defined as “the realization that each random passerby is living a life as vivid and complex as your own—populated with their own ambitions, friends, routines, worries and inherited craziness”. When I am riding the bus on my way back home after a particularly awful day, I look around at the other passengers on the bus and … sonder washes over me…. “

Evolve

ECS 189L: Topics in Computer Science: Black Mirror

MWF 12-1 + W 1-2 in Giedt 1007

CRN 57170 4 units

Prof. Phillip Rogaway rogaway@ucdavis.edu

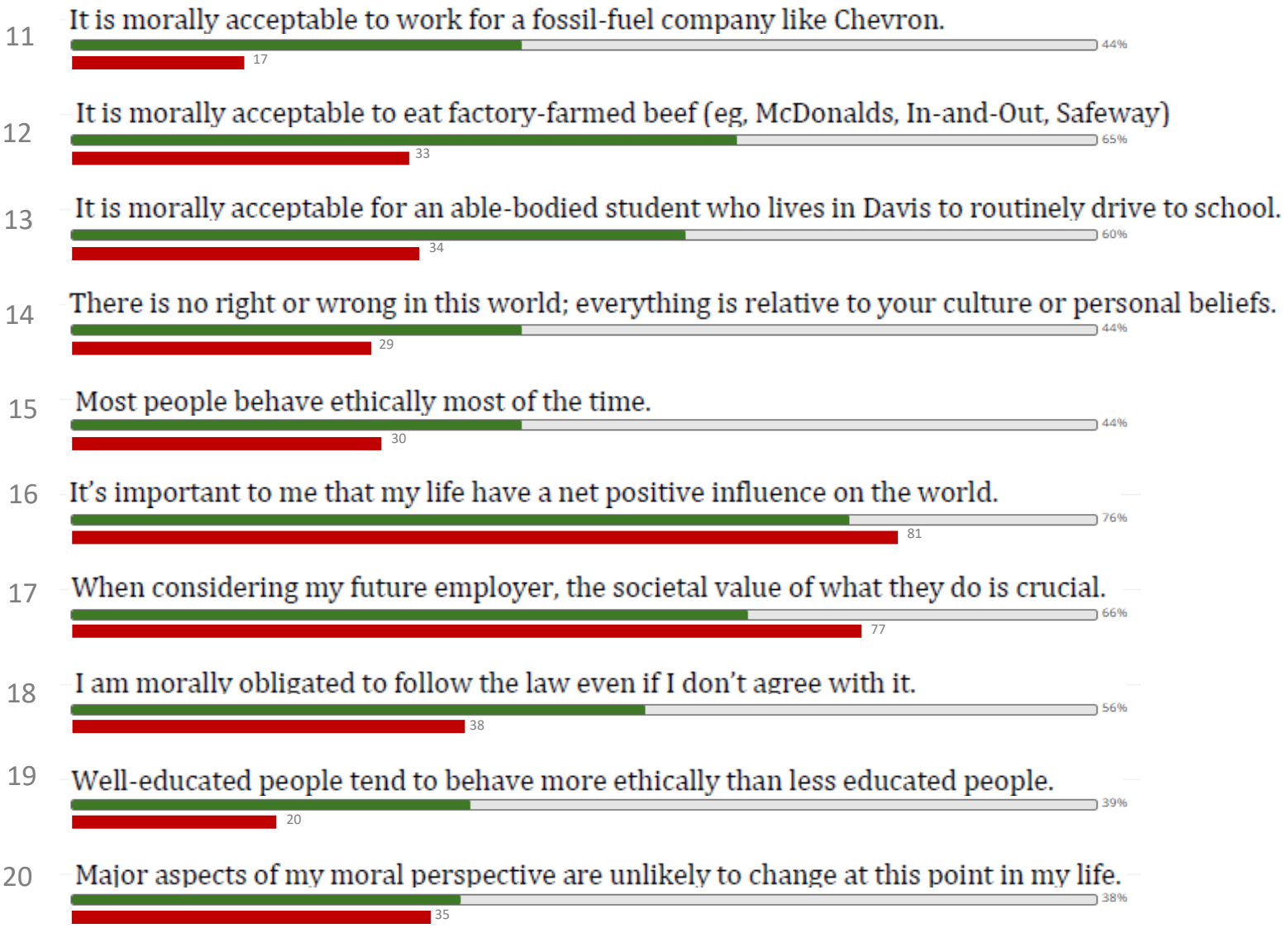
The technologies that us computer scientists develop have profound social, ethical, and political implications—all of which we, the technical people who do the work, routinely ignore. When we *do* think about our impact, we are often superficial in our analysis, overstate the positives, and downplay the negatives. This class will use the science-fiction series Black Mirror (2011-2023) to ground an exploration of recent and future impacts of CS. Is our work improving the world or wrecking it?

Our class will be run as a seminar in which we will watch Black Mirror episodes at home and then, in class, discuss the issues they implicate. The course is intended for CS and CSE students who are introspective, engaged, and communicative. Grading is based on (mandatory) attendance, participation, student writing, pop quizzes, a classroom discussion you will lead, and a creative final project, which you will present.

The class (along with ECS 127) will be the last course I teach at UCD. It serves as my final and desperate plea that we upend our disciplinary culture and start to think as much about *if* we should do stuff as *how* to get it done.

Do attitudes shift? Beginning — End (SQ23)

Do attitudes shift? Beginning — End (SQ23)



Things a student might gain

1. Becoming more introspective, humble, and skeptical
2. Licensing emotion / intuition, questioning rationalism
3. Feeling the presence of a community
4. Reconceptualizing CS as more than a technical endeavor pursued for self-interest

Concluding remarks

I spent my career writing technical papers – giving talks on them – obsessing over them.

They're mostly pretty good – I don't dislike them, or regret having written any of them.

But when I look back on my career, the two things I feel best about are

- *The Moral Character of Cryptographic Work*
- My ethics-and-technology class

The later, I believe, changed many students' lives.

Maybe you might like to teach an ethics-and-technology class, too?

Acknowledgments and References

1. Ian Barber: *Ethics in an Age of Technology* (The Gifford Lectures: Volume Two) (2013). First book I read that struck a tone I liked for my class. An assigned book that first term I taught.

2. Jacob Bronowski: *The Ascent of Man* (1973). BBC TV series that laid the roots for my wanting to teach an ethics class 30 years later.

3. Jonathan Haidt: *The Righteous Mind: Why Good People are Divided by Politics* (2012). Relied on heavily in preparing this talk. Provides some *post hoc* justification for my practices and ideas.

4. Anatoly Gromyko and Martin Hellman: *Breakthrough: Emerging New Thinking: Soviet and Western Scholars Issue a Challenge to Build a World Beyond War* (1988).
   My evidence that a serious scientist *could* refocus on social-ethical concerns.

5. Krzysztof Kieślowski: *Dekalog I* (1988). The 50-minute film I start my class with.
   Sets the tone for the entire term.

6. Neil Postman: *Teaching as a Subversive Activity* (1971).
   First pedagogy book I ever read, inspired how I aimed to run the classroom.

7. Daniel Quinn: *Ishmael* (1995). A book I now have the kids read in full.
   Inspired the "STN" label, trying to *name* the story that my students, I felt, had lived.
   .
8. Phillip Rogaway, editor, *Beyond Computer Ethics* (2004-2024). Living reader, changes with each rendition of the class. Course webpages enumerate what we actually did.

Title: Can Ethics Be Taught?

Speaker: Phillip Rogaway

Abstract: Beyond my technical work in cryptography, for 20 years I taught a class on ethics and technology. My students, who would soon graduate in one of our two Computer Science majors, had lived and breathed the Silicon Valley story, which centers technical skill and techno-optimism, individuality, disruption, and wealth. I wanted to turn them into skeptical, ethically focused, socially responsible people. So pretty much their polar opposites. Trying to do this year after year was challenging, depressing, and intensely personal. A fool's errand? What, if anything, was it all for?

Biography: Phil is, or maybe was, a well-known cryptographer. After graduating from MIT he spent much of his career at the University of California, Davis. He developed, along with Mihir Bellare, the approach that brings definitions and proofs to real-world cryptography. His schemes are in numerous cryptographic standards. For decades, however, Phil would fret over the social, political, and environmental threats that modern technologies bring forth. Sometimes he would write or speak about that, too.