

To all readers: Please note that this is a **work in progress**.

You can see all versions by clicking Ctl-Alt-Shift-H to view the version history.

The versions from November were used for the presentation. I am continuing to work on this.

Note to readers and commenters: Your feedback is welcome.

Use the 'Add Comment' button (or Ctl-Alt-M) to add a comment

Here are some ideas:

- have I missed a line of reasoning,
- is there a body of literature or discussion on a certain point,
- or a particular author I should consider here,
- are there examples (pro or con) of the point I'm making,
- is there data (pro or con) related to a point I'm making,
- is my reasoning sound,
- are there counter-examples to conclusions I draw,
- is my writing unclear or ambiguous

Note also that this is a document in progress, which means it may still be in the early stages of resource-gathering and organizing, rather than text-writing

Most of Chapters 1-4 (ie., everything up to 'Ethical Codes') are my own words and fully references; most of the rest is clippings from other articles, often only partially referenced.

"The street finds its own uses for things" - William Gibson, [Burning Chrome](#)

Contents

1. [Introducing](#)
2. [Applications](#)
3. [Issues](#)
4. [Codes](#)
5. [Approaches](#)
6. [Care](#)
7. [Machine](#)
8. [Practices](#)
9. [References](#)

Ethics, Analytics and the Duty of Care

Stephen Downes

Approaches to Ethics

Ethical Theory

- First of all, we can't just talk about "the ethics of such and such" as though we have solved ethics. There are multiple perspectives on ethics.

In 'The Ethics of Belief' (1877), Clifford gives three arguments as to why we have a moral obligation to believe responsibly, that is, to believe only what we have sufficient evidence for, and what we have diligently investigated.

- Beliefs lead to actions, and false beliefs about physical or social facts lead us into poor habits of action that in the most extreme cases could threaten our survival.
- poor practices of belief-formation turn us into careless, credulous believers.
- in our capacity as communicators of belief, we have the moral responsibility not to pollute the well of collective knowledge.

Francisco Mejia Uribe. (2018). Believing without evidence is always morally wrong. Aeon (Weblog). November 5, 2018.

<https://aeon.co/ideas/believing-without-evidence-is-always-morally-wrong>

The point here is that ethics is typically thought of as using reason to determine what is ethical and what we ought to do

Some of the major ethical theories, sclassified as per

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

:

Starting Points

What is the basis for statements about the ethics of learning analytics. Drew (2016), for example, states that “people think that it would be irresponsible not to use data science in certain cases, and that we should not lag behind other countries.” Are we to draft our ethical principles based on “what people think” or on whether we “lag behind” other countries? If not, why not?

MyNotes: Exploring Humanism, Chapters 1-2

Miguel Guhlin, Around the Corner, 2021/09/07

Icon

Miguel Guhlin shares his summary of Greg Epstein's book, Good Without God, and I thought I'd include it here as complexion and nuance for my upcoming MOOC. It follows a path well trodden by people like Kai Nielson and J.L. Mackie. I appreciated Guhlin's honest and forthright consideration of the work. There are three posts, covering chapters one and two, chapter three,

and chapters four and five. It considers the questions not only of whether humanists can be good, but even questions like "what are we striving for", including such things as human dignity and compassion.

Web: [Direct Link] [This Post] <https://www.downes.ca/post/72711>

Ethics as...

"To make sense of these complexly polysemous meanings of "ethics" in tech, we have identified four overlapping meanings of the word "ethics" among those who use the term most forcefully:

Moral justice

Corporate values

Legal risk

Compliance"

"Compliance, legal risk, corporate values, and moral justice are all sometimes described as "ethics," and at times work in parallel to pursue common goals, and at other times come into tension with each other. In the best case scenario those goals are complementary. In the worst case, the procedural and internal-facing nature of ethical technology development, and the marketing and public engagement around these capabilities, can operate as "ethics-washing" and obscure the more fundamental societal changes that are needed. However, we argue that the proceduralism and internalism of these processes does not mean that they aren't valuable. Indeed, the bold goals of those using the language of moral justice are likely impossible without organizations using proceduralism and internalism to render those goals as visible, achievable, and accountable."

Too Big a Word. Emanuel Moss and Jacob Metcalf.

<https://points.datasociety.net/too-big-a-word-13e66e62a5bf>

Ethics as requirements

"Ethical requirements are requirements for AI systems derived from ethical principles or ethical codes (norms)" (Guizzardi, et.al., 2020)

1.5 GUIDING PRINCIPLES

As Voltaire's *Candide* might have reflected, we are faced with the imperative to seek out the 'best of all possible worlds':

In assuring educational benefits, not least supporting student progression, maximising employment prospects and enabling personalised learning, it is incumbent on institutions to adopt key principles from research ethics.

As businesses, post-compulsory educational institutions are facing the same business drivers and globalised competitive pressures as any organisation in the consumer world.

To satisfy expectations of the 'born digital' / 'born social' generations, there is a likely requirement to take on ethical considerations, which may run contrary to the sensibilities of previous generations, especially in respect of the trade-off between privacy and service.

Notwithstanding these tensions, we conclude that there are common principles that provide for good practice:

Clarity, open definition of purpose, scope and boundaries, even if that is broad and in some respects open-ended.

Comfort and care, consideration for both the interests and the feelings of the data subject and vigilance regarding exceptional cases.

Choice and consent, informed individual opportunity to opt-out or opt-in.

Consequence and complaint, recognition that there may be unforeseen consequences and therefore providing mechanisms for redress.

Kay, D., Korn, N. & Oppenheim, C., 2012. Legal, Risk and Ethical Aspects of Analytics in Higher Education, Available at: <http://publications.cetis.org.uk/2012/500>

Schools of Thought

Many readers will find this section unnecessary, but for many others the range and variety of ethical theories extant may be new to them. It is my objective here to show that a significant number of questions and assumptions in dialogue around ethics are open for discussion. Ethics is by no means a complete or closed discipline; it is a living study that has been shaped and formed by thinkers from the ancient world through to the modern era.

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

(as good a statement of this as any)

Five Sources of Ethical Standards

1. The Utilitarian Approach - the ethical action is the one that provides the most good or does the least harm, or, to put it another way, produces the greatest balance of good over harm.
2. The Rights Approach - the ethical action is the one that best protects and respects the moral rights of those affected. This approach starts from the belief that humans have a dignity based on their human nature per se or on their ability to choose freely what they do with their lives. On the basis of such dignity, they have a right to be treated as ends and not merely as means to other ends.
3. The Fairness or Justice Approach - all equals should be treated equally. Today we use this idea to say that ethical actions treat all human beings equally-or if unequally, then fairly based on some standard that is defensible.
4. The Common Good Approach - life in community is a good in itself and our actions should contribute to that life. This approach suggests that the interlocking relationships of society are the basis of ethical reasoning and that respect and compassion for all others -- especially the vulnerable -- are requirements of such reasoning.
5. The Virtue Approach - A very ancient approach to ethics is that ethical actions ought to be consistent with certain ideal virtues that provide for the full development of our humanity. These virtues are dispositions and habits that enable us to act according to the highest potential of our character and on behalf of values like truth and beauty. Honesty, courage, compassion, generosity, tolerance, love, fidelity, integrity, fairness, self-control, and prudence are all examples of virtues.

Virtue and Character

Ethics is in the first instance the study of virtue in a person, in a person's actions, or in a society. But what is a virtue? The SEP says, "A virtue is an excellent trait of character. It is a disposition, well entrenched in its possessor—something that, as we say, goes all the way down, unlike a habit such as being a tea-drinker—to notice, expect, value, feel, desire, choose, act, and react in certain characteristic ways." ([SEP, 2017](#))

While we typically characterize virtue by means of various traits - honesty, frugality, piety, humility, caring, courage, generosity, moderation - the concept of virtue is not defined by those traits. It might be derived from some sense of ideals or perfection, as Plato might say, or it might be derived from the Greek notion of arete (ἀρετή) - “be all that you can be”.

The achievement of virtue is essentially tied up with the development of character. As Aristotle says, the achievement of virtue might be a lifetime task. Virtue is the opposite of what might be termed the “weakness of the will” - our succumbing to the temptation to indulge, to become intemperate, dishonest, or violent. ([Aristotle, 1959](#))

We see this perspective reflected in modern ethics by writers such as Michael Foucault (1985). In *The Use of Pleasure* he talks of morality as “self-formation as an ‘ethical subject,’ a process in which the individual delimits that part of himself that will form the object of his moral practice, defines his position relative to the precept he will follow, and decides on a certain mode of being that will serve as his moral goal.”

<https://slidetodoc.com/aristotle-virtue-ethics-character-virtue-courage-friendship-todays/>

<https://cpled.ca/about-cpled/competency-framework/professional-ethics-and-character/>

Note different between ‘act honestly’ and ‘*be* honest’

<https://studylib.net/doc/5803513/virtue-ethics-show> StudyLib

Principles

Virtue ethics teaches:

- An action is only right if it is an action that a virtuous person would carry out in the same circumstances.

- A virtuous person is a person who acts virtuously
- A person acts virtuously if they "possess and live the virtues"
- A virtue is a moral characteristic that a person needs to live well.

Most virtue theorists would also insist that the virtuous person is one who acts in a virtuous way as the result of rational thought (rather than, say, instinct).

Introduction to Virtue Ethics - BBC

<https://www.bbc.co.uk/ethics/introduction/virtue.shtml>

What are the Virtues?

The Cardinal Virtues

The cardinal virtues are four virtues of mind and character in both classical philosophy and Christian theology. They are prudence, justice, fortitude, temperance. They form a virtue theory of ethics. The term cardinal comes from the Latin *cardo* (hinge);[1] virtues are so called because they are regarded as the basic virtues required for a virtuous life.

Plato in *Republic* Book IV, 426–435

https://www.gutenberg.org/files/1497/1497-h/1497-h.htm#link2H_4_0007

(see also *Protagoras* 330b, which also includes *piety* (*hosiotēs*)). (

<https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0178%3Atext%3DProt.%3Asection%3D330b>) <https://www.gutenberg.org/cache/epub/1591/pg1591-images.html>

- Prudence (φρόνησις, *phrónēsis*; Latin: *prudentia*; also Wisdom, *Sophia*, *sapientia*), the ability to discern the appropriate course of action to be taken in a given situation at the appropriate time.
- Justice (δικαιοσύνη, *dikaiosýnē*; Latin: *iustitia*): also considered as fairness;[3] the Greek word also having the meaning righteousness

- Fortitude (ἀνδρεία, andreía; Latin: fortitudo): also termed courage: forbearance, strength, endurance, and the ability to confront fear, uncertainty, and intimidation
- Temperance (σωφροσύνη, sōphrosýnē; Latin: temperantia): also known as restraint, the practice of self-control, abstention, discretion, and moderation tempering the appetite. Plato considered Sōphrosynē, which may also be translated as sound-mindedness, to be the most important virtue.

https://en.wikipedia.org/wiki/Cardinal_virtues

Aristotle's List of the Moral Virtues

<https://kevinhabits.com/aristotles-12-virtues-from-courage-to-magnificence-patience-to-wit/>

Stoicism

https://slideplayer.com/4319715/14/images/slide_1.jpg

https://www.reddit.com/r/Stoicism/comments/bq1njj/easy_way_to_explain_stoicism_the_stoic_happiness/

Confucianism

Within Confucianism there are five constant virtues or wu chang (五常). In descending order of importance, the virtues are benevolence or ren (仁), righteousness or yi (义), propriety or li (理), wisdom or zhi (智) and fidelity or xin (信).

Taoism

<https://www.britannica.com/topic/Daoism/Concepts-of-human-being-and-society>

The “superior virtue” of Daoism is a latent power that never lays claim to its achievements; it is the “mysterious power” (xuande) of Dao present in the heart of the sage—“persons of superior virtue never act (wuwei), and yet there is nothing they leave undone.”

“Wuwei is neither an ideal of absolute inaction nor a mere “not-overdoing.” It is actions so well in accordance with things that their authors leave no traces of themselves in their work: “Perfect activity leaves no track behind it; perfect speech is like a jade worker whose tool leaves no mark.” It is the Dao that “never acts, yet there is nothing it does not do.” There is no true

achievement without wuwei because every deliberate intervention in the natural course of things will sooner or later turn into the opposite of what was intended and will result in failure.”

“Daoist sagehood, however, is internal (neisheng), although it can become manifest in an external royalty (waiwang) that brings the world back to the Way by means of quietism: variously called “non-intervention” (wuwei), “inner cultivation” (neiye), or “art of the heart and mind” (xinshu).”

James F Keenan

<https://www.bbc.co.uk/ethics/introduction/virtue.shtml>

The modern theologian James F Keenan suggests:

- Justice
 - Justice requires us to treat all human beings equally and impartially.
- Fidelity
 - Fidelity requires that we treat people closer to us with special care.
- Self-care
 - We each have a unique responsibility to care for ourselves, affectively, mentally, physically, and spiritually.
- Prudence
 - The prudent person must always consider Justice, Fidelity and Self-care.
 - The prudent person must always look for opportunities to acquire more of the other three virtues

“As persons, we are relational in three ways: generally, specifically, and uniquely. And each of these relational ways of being demands a cardinal virtue. As a relational being in general, we are called to justice. As a relational being specifically, we are called to fidelity. As a relational being uniquely, we are called to self-care... The fourth cardinal virtue is prudence, which determines what constitutes the just, faithful, and self-caring way of life for an individual.” pp. 723-724. <http://cdn.theologicalstudies.net/56/56.4/56.4.5.pdf> Theological Studies 56 (1995) PROPOSING CARDINAL VIRTUES, JAMES F. KEENAN, S.J.

Moulin Rouge

<https://letsmakesomeartdammit.blogspot.com/2014/10/moulin-ching.html>

Truth, Beauty, Freedom, Love

Character and Mindset

Nietzsche - Man and Superman

- Beyond Good and Evil
- Man and Superman
 - <https://en.wikipedia.org/wiki/Superman>
 - In Superman's new slogan, he no longer fights for "the American way" - <https://chicago.suntimes.com/entertainment-and-culture/2021/10/18/22732108/superman-slogan-american-way-new-truth-justice-better-tomorrow-dc-comics> - "The Man of Steel's new motto — 'Truth, Justice and a Better Tomorrow' — is meant to inspire 'people from around the world,' DC Comics says." By [Darel Jevens](#) Oct 18, 2021, 5:30am CDT Chicago Sun-Times

Role Models

- Aaron Rogers vs Terry Bradshaw - Terry Bradshaw, FOX NFL Sunday Crew Rip Aaron Rodgers: 'You Lied to Everyone' - <https://www.si.com/nfl/2021/11/07/terry-bradshaw-fox-nfl-sunday-crew-rips-aaron-rodgers> - Nov 7, 2021 - Sports Illustrated - [Andrew Gastelum](#)
-

Personality Types

This personality type knowingly shares fake news

<https://www.futurity.org/personality-type-fake-news-chaos-2653532-2/>

“The study, published in the Journal of Experimental Psychology: General, showed participants with liberal and conservative political beliefs both shared erroneous news stories to a certain degree. But conservatives who also scored low on conscientiousness engaged in such behavior to a greater extent—they were more likely than liberals or more conscientious conservatives to share misleading information, the research finds.”

Mindsets

<https://characterlab.org/playbooks/>

- Eg. ‘growth mindset’ - <https://characterlab.org/playbooks/growth-mindset/>

<https://www.lifehack.org/articles/productivity/11-characteristics-highly-effective-mindset.html>

<https://www.thethinkingbusiness.com/wp-content/uploads/2014/03/Mindset.png>

George Couros - Innovator’s Mindset - <https://georgecouros.ca/blog/archives/4783>

The Limits of Virtue Ethic

Simply developing one’s own character, though, might seem selfish to some. It’s self-indulgent, at the very least. And one might question whether the cultivation of virtue constitutes a *basis* for ethical action. We need a sense of *normative* virtue ethics, such that the virtues not only describe good character, but *prescribe* right actions. ([Hursthouse, 1998](#))

Travis Timmerman, Yishai Cohen: [The Limits of Virtue Ethics](#) (pdf, 9550 words)

A wide variety of normative ethical views have been developed as the field of ethics has progressed. Proponents of each view disagree with one another about the deontic status of acts and about the exact right-making features of acts. Utilitarians, for instance, believe that an action is right because it maximizes utility, while Kantians believe an action is right because it accords with the categorical imperative.

Virtue ethics is often understood as a set of normative ethical views that are purported rivals to versions of consequentialism, deontology, contractualism, and other normative ethical views. To be sure, not all accounts of virtue ethics are developed to fit this role.

- virtue-theoretic normative ethical theory (VNET)
- the criteria of rightness is complete to the degree to which it satisfies two conditions:
 - (A) taking a stance on the deontic status of anything that can have a deontic status in any possible circumstance, and
 - (B) identifying the maximal set of fundamental right-making features (if the theory in question thinks anything has a deontic status).
- satisfying condition (B) poses a dilemma for each VNET. Either it will fail to be a distinctively virtue ethical position or it will be committed to an implausible or mysterious set of fundamental right-making features.

Deontology and Duty-Based Ethics

Duty

Duty is a requirement to moral action

The word 'deontological' comes from the Greek word *deon*, which means 'duty'.

Duty-based ethics teaches that some acts are right or wrong because of the sorts of things they are

For Kant, morality poses the question of what would constitute a *duty* to act. This is found in the bases of Kantian morality autonomy and freedom. It is only through autonomy and freedom that

we have the possibility of making moral choices. As we would say today, “ought implies can”. The morality of making a choice entails the *possibility* of making a choice. (Kant, 1956)

So morality applies to any rational being, and the nature of morality can be known through reason (indeed, it is this very fact that makes morality possible at all). There are several elements to Kantian ethics; one of the most significant is the categorical imperative.

In a nutshell, this is the principle that we must act in a way that we would imagine the action being a universal law. This is not the principle your mother appeals to when she says “what if everybody did that?” Rather, it’s the idea that you would *will* people to act in such a way because such actions are inherently good. (Kant, 1998)

What sort of actions could be universalized in such a way. Many typical actions, those based merely in our own pleasures, where we use other people as a *means* to an end, would not qualify. The only consistent universal principle of morality imposes on us the duty to treat people as *ends* in themselves, rather than as a means to an end.

Deontology

Radford University Core Handbook: Core 202- Ethical Reasoning and Analysis

<https://lcubbison.pressbooks.com/chapter/core-202-ethical-reasoning/>

Deontology is a universal ethical theory that considers whether an action itself is right or wrong. Deontologists argue that you can never know what the results will be so it doesn’t make sense to decide whether something is ethical based on outcomes. You can consider it the opposite of consequentialism and utilitarianism in many ways. Deontologists live in world of moral rules: It is wrong to steal. It is right to keep promises. Deontology is also concerned with intentions. If you intended good through your action, then the action is good, no matter what actually happened as a result. Deontology encompasses two kinds of approaches: duty-based and rights-based.

Duty-based ethics says that there are universal moral norms or rules, and it is essential that everyone follows them. If you've ever said, "I did it because it was the right thing to do," then you've employed duty-based ethics.

The duties themselves may be tied to professional roles, too. Teachers have a duty to grade students fairly; police officers have a duty to enforce the law; psychologists have a duty to respect the confidentiality of their patients. When you encounter codes of professional conduct—either written or unwritten—likely you are dealing with duty-based ethics.

Moral Intuitions

Derived from the 'natural law theory' of the Middle Ages

E.g. Aquinas: "every substance seeks the preservation of its own being, according to its nature: and by reason of this inclination, whatever is a means of preserving human life, and of warding off its obstacles, belongs to the natural law." (Q94:2)

Act intuitionism, rule-intuitionism

Natural Law Theory

From John Finnis, Germain Grisez, and others, a set of seven basic, self-evident values from which moral norms could be safely derived

- Life, health, and safety
- Knowledge of reality and appreciation of beauty
- Some degree of excellence in work and play
- Living at peace with others, neighborliness, friendship
- Aesthetic experience as harmony and inner peace
- Harmony between one's choices, judgements and performances
- Religion and pursuit of ultimate questions of meaning and value

Rousseau on Human Freedom

- The meaning and importance of human dignity
- The primacy of freedom and autonomy
- The intrinsic worth of human beings

“The hope of creating a stable and just political society on the basis of narrow self-interest is a soul-shrinking and self-destructive dogma masquerading as a science of politics.” (p.19)

- <https://library.oapen.org/bitstream/handle/20.500.12657/30072/650028.pdf?sequence=1&isAllowed=y>
-
-

Duty and Necessity

Kant:

- “Nothing in the world—or out of it!—can possibly be conceived that could be called ‘good’ without qualification except a good will.”
- “Good will is good because of how it wills—i.e. it is good in itself.
- Morality should not depend on human nature and hence be subject to the fortune of change and the luck of empirical discovery

<https://www.earlymoderntexts.com/assets/pdfs/kant1785.pdf>

Yes but: “the road to Hell is paved with good intentions” (Pojman & Fieser, 2012, p. 127)

Kant’s Categorical Imperative

Hypothetical Imperative: If you want A, you must do B

Categorical Imperative: Do B

“Act only according to the maxim by which you can at the same time will that it would become a universal law.”

All specific duties can be derived from this one imperative

All from: <https://www.longwoods.com/content/17389>

Kant formulates his categorical imperative in at least three different ways:

- His first formulation - "I ought never to act except in such a way that I can at the same time will that my maxim should become a universal law" - is a test for logical consistency.
- Kant's second formulation states, "Act in such a way that you always treat humanity, whether in your own person or in the person of another, never simply as a means, but always at the same time as an end." For Kant, all rational beings are genuine participants in the realm of morality.
- The third formulation of Kant's categorical imperative is, "So act as if you were through your maxims a lawmaking member of a kingdom of ends." Treat yourself and every other rational being as jointly constituting a community of agents who, as ends in themselves, can accept only those laws that they have given themselves.

In his thesis, *Healthcare Workers' Duty to Care and Severe Infectious Diseases*, (Sokol, 2004) Daniel K. Sokol writes, "Religion, financial gain, reputation, personal character, social context, geographical location, severity and nature of disease, the climate of fear - these are all influential factors in doctors' decision to treat, perhaps more so than in any other historical period." Kant would disagree with this statement. According to Kant, persons have "autonomy of will" if and only if they knowingly act in accordance with the universally valid moral principles that pass the requirements of the categorical imperative.

Sokol D. 2004. Healthcare Workers' Duty to Care and Severe Infectious Diseases.
www.medicalethicist.net/Thesis3.doc .

Criticisms

Radford University Core Handbook: Core 202- Ethical Reasoning and Analysis

<https://lcubbison.pressbooks.com/chapter/core-202-ethical-reasoning/>

Both duty and rights-based ethics... rely on absolute principles regarding duties and rights. But there's no definitive list recorded anywhere.

Duty ethics allows little room for context. In *Les Misérables*, was Jean Valjean wrong to steal bread to feed his starving sister's children? Would it have been wrong to lie to a Gestapo officer asking where Jews were hidden or to slave-catchers in pursuit of runaways in the pre-war South? Some would say that the answers depend upon the circumstances and options available to us, rather than on it being the case that certain types of actions are always and necessarily wrong.

What about Futility?

Futility is defined as "inadequacy to produce a result or bring about a required end; ineffectiveness" [13]. Medically, the concept of "futility," according to the American Medical Association, "cannot be meaningfully defined" [14]. Essentially, futility is a subjective judgment, but one that is realistically indispensable [15]. There is consensus within the medical community that at specific times during the course of an illness some treatments are medically futile; consensus ends however, when attempts are made to formulate a fully objective and concrete definition.

Peter A. Clark (2007) Medical Futility: Legal and Ethical Analysis. *Virtual Mentor*. 2007;9(5):375-383.

<https://journalofethics.ama-assn.org/article/medical-futility-legal-and-ethical-analysis/2007-05>

DOI

10.1001/virtualmentor.2007.9.5.msoc1-0705.

Professional Duties

Groundwork, Kant's principle of morality classifies duties into four categories: duties to oneself and to others, and perfect and imperfect duties. Perfect duties are prohibitions of specific kinds of actions, and violating them is morally blameworthy. Imperfect duties are recommendations of certain acts, and fulfilling them is praiseworthy. In the later *Metaphysics of Morals*.

According to Kant, "All duties are either duties of right, that is, duties for which external law giving is possible, or duties of virtue, for which external lawgiving is not possible."

"Duty to care for a patient by physician is a duty of virtue. A person who decides to become a physician takes on the promise of using his or her abilities to the best advantage. Such a promise can be explicit (e.g., taking an oath) or implicit, but there is no denying its existence."

"One of Kant's most important claims is that the moral worth of an individual's actions depends exclusively on the moral acceptability of the maxim on which the person acts."

Ethical Rules

To prescribe right behaviour, one might appeal to a set of rules describing the virtues. A classic example of this is the Ten Commandments, which requires that adherents be honest, to not covet, to not kill, and the like. ([Bible: Exodus 20](#))

With rules one encounters almost immediately what has come to be known as 'the conflict problem'. In a case where the application of different rules produces different conclusions, which rule takes priority? Additionally, we encounter what might be called 'the exception problem' - the

rule may say, for example, that you must not kill - but what if this is the only possible result of defending oneself?

But more significantly, morality doesn't seem to simply be a matter of following the rules. "If right action were determined by rules that any clever adolescent could apply correctly, how could this be so? Why are there not moral whiz-kids, the way there are mathematical (or quasi-mathematical) whiz-kids?" ([Hursthouse, 1998](#))

The End of Principles?

"The concept of "ethical principles" for AI has encountered pushback both from ethicists, some of whom object to the imprecise usage of the term in this context, as well as from some human rights practitioners, who resist the recasting of fundamental human rights in this language." (Fjeld, et.al., 2020)

Consequentialism

the ethical action is the one that provides the most good or does the least harm, or, to put it another way, produces the greatest balance of good over harm.

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

Most common form: utilitarianism

The pursuit of pleasure – hedonism

The 'higher' pleasures – John Stuart Mill

The 'right of each person to pursue his own good in his own way, provided this does not infringe on the right of the other doing the same'

From intro: Perhaps, it may be argued, we should focus specifically on *outcomes*. This is a common line of reasoning education circles, focusing for example on 'what works' (Serdyukov, 2017) and 'effect sizes' (Hattie, 2008). But what are successful outcomes? Will it work next time? What happens when we can't predict what the consequences will be, and what happens when we can't repair bad consequences after the fact?

See <https://www.erudit.org/en/journals/mje/2017-v52-n1-mje03166/1040816ar.pdf>

- That said, there seems to be in general a recognition that consequentialist approaches to ethics are not satisfactory in the case of fields like learning analytics
 - We can't predict what the consequences will be (principle of unintended effects)
 - We can't repair bad consequences after the fact (principle of closing the barn door)

Hedonism, Egoism

Causal decision theory (CDT) says that you should do your best to improve the world in which you find yourself. An act is choiceworthy to the extent that you expect it to promote valuable ends. In contrast, evidential decision theory (EDT) says that you should give yourself good news about the world. An act is choiceworthy to the extent that it indicates valuable ends. Causal decision theorists complain that EDT prescribes an irrational policy of 'managing the news'

<http://pitt.edu/~jdg83/publication/pdfs/tcdtgtmtn.pdf>

Utilitarianism

Utilitarianism is sometimes known as 'the happiness principle'. The simplest statement of utilitarianism is that something is morally good according to whether it produces pleasure and

avoids pain. In a society, a morally good action is that which produces the greatest good for the greatest number. (J.,S.Mill, 1957) Utilitarianism is therefore an important statement of ethical *consequentialism*, that is, the idea that the effect of one's actions are relevant to ethical appraisal. It is worth noting that utilitarianism is concerned with the *goodness* of an act, as opposed to the Kantian concept of duty to act.

With utilitarianism come several immediate objections. (Smart & Williams, 1973) For one, there is the concern that utilitarianism caters to our lowest desires; for example, in *hedonism* we find the ethic of personal pleasure. Another is there is the question of how consequences may be measured (the unit of measurement sometimes derisively called a 'hedon'). Indeed, we might not be able to know, or to calculate in time, the 'unintended consequences' of an action.

Many of these objections are answered by John Stuart Mill. The cultivation of taste, he writes, leads one to enjoy the 'higher pleasures'. Better to be a discontented man than a contented sheep. As well, we need not evaluate each act individually. We may distinguish between 'act utilitarianism', which looks at the consequences of individual acts, and 'rule utilitarianism', which looks at the consequences of types of behaviour generally.

But a final critique of utilitarianism is that it is cold and unfeeling. Do the needs of the many genuinely outweigh the needs of the few? If seven billion people could be made to feel slightly better by the life-long torture of one person, is this act morally permissible? Intuitively this seems wrong, though a utilitarian calculation might say otherwise.

[Philosophy, et cetera](#)

[Updates to utilitarianism.net](#)

www.philosophyetc.net/2021/10/updates-to-utilitarianismnet.html (HTML, 247 words)

Back in July, I mentioned our new introduction to population ethics. Since then, I've also added a chapter on Theories of Well-being, and -- brand new as of today -- Arguments for Utilitarianism. I'm inclined to think the best case for utilitarianism stems from simply reflecting on what fundamentally matters (and one who doesn't find the utilitarian answer here intuitively compelling is unlikely to be much moved by any other argument in support of the view). ...

Found 2 days, 20 hours ago on [Philosophy, et cetera](#)

Egoism

Another form of consequentialism, *Egoism* is the philosophy that one is required only to act in their own self-interest. This is the philosophy often associated with Ayn Rand under the heading of 'objectivism' (Rand, 1970), and though Rand's arguments in favour are incoherent, reasoned argumentation for egoism is not rare.

Egoism can be expressed in different ways. "Psychological egoism asserts that it is impossible for anyone to do anything other than seek his own good. Ethical egoism tells us that a person ought to promote his own interests." ([McConnell, 1978](#)) Both of these suggest that whatever the status of ethical theory, it is not really *possible* for a person to adopt any ethics other than personal self-interest.

Egoism forms the foundation of modern economics. As Adam Smith Writes, "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages" (Smith, 1937, I.ii.2).

James Rachels - Our triumphs seem hollow unless we have friends to share them, and our failures are made bearable by their understanding.

- *The Elements of Moral Philosophy* (1999), p. 183

Social Contract

While we usually associate consequentialist theories with the pursuit of pleasure and avoidance of pain, consequentialist theories can identify other goods, for example, justice, fairness, and equality. However these are even more difficult to define and measure than pleasure and pain. An alternative mechanism is required; historically this has been the social contract.

The social contract appears first with any significance in modern philosophy, and in particular the work of Hobbes, Locke and Rousseau.

John Locke depicts the contract as a mechanism to defend the rights of citizens against the sovereign, and in particular, to protect their right of property, which they acquire by removing goods from the state of nature and adding their own labour to them. Failing this, writes Locke, the recourse is either legitimate revolution to overthrow the sovereign, or emigration to unoccupied land. (Locke, 1821)

“Man is born free,” writes Rousseau at the beginning of the *Social Contract*, “yet everywhere he is in chains.” Rousseau depicts a ‘state of nature’ quite opposite to Hobbes, where people lived in peace and plenty, and the net effect of society was to constrain this freedom and enslave people to serve the individual will of the master. The objective of the social contract is to ascertain ‘the general will’ expressed by the unanimity of citizens. (Rousseau, 1950)

A significant and influential modern version of social contract theory emerges with John Rawls’ *A Theory of Justice*. Rather than postulate an ethically dubious ‘state of nature’, Rawls proposes that we imagine what sort of contract we would negotiate with each other if we were not aware of where we would be in society. What results, he argues, is a theory of “justice as fairness” (which doesn’t sound remarkably different than Plato’s version, “to everyone his due.” (Rawls, 1971)

Fairness

all equals should be treated equally.

today we use this idea to say that ethical actions treat all human beings equally-or if unequally, then fairly based on some standard that is defensible.

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

Has roots in the law codes of Solon and the idea of universal law

Also based in social contract theory – Locke, Hobbes

Current: John Rawls, *A Theory of Justice* (‘justice as fairness’)

A lovely summary of Rawls:

<https://medium.com/patrickdaniel/a-theory-of-justice-an-introduction-to-john-rawls-208594ba3b75>

Good stuff here:

Saif M. Mohammad. (2019). Examining Fairness in Language Through Emotions. National Research Council Canada.

<http://saifmohammad.com/WebDocs/Saif-2019-EIAI-talk-web.pdf>

“The Right to Freedom from Discrimination One of the most preoccupying aspects of artificial intelligence is how it will exacerbate local and global economic inequality. According to the World Economic Forum, seventy percent of the global economic impact of AI by 2030 will be gained by North America and China, while developing countries in Asia, Latin America and Africa will see less than six percent of the overall gain.⁵¹ Artificial intelligence systems will further be used to determine who has access to credit; this may negatively affect the lives of families around the world as “there is a significant risk that the information used to generate credit scores is systematically biased against minority communities.”⁵² These two examples illustrate the direct potential economic impact of new applications of artificial intelligence on people’s lives, which is on a track to broaden economic inequalities and thus may disparately and negatively affect the lives of children who are born in already-difficult economic and social conditions.” (UC Berkeley, 2019)

“The “fairness” principle was defined as equitable and impartial treatment of data subjects by AI systems. We used this definition, drawn from common usage, over a technical one because articulations of fairness in the documents coded under this principle are not especially technical or overly specific in spite of the rich vein of academic research by AI and machine learning academics around competing mathematical formalizations of fairness.²³⁴ However, Microsoft adds to its principle “AI systems should treat all people fairly” the further elaboration that “industry and academia should continue the promising work underway to develop analytical techniques to detect and address potential unfairness, like methods that systematically assess the data used to train AI systems for appropriate representativeness and document information about its origins and characteristics.””: (Fjeld, et.al., 2020:49)

EqualityThe principle of “equality” stands for the idea that people, whether similarly situated or not, deserve the same opportunities and protections with the rise of AI technologies. “Equality” is similar to “fairness” but goes farther, because of fairness’s focus on similar outcomes for similar inputs. As the European High Level Expert Group guidelines puts it: “Equality of human beings goes beyond non-discrimination, which tolerates the drawing of distinctions between dissimilar situations based on objective justifications. In an AI context, equality entails that the same rules should apply for everyone to access to information, data, knowledge, markets and a fair distribution of the value added being generated by technologies.”²³⁸ (Fjeld, et.al., 2020:50-51)

“Fair Risk Distribution: Autonomous vehicles ought to be programmed so that, to the greatest extent possible, the risks of the road are distributed equally among all of those who might be harmed as a result of the use of motor vehicles, unless there is a morally compelling reason for deviating from this aim.” (Berkey, forthcoming).

Brian Berkey. Forthcoming. Autonomous Vehicles, Business Ethics, and Risk Distribution in Hybrid Traffic. *Autonomous Vehicle Ethics: Beyond the Trolley Problem* (OUP).
<https://faculty.wharton.upenn.edu/wp-content/uploads/2020/01/Autonomous-Vehicles-Business-Ethics-and-Risk-Distribution-in-Hybrid-Traffic.pdf>

--

“Serious thinkers in academia and business have swarmed to the A.I. bias problem, eager to tweak and improve the data and algorithms that drive artificial intelligence. They’ve latched onto *fairness* as the objective, obsessing over [competing constructs of the term](#) that can be rendered in measurable, mathematical form. If the hunt for a [science of computational fairness](#) was restricted to engineers, it would be one thing. But given our contemporary exaltation and deference to technologists, it has limited the entire imagination of ethics, law, and the media as well.(Powles and Helen Nissenbaum, 2018))

What is Fairness?

Is fairness something that can be addressed algorithmically? Eg. FairSight: Visual Analytics for Fairness in Decision Making. <https://arxiv.org/pdf/1908.00176.pdf>

Is fairness just a lack of bias? As suggested in numerous papers, eg.
<https://arxiv.org/pdf/1802.01029.pdf>

there are several incompatible standards for what fairness and de-biasing might mean (see J. Kleinberg, S. Mullainathan and M. Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores, ITCS 2017 <https://arxiv.org/abs/1609.05807>)

--

Fair Data Economy

The fair data economy is a new Finnish-led European initiative for better, trust-based use of personal data, which contributes to more personalized services, market competitiveness, better public services and overall well-being.

In the fair data economy, people are in control of how their data is used and shared, while all businesses, no matter the size, have access to larger data pools, maximising their ability to succeed.

Believe it or not - instead of a handful of super platforms there is a myriad of new winners in the fair data economy!

--

FAIRsFAIR - <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

Fairness and Machine Learning

Fairness and machine learning

Solon Barocas, Moritz Hardt, Arvind Narayanan, 2020/08/21

<https://fairmlbook.org/>

The sections on classification and causality remind me of the work I did as a graduate student learning the foundations of probability theory. It also reminds me of Clark Glymour's 'Tetrad Automated Causal Discovery Platform' for discovering "valid, novel, and significant causal relationships" in data.' These are the sorts of things bringing together ethicists and epistemologists today. Anyhow, this draft textbook (several chapters are still incomplete) reads well and is relatively accessible, though readers will find themselves in some heady territory in

the later stretches of both sections mentioned above. But there's some really good discursive content here, for example, the section on 'no fairness through unawareness', which describes the sort of issues that can arise when using (say) DNA to identify potential academic achievement.

Redressing Inequality

“Bias and Fairness in the ML community

In the past years we've had a growing discussion about how to reduce social bias and increase fairness, which is good. We wouldn't want our models to be discriminatory because of discriminatory historical data.

However what I've noticed is that some have taken this as an opportunity to poison the well by suggesting active discriminatory practices in seductive language and insidious framing.”

https://www.reddit.com/r/MachineLearning/comments/f0kqbw/d_bias_and_fairness_in_the_ml_community/

Brian Berkey: [The Value of Fairness and the Wrong of Wage Exploitation](#) (pdf, 8960 words)

“It seems that ProPublica was tying fairness to one set of criteria, while Northpointe and Flores et al. were tying fairness to a different set of criteria. How should we decide which side was right? How should we decide whether COMPAS was really unfair or biased against black people? More generally, how should we decide whether an algorithm is unfair or biased?”

https://www.kevindorst.com/stranger_apologies/how-not-to-test-for-algorithmic-bias-guest-post

Phenomenal World August 23rd, 2019

Is it impossible to be fair?

<https://phenomenalworld.org/analysis/impossible-to-be-fair>

Impact

A consistent application of the principle of fairness would result in a very different social, economic, and cultural environment. For example, consider the result when the principle is applied to academic culture. “ An analysis of the ethicality of current arrangements in academic publishing shows that monetization and the sequestering of scientific knowledge behind paywalls breach the principle of fairness and damage public interest... there are ethical grounds for consumers of scientific knowledge to invoke the right of collective civil disobedience, including support for pirate open access” (James, 2020).

Limitations

The Problem with Too Much Fairness

Abhishek Chakraborty

<https://coffeeandjunk.com/fairness/>

The Common Good

Plato, Rousseau

Communitarianism

<https://plato.stanford.edu/entries/communitarianism/>

This article, a substantial revision of the version first published two decades ago, describes 'communitarianism' as a philosophical and ethical response to some of the principle tenets of liberalism as expressed, say, in Rawls's [A Theory of Justice](#), disputing the idea that "the principal task of government is to secure and distribute fairly the liberties and economic resources individuals need to lead freely chosen lives." The article as a whole sets out three major responses to liberalism:

1. One ground of objection disputes the idea that there is a universal set of principles; societies may order the principles differently, justify them differently, or even identify principles unique to their culture.
2. Another set of objections is based on the argument that "liberalism rests on an overly individualistic conception of the self" and that "this view neglects the extent to which individuals are embodied agents in the world." In particular, this argument opposed the view that morality is the product of individual choice, and that there are wider social

contexts to be considered, for example, a "communal attachment so constitutive of one's identity that it cannot be revised and rejected."

3. The third argument states that liberalism has failed, pointing to the "negative social and psychological effects related to the atomistic tendencies of modern liberal societies." In contrast, "community is valuable—at least as valuable as the need for freedom, if not more so."

I think this article is valuable because it draws out the line distinguishing the recent 'ethics of care' and other communitarian trends in open pedagogy and the prevailing neo-liberal and capitalist ethic of the marketplace. At the same time, it points to some of the ways the communitarian response actually borrows from liberal tradition, for example, by arguing for universalist conceptions of natural justice based in fairness or individual freedom of choice. But the article as a whole should be criticized; it functions mostly as an argument *against* communitarianism, frequently raising straw man objections (gated communities, karoshi (death from overwork), amputations) and ultimately defending "good forms of hierarchy that express morally defensible values."

life in community is a good in itself and our actions should contribute to that life.

the relationships of society are the basis of ethical reasoning

respect and compassion for all others -- especially the vulnerable

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

Examples include Ubuntu philosophy, cooperativism, and collectivism. Social cohesion and social order are primary values. E.g. Xi Jinping - "seven years of rural life [that] gave me something mysterious and sacred."

http://www.ce.cn/xwzx/gnsz/szyw/201801/30/t20180130_27970996.shtml

Via:

Francisco Mejia Uribe. (2018). Believing without evidence is always morally wrong. Aeon (Weblog). November 5, 2018.

<https://aeon.co/ideas/believing-without-evidence-is-always-morally-wrong>

In '[The Ethics of Belief](#)' (1877), Clifford gives three arguments as to why we have a moral obligation to believe *responsibly*, that is, to believe only what we have sufficient evidence for, and what we have diligently investigated.

- Beliefs lead to actions, and false beliefs about physical or social facts lead us into poor habits of action that in the most extreme cases could threaten our survival.
- poor practices of belief-formation turn us into careless, credulous believers.
- in our capacity as communicators of belief, we have the moral responsibility not to pollute the well of collective knowledge.

Note that most of what follows is currently cut-and-paste notes from other sources that I've used to organize my thoughts, and not actual text I've written.

Inclusiveness

"Inclusiveness in Impact" "Inclusiveness in impact" as a principle calls for a just distribution of AI's benefits, particularly to populations that have historically been excluded. There was remarkable consensus in the language that documents employed to reflect this principle, including concepts like "shared benefits" and "empowerment". (Fjeld, et.al., 2020:51)

"Human Values and Human Flourishing" The principle of "human values and human flourishing" is defined as the development and use of AI with reference to prevailing social norms, core cultural beliefs, and humanity's best interests. As the Chinese AI Governance Principles put it,

this principle means that AI should “serve the progress of human civilization.”³⁰⁸ (Fjeld, et.al., 2020:61)

MetaEthics

The study of meta-ethics is the study of what *grounds* an ethical argument. To some degree this discussion is already present in the range of ethical theories described above (and many writers place the discussion of meta-ethics prior to the list of ethical theories). I have chosen to place it here because, after reflection on the different theories, it is relevant to ask about the bases or grounds for one approach or another.

For example, as we consider these different theories, we see that even what counts as *ethical* can vary from one viewpoint to another. Some see it as a form of excellence in individuals, others see it as defined in terms of duties and responsibilities, still others characterize ethics in terms of good and bad or right and wrong, while others see ethics expressed in terms of value and worth.

Does Might Make Right?

Suppose Gyges has a ring, says Glaucon in Plato’s Republic, where this ring makes him invisible and hence essentially free of retribution for any act. He can take whatever he wants, lie with anyone he wants, even murder anyone he wants, and there will be no retaliation. Why then would he act in a moral manner at all, no matter how we define morality? (Plato, 2000)

Friedrich Nietzsche makes a compelling modern case for this argument. He argues that if a man becomes ‘Superman’ (*ubermensch*), then whatever he does is by that fact moral. (Nietzsche, 1900) We see echoes of this today in the proclamations of Donald Trump when he observes that the President *can’t* be in a conflict of interest. (Voskuhl & Melby, 2016)

Conversely, if a person must behave ethically because of the power of an authority (whether it is the will of God or the dictates of a King) and is unable to do otherwise, on what grounds would we call behaving in this manner moral at all? If I am falling, and will kill someone when I land on him, I am powerless to stop or to change direction. Am I still responsible for the man’s death?

The relation between power and morality is a complex one. If morality is based on subservience to power, this takes away the element of choice, which seems essential to morality. But if the element of power is removed, what then makes an act moral or immoral?

Power and Control

Foucault here

Rights and Duties

-
- the ethical action is the one that best protects and respects the moral rights of those affected.
- starts from the belief that humans have a dignity based on their human nature per se or on their ability to choose freely what they do with their lives.
- on the basis of such dignity, they have a right to be treated as ends and not merely as means to other ends

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/>

This is, in essence, Immanuel Kant

Also has a basis in Constitution Theory, Libertarianism

Rights of the Child, UN

--

Rights. <https://plato.stanford.edu/entries/rights/>

--

Why human rights are not enough

To combat the narrow-minded view of whom society belongs to, we must recognise a more sophisticated idea of what it means to be a citizen. - By Will Kymlicka (Kymlicka, 2020)

“the restriction of rights to members seems morally arbitrary. Why should someone’s life chances be so powerfully determined by the accident of which side of the Rio Grande they were born on? Ayelet Shachar calls this the “[birthright lottery](#)”, equivalent to the medieval idea of being born into a particular caste.”

“many philosophers also think there may be something valuable in the idea of membership rights. A good society, we might think, should have some sense of solidarity: members would think of themselves as having special duties to one another, above and beyond the basic respect owed to all human beings. A one-track conception of justice, relying solely on universal humanitarianism, might not lead to a levelling-up in our treatment of outsiders – treating them as well as we currently treat insiders – but rather to a levelling-down, treating everyone as if they were strangers to whom we have no special obligations.”

Will Kymlicka is the Canada Research Chair in Political Philosophy. He is the author of [Multicultural Citizenship](#).

Rationalism

Autonomy and Rationality

<https://press.rebus.community/intro-to-phil-ethics/>

“Kant’s theory, a model of cool and detached reasoning, unconcerned with personal or familial commitments, became a requirement of independent decision-making. However, this way of thinking

about autonomy is problematic because, under such requirements, one must either acknowledge that

no person fully meets the criteria, or willfully ignore that any person’s ability to be independent is facilitated by the ongoing care provided to them by others.”

“Relational theories of autonomy generally start with the minimal acknowledgment

that we begin as non-autonomous beings, as infants, and develop into autonomous beings gradually

as we learn various sets of skills and gain specific abilities central to making our own decisions,

from the mundane to the momentous. Many relational theories of autonomy also take into account

that our autonomy is impacted by the process of socialisation (Benson 1991; Meyers 1987), or may

be suspended at various times in our lives.“

“Buckingham Shum and Ferguson (2012, p.4) have argued that “new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration”. In consequence, LA in social systems (e.g., in the context of a classroom at a school) “must account for connected and distributed interaction activity”.

Buckingham Shum and Ferguson therefore propose social learning analytics as a domain in its own right (2012). Similar, gamification or gameful design for learning is considered as an on own domain using”

https://link.springer.com/epdf/10.1007/s10758-019-09399-5?shared_access_token=-om60iFmL_P9pkL3UF1JeGve4RwIQNchNByi7wbcMAY4q_htjt1aLUg4VNoQvnXPvI85d31qUIbZ9Mw6kRF2axjZoKvFRuLoU1QN2CmSdwRAoAUHOeVYWwj76wRIJXVLRVM0sua0j6sg-yBpzlwbAVQ%3D%3D

The coincidence between the right and the beneficial

<http://alexanderpruss.blogspot.com/2020/08/the-coincidence-between-right-and.html>

Are all practical reasons based on value?

Benjamin Kiesewetter, PhilPapers, 2020/09/02

The value-based theory of practical reasons (VBT) is the idea that any practical reason that can be given for performing a task can be explained in terms of value. 'Value' here means not just monetary value, but also instrumental value, such as pleasure or companionship. This article (29 page PDF) gives an account of the theory and then several reasons to question it with relation to (say) keeping promises, following orders or ensuring equitable distribution - reasons why actions may be good even though the action produces no particular value. This (to me, and in a broader sense) undermines the idea that 'good actions' are based on reason at all, not even practical reasons. The paper is an interesting read, gives you background into VBT and its objections, though it's written in that typically analytical philosophical language. Image: Schwartz theory of basic values.

<https://www.downes.ca/post/71303>

Agency

[Agency and self-knowledge](#)

Brie Gertler, [The Routledge Handbook for the Philosophy of Agency](#), 2020/03/06

In [Routledge Handbook for the Philosophy of Agency](#), Luca Ferrero (ed.), Routledge (forthcoming)

“The main role of reasons is not to motivate or guide us in reaching conclusions but to explain and justify after the fact the conclusions we have reached” (Mercier & Sperber, 2017, p. 121)

--

“Our rational beliefs and intentions are not mere mental attitudes, but active states of normative commitment” (Korsgaard 2009, 39). Korsgaard, C.M. 2009. The Activity of Reason. Proceedings and Addresses of the American Philosophical Association 83: 23-43.

The most prominent agentialist accounts (Burge 1996, Moran 2001, Boyle 2009) take our epistemic authority, relative to our attitudes, to derive from our agential authority—specifically, our authority as rational thinkers. These accounts are motivated by the view that beliefs and intentions are active insofar as they are responsive to reasons.

Burge, Tyler. 1996. Our Entitlement to Self-Knowledge. Proceedings of the Aristotelian Society 96: 91-116.

Tyler Burge’s version of agentialism centers on the idea that, insofar as we are rational thinkers, we have certain obligations. We are obligated to (try to) avoid inconsistent beliefs, to conform our beliefs to our evidence, and to strive to satisfy other rational norms. Burge claims that we could not hope to satisfy these obligations unless we had the capacity for self-knowledge, since we must know what we believe in order to assess our beliefs for consistency with each other and for conformity with our evidence.

...

The idea that avowing a belief or intention is commissive forms the basis for Annalisa Coliva’s version of agentialism (Coliva 2012). She argues that, so long as a thinker is rational and possesses the relevant concepts, self-ascriptions of beliefs or intentions will create the attitudes ascribed, by committing the thinker to those attitudes. She extends this “constructivist” view to rationally held conative attitudes, such as desires.

Coliva, Annalisa. 2012. One Variety of Self-Knowledge: Constitutivism as Constructivism. In *The Self and Self-Knowledge*, A. Coliva (ed.), Oxford: Oxford University Press, pp. 212-242.

...

The well-established fact that attitudes are often shaped by non-rational factors challenges the epistemic basis of rational agency accounts of self-knowledge.

...

Victoria McGeer presses this kind of objection (McGeer 1996, 2007). She argues that a more accurate depiction of our cognitive lives will rely on a more expansive understanding of first-person agency, beyond the agency involved in responsiveness to reasons. Given that we are not ideally rational cognizers, she says, fulfilling our responsibility for our own attitudes

sometimes requires adopting a detached, third-person perspective on our cognitive lives. (Moran regards this as a kind of self-alienation.) From this perspective, we can better identify biases, impulses, and other non-rational factors affecting our sense of our reasons. Recognizing these non-rational factors is required if we are to curb their influence (McGeer 2007, 102).

McGeer, Victoria. 1996. Is "Self-Knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy* 93: 483-515.

Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.

Consent

'It's never okay to say no to teachers': Children's research consent and dissent in conforming schools contexts - <https://onlinelibrary.wiley.com/doi/full/10.1002/berj.3638?af=R>

Naturalism

There is a long tradition in ethics, often depicted as a variation of rationalism, to the effect that right and wrong are defined by natural law. This can be expressed in different ways. For example, there is the argument that human rights are based in natural law, as evidenced in the U.S. Declaration of Independence: "We hold these truths to be self-evident: That all men are created equal; that they are endowed by their Creator with certain unalienable rights..." ([Stoner, 2017](#))

There is also an interpretation of naturalism and natural law to the effect that we should behave according to our nature, or (variously) according to our best nature. Thomas Aquinas, for example, places the creation of our nature in the hands of God, which therefore makes behaving according to that nature. ([Magee, 1996](#)) Flavours of naturalism can also be found in Taoist and Confucian thought. ([Nelson, 2009](#))

But can we deduce moral facts from nature, or even from human nature? David Hume argued famously that one cannot deduce an 'ought' from an 'is'. If it's the *nature* of something to do something, there is no right or wrong about it. (Hume, 2003) G.E. Moore called such an

inference “The Naturalistic Fallacy.” Specifically, the fallacy is “the assumption that because some quality or combination of qualities invariably and necessarily accompanies the quality of goodness, or is invariably and necessarily accompanied by it, or both, this quality or combination of qualities is identical with goodness.” (Moore, 1903)

There is, after all, no means of determining *which* natural properties are identical (or opposite to) goodness. If flight is not natural, is flight a sin? If violence is natural, is violence ethically acceptable?

Relativism

Kant argues that morality is based on the categorical imperative, the duty that arises out of universal moral precepts. But what if morality exists only in relation to some purpose, goal or outcome. Then they become *hypothetical* imperatives. In her paper of the same name, Philippa Foot asks what we are to say to the man who does not care about the ends we would ascribe to the moral man - justice, liberty, etc. (Foot, 1972) If he does care about them, it is because he values them as an end, not because he must (in absolute sense) *ought* to care.

Care ethics is a type of morality that can be understood as a hypothetical imperative. Drawn from feminist theory, which stresses nurturing and relationships, “care ethics affirms the importance of caring motivation, emotion and the body in moral deliberation, as well as reasoning from particulars.” (IEP, 2017) What’s significant about care ethics is that it addresses not only motivations and actions, but also attitudes and motivations. (Held, 2006)

A final question concerning relativism is whether it is feasible. While some argue there can be no compromise on ethical principle, relativists will generally hold that different perspectives can (to a certain degree) be *compatible* with each other. For example, in a society *some* people may subscribe to care ethics, but it does not follow that *all* people must entertain the same attitudes and motivations. In economics we have the concept of “incentive compatibility”, which expresses a similar idea, where people may have different interests, provided they are consistent with the principles of exchange adopted by the group. (Myerson, 2009)

<https://www.csus.edu/indiv/g/gaskilld/ethics/emp%20chapter2.htm>

Thomistic Philosophy - the Philosophy of Saint Thomas Aquinas. Web. 17 Jan. 2017.

Aristotle, and W. D. Ross. *The Nichomachean Ethics*. London: Oxford UP, 1959. Print.

"BibleGateway." *Exodus 20 - - Bible Gateway*. Web. 17 Jan. 2017.

Boisvert, Daniel R. "Charles Leslie Stevenson." *Stanford Encyclopedia of Philosophy*.

Stanford University, 15 Apr. 2011. Web. 18 Jan. 2017.

"The Declaration of Independence | Natural Law, Natural Rights, and American Constitutionalism." *The Declaration of Independence | Natural Law, Natural Rights, and American Constitutionalism*. Web. 17 Jan. 2017.

"The Ethics of Tenuous Faculty." *The Ethics of Tenuous Faculty | AAUP*. Web. 20 Jan. 2017.

Foot, Philippa. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review* 81.3 (1972): 305. Print.

Foucault, Michel. *The Use of Pleasure*. New York: Pantheon, 1985. Print.

Held, Virginia. *The Ethics of Care: Personal, Political, and Global*. Oxford: Oxford UP, 2006. Print.

Hobbes, Thomas, and C. B. Macpherson. *Leviathan*. Harmondsworth, Eng.: Penguin, 1986. Print.

Hume, David. "An Enquiry Concerning the Principles of Morals." *The Clarendon Edition of the Works of David Hume: An Enquiry concerning the Principles of Morals* (1751):

1-2. Print.

Hume, David. "A Treatise of Human Nature." *David Hume: A Treatise of Human Nature (Second Edition)* (1739). Print.

Hume, David, and Tom L. Beauchamp. *An Enquiry concerning the Principles of Morals: A Critical Edition*. Oxford: Clarendon, 2003. Print.

Hursthouse, Rosalind. "Normative Virtue Ethics." *How Should One Live?* (1998): 19-36. Print.

Hursthouse, Rosalind. "Virtue Ethics." *Stanford Encyclopedia of Philosophy*. Stanford University, 18 July 2003. Web. 17 Jan. 2017.

Kant, Immanuel. *Critique of Practical Reason*. New York: Liberal Arts, 1956. Print.

Kant, Immanuel, and Mary J. Gregor. *Groundwork of the Metaphysics of Morals*. Cambridge, U.K.: Cambridge UP, 1998. Print.

Locke, John. *Two Treatises on Government*. London: Printed for R. Butler, 1821. Print.

McBride, Kelly. "The New Ethics of Journalism: About This Blog." *Poynter*. Poynter, 25 Nov. 2014. Web. 19 Jan. 2017.

McBride, Kelly, and Tom Rosenstiel. *The New Ethics of Journalism: Principles for the 21st Century*. Los Angeles: SAGE, 2014. Print.

McConnell, Terrance C. "The Argument from Psychological Egoism to Ethical Egoism."

Australasian Journal of Philosophy 56.1 (1978): 41-47. Print.

Mill, John Stuart, and Oskar Piest. *Utilitarianism*. Indianapolis: Bobbs-Merrill, 1957. Print.

Moore, G. E. *Principia Ethica*. Cambridge: At the UP, 1903. Print.

Myerson, Roger B. "Fundamental Theory of Institutions: A Lecture in Honor of Leo

Hurwicz." *Review of Economic Design* 13.1-2 (2009): 59-75. Print.

Nietzsche, Friedrich Wilhelm, and Thomas Common. *Thus Spake Zarathustra*. New York:

Modern Library, 1900. Print.

Plato, G. R. F. Ferrari, and Tom Griffith. *The Republic*. Cambridge: Cambridge UP, 2000.

Print.

Rand, Ayn, and Nathaniel Branden. *The Virtue of Selfishness: A New Concept of Egoism*.

New York: Signet/New American Library, 1970. Print.

Rousseau, Jean-Jacques, and G. D. H. Cole. *The Social Contract: And Discourses*. New

York: E.P. Dutton and, 1950. Print.

Smart, J. J. C., and Bernard Williams. *Utilitarianism; for and against*. Cambridge: U, 1973.

Print.

"Statement on Professional Ethics." *Statement on Professional Ethics | AAUP*. Web. 20

Jan. 2017.

Stevenson, Charles Leslie. "li—The Emotive Meaning Of Ethical Terms." *Mind* XLVI.181

(1937): 14-31. Print.

Voskuhl, John, and Caleb Melby. "Trump Says 'Can't Have a Conflict of Interest' as President." *Bloomberg.com*. Bloomberg, 22 Nov. 2016. Web. 17 Jan. 2017.

One Ethics to Rule Them All?

There are often calls for "consensus" on ethics. But this is an unlikely result.

How does society decide?

Not by reason. Not by argument.

Not based on number of people. One egoist can undo the efforts of many communitarians (creating the effect known as the 'tragedy of the commons').

Universality

- "The 12 Guiding Principles of Exceptional Character," developed by the International Center for Leadership in Education, are *adaptability, compassion, contemplation, courage, honesty, initiative, loyalty, optimism, perseverance, respect, responsibility, and trust-worthiness* (www.leadered.com/guiding_princ.html).
- "The Seven Universal Ethical Attributes," developed by the Heartwood Institute, are *courage, loyalty, justice, respect, hope, honesty, and love* (<http://heartwoodethics.org/1-approach/framework.asp>). The Heartwood website also includes links to "Other Ethical Attributes" and "Other Ethical Frameworks."

Via:

<http://www.ascd.org/publications/educational-leadership/feb11/vol68/num05/Character-Education-for-the-Digital-Age.aspx> (Character Education for the Digital Age, Jason Ohler, 2011, ASCD)

“Universalist arguments apply across lines of social and economic class, but their generality is not well suited to combating unequal and particular harms. Members of our most disadvantaged communities have long asserted that they fall under heavier and more hostile scrutiny, and empirical research supports them. Meaningful oversight of state surveillance cannot be designed with only common conditions in mind.”

(Gellman and Adler-Bell, 2017)

Killing

Vs Murder

Lying

<https://journals.sfu.ca/jalt/index.php/jalt/article/view/293/253>

Spying

Axiology

<https://www.slideshare.net/robertfarrow/research-methods-in-open-education-insights-from-the-global-oer-graduate-network>

Ethical Universalism

Moral universals: A machine-reading analysis of 256 cultures

<https://public.tableau.com/profile/mark.alfano#!/vizhome/Virtuesandvicesfromtheperspectivesof256cultures/Virtuesandvicesaroundtheworld>

Ethics and morality as social phenomena

Metcalf, Moss & Boyd contrast 'ordinary ethics', ie., ethics as it is traditionally done, with ethics as a social phenomenon.

From: *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics* by Jacob Metcalf, Emanuel Moss and Danah Boyd ([14 page PDF](#)).

<https://datasociety.net/library/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/>

a central lesson of the ordinary ethics model is that ethics as practice is foundationally a tension between the everydayness of the present and the possibility of a different, better everydayness (Keane 2017). This dynamic was palpable in interviews, such as when, almost with a sense of exasperation, an engineer remarked, I don't think there are enough people talking about ethics in the tech industry.... There are people who are impacted by technology who talk about the tradeoffs of technology and how it has affected their lives or their communities.... There are people who have made a career out of being disaffected tech people who regret what they built [even though they] profited from it. And there are those who are coming from the outside—"professional ethicists or philosophers." Yet, even though these voices are loud and growing, from her perspective, the vast majority of people in tech "are not yet moved by ethics."

Zigon (2007 and 2014) argues that while ordinary ethics ethnography is useful for locating ethics in a culture, this method struggles to recognize when ethical claims are made as a way to change a culture, as happened in crises like Cambridge Analytica. He emphasizes the importance of the moment of “moral breakdown,” where a mode of life becomes untenable, which opens the possibility of hermeneutic interrogation of everyday moral frameworks. This echo work by Dave (2012) and Mahmood (2012) demonstrating how everyday ethical practices—such as queer activism or forms of religious piety—are leveraged as a form of critique against dominant modes of morality and political power. Establishing a new mode of everyday ethical practice is necessary—because no one can live in a state of breakdown forever—and ideally would concretize a more just and fulfilling mode of life. As Zigon (2007) puts it, “ethics is the process of once again returning to the unreflective mode of everyday moral dispositions. But this return from the ethical [breakdown] moment is never a return to the same unreflective moral dispositions.”

Disciplinary Ethics

Scientific Virtues

Theoretical Virtues in Scientific Practice: An Empirical Study Moti Mizrahi

<https://philpapers.org/archive/MIZTVI.pdf>

It is a common view among philosophers of science that theoretical virtues (also known as epistemic or cognitive values), such as simplicity and consistency, play an important role in scientific practice. In this paper, I set out to study the role that theoretical virtues play in scientific practice empirically. I apply the methods of data science, such as text mining and corpus analysis, to study large corpora of scientific texts in order to uncover patterns of usage. These patterns of usage, in turn, might shed some light on the role that theoretical virtues play in scientific practice. Overall, the results of this empirical study suggest that scientists invoke theoretical virtues explicitly, albeit rather infrequently, when they talk about models (less than 30%), theories (less than 20%), and hypotheses (less than 15%) in their published works. To the extent that they are mentioned in scientific publications, the results of this study suggest that accuracy, consistency, and simplicity are the theoretical virtues that scientists invoke more frequently than the other theoretical virtues tested in this study. Interestingly, however, depending on whether they talk about hypotheses, theories, or models, scientists may invoke one of those theoretical virtues more than the others.

Business Ethics

From: Owing Ethics: [Corporate Logics, Silicon Valley, and the Institutionalization of Ethics](#) by Jacob Metcalf, Emanuel Moss and Danah Boyd ([14 page PDF](#)).

<https://datasociety.net/library/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/>

Given the conceptual and organizational strain that ethics poses for the tech industry, what should

the work of ethics owners consist of? What should the doing of ethics look like? An obvious place to turn

would be applied business ethics research literatures—after all, contemporary tech corporations are not the

first companies to face ethical dilemmas. Business ethics literatures typically index normative concerns to the

need for coordination between individual behavior and organizational goals. For example, Sims (1991)

describes a primary method for institutionalizing ethics within organizations as the mutual understanding

and alignment of organizational expectations and individual behavior. Other top priorities Sims outlines focus

on the personal psychological needs of employees to receive adequate attention and rewards, identifying

employees with their work so they are publicly accountable for ethical lapses, and making public and internal

expectations clear.

The alignment of personal psychological states and explicit organizational priorities is a major theme

of business ethics. For example, Trevino (1986) shows how employees respond to ethically ambiguous

situations in terms of their own stage of cognitive-moral development, and Fleischman et al. (2017)

experimentally demonstrate that managers are less likely to intervene in formally unethical behavior of

subordinates when they perceive those behaviors as financially benefiting the organization.

Silicon Valley Ethics

Ethics in Silicon Valley

Signe Agerskov, Data Ethics, 2020/02/24

Icon

Signe Agerskov summarizes *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics* by Jacob Metcalf, Emanuel Moss and Danah Boyd (14 page PDF). This paper, which is quite good and well worth reading in its own right, argues that the technology industry views ethics as a technological problem. As Agerskov summarizes. "ethics is seen by many tech workers as something that arises from imperfect products and is not understood in a social context as something that structures social life. For tech people the apparent solution to ethical problems is therefore technology in the form of improved products, not changes of fundamental structures in the organization or industry." This is not the first time this has happened; as the authors argue, business ethics is designed the same way. "Business ethics literatures typically index normative concerns to the need for coordination between individual behavior and organizational goals."

Web: [Direct Link] [This Post] <https://www.downes.ca/post/70529>

Alkso:

ethics in the tech industry is based on the underlying logic of meritocracy, technological solutionism and market fundamentalism. Meritocracy is a form of government that argues that power should be given to those with the highest skills and qualifications. Technological solutionism trusts innovation and technological solutions to solve broad social problems and

market fundamentalism reason that a free and unregulated market can solve social and economic problems.

From: Owing Ethics: [Corporate Logics, Silicon Valley, and the Institutionalization of Ethics](#) by Jacob Metcalf, Emanuel Moss and Danah Boyd ([14 page PDF](#)).

<https://datasociety.net/library/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/>

Yet our ethnographic fieldwork, interviews, and textual sources indicate that as ethical product design and governance goals are becoming institutionalized by tech firms, the practices associated with these

goals are being crafted and executed according to the existing logics and structures of the technology industry,

even as they are responding to outside critiques of these logics and structures. In doing so, they are producing

pitfalls that threaten to prematurely foreclose what can be thought or done under the heading of “ethics.” As

companies hasten towards practices that can detect crises before they erupt, they also attempt an updated

return to “business-as-usual.” This is not to say that business-as-usual cannot be ethical; rather, we point to

the potential for the pitfalls introduced by tech logics and organizational structures to prevent the work of

ethics owners from addressing the broader goal of a more just algorithmic and data-driven world.

Ethics Of...

Educational Ethics

Theories?

Ethics of Technology

Have to mention Dreyfus here

Neuroethics

<http://philsci-archive.pitt.edu/19734/1/veitbrowning2021neuroethics.pdf>

The End of Ethics?

The Golden Rule

The thing I read in HBS that bothered me: “The CEOs in the case study need to take the “front page” test: If the headline on the front page of the newspaper were reporting abuse of customer data (yours included), how would you react? If you wouldn’t want your personal data used in a certain way, chances are your customers wouldn’t, either.”

But what if the CEO in question is Mark Zuckerberg, and genuinely does not care?

Thomas H. Davenport and Jeanne Harris. (2007). The Dark Side of Customer Analytics. Harvard Business Review. May, 2007.

<https://hbr.org/2007/05/the-dark-side-of-customer-analytics>

Ethics as a Branch of Something Else

Ethics and Social Structure

Amartya Sen

In 1981, Sen published *Poverty and Famines: An Essay on Entitlement and Deprivation* (1981), a book in which he argued that famine occurs not only from a lack of food, but from inequalities built into mechanisms for distributing food. Sen also argued that the Bengal famine was caused by an urban economic boom that raised [food prices](#), thereby causing millions of rural workers to starve to death when their wages did not keep up.^[19]

Sen's revolutionary contribution to development economics and social indicators is the concept of "[capability](#)" developed in his article *Equality of What*.^[22] He argues that governments should be measured against the concrete capabilities of their citizens. This is because top-down development will always trump [human rights](#) as long as the definition of terms remains in doubt (is a "right" something that must be provided or something that simply cannot be taken away?). For instance, in the United States citizens have a right to vote. To Sen, this concept is fairly empty. In order for citizens to have a capacity to vote, they first must have "functionings". These "functionings" can range from the very broad, such as the availability of education, to the very specific, such as transportation to the [polls](#). Only when such barriers are removed can the citizen truly be said to act out of personal choice. It is up to the individual society to make the list of minimum capabilities guaranteed by that society. For an example of the ^[23]

In 2009, Sen published a book called [The Idea of Justice](#).^[41] Based on his previous work in welfare economics and social choice theory, but also on his philosophical thoughts, Sen presented his own theory of justice that he meant to be an alternative to the influential modern theories of justice of [John Rawls](#) or [John Harsanyi](#). In opposition to Rawls but also earlier justice theoreticians [Immanuel Kant](#), [Jean-Jacques Rousseau](#) or [David Hume](#), and inspired by the

philosophical works of [Adam Smith](#) and [Mary Wollstonecraft](#), Sen developed a theory that is both comparative and realisations-oriented (instead of being transcendental and institutional).

Wikipedia: https://en.wikipedia.org/wiki/Amartya_Sen

Critical Race Theory

Roy L. Brooks defined critical race theory in 1994 as "a collection of critical stances against the existing legal order from a race-based point of view".^[24] More specifically, race is a [social construct](#) and racism is neither an individual bias nor prejudice, but rather embedded in the legal system and supplemented with policies and procedures.^[25]

Critical race theory (CRT) is a body of legal scholarship and an academic movement of US [civil-rights](#) scholars and activists who seek to critically examine the intersection of [race](#) and [US law](#) and to challenge mainstream [American liberal](#) approaches to [racial justice](#).^{[1][2][3][4]} CRT examines social, cultural, and legal issues primarily as they relate to race and [racism in the US](#).^{[5][6]} A tenet of CRT is that racism and disparate racial outcomes are the result of complex, changing, and often subtle social and institutional dynamics, rather than explicit and intentional prejudices of individuals.^{[7][8]}

CRT originated in the mid-1970s in the writings of several American legal scholars, including [Derrick Bell](#), Alan Freeman, [Kimberlé Crenshaw](#), [Richard Delgado](#), [Cheryl Harris](#), Charles R. Lawrence III, [Mari Matsuda](#), and [Patricia J. Williams](#).^[1] It emerged as a movement by the 1980s, reworking theories of [critical legal studies](#) (CLS) with more focus on race.^{[1][9]} CRT is grounded in [critical theory](#)^[10] and draws from thinkers such as [Antonio Gramsci](#), [Sojourner Truth](#), [Frederick Douglass](#), and [W. E. B. DuBois](#), as well as the [Black Power](#), [Chicano](#), and [radical feminist](#) movements from the 1960s and 1970s.^[1]

CRT scholars view race and [white supremacy](#) as an intersectional [social construct](#)^[7] that advances the interests of [white people](#)^[11] at the expense of persons of other [races](#).^{[12][13][14]} In the field of legal studies, CRT emphasizes that formally [colorblind](#) laws can still have [racially discriminatory](#) outcomes.^[15] A key CRT concept is [intersectionality](#), which emphasizes that race can intersect with other identities (such as gender and class) to produce complex combinations of power and advantage.^[16]

Academic critics of CRT argue that it relies on [social constructionism](#), elevates storytelling over evidence and reason, rejects the concepts of truth and merit, and opposes liberalism.^{[17][18][19]} Since 2020, conservative US lawmakers have sought to ban or restrict critical race theory

instruction along with other anti-racism programs.^{[8][20]} Critics of these efforts say the lawmakers have poorly defined or misrepresented the tenets and importance of CRT and that the goal of the laws is to more broadly silence discussions of racism, equality, social justice, and the history of race.^{[21][22][23]}

Wikipedia - https://en.wikipedia.org/wiki/Critical_race_theory

Structural Discrimination

Why Hard Work isn't Enough to Get By - BBC -

<https://www.bbc.com/worklife/article/20210921-why-hard-work-alone-isnt-enough-to-get-ahead>

“You need to be liked: by people at your level, by people above you and by people below you. When you evaluate work, the research is quite clear – [people who are liked get better ratings](#), even if their work is the same.” And therein lies the unfair truth: you and a colleague may have the exact same skills and work ethic, but if they’ve spent more time making friends and influencing people, they’ll look better at their job.

Military leaders saw pandemic as unique opportunity to test propaganda techniques on Canadians, Forces report says. A plan devised by the Canadian Joint Operations Command relied on propaganda techniques similar to those employed during the Afghanistan war. David Pugliese • Ottawa Citizen Sep 27, 2021

<https://ottawacitizen.com/news/national/defence-watch/military-leaders-saw-pandemic-as-unique-opportunity-to-test-propaganda-techniques-on-canadians-forces-report-says>

Pedagogy of Oppression - Paulo Friere

Could a dedicated disinformation beat help address the infodemic in Canada?

By Emma Buchanan - J-Source - 2021-09-24

<https://j-source.ca/could-a-dedicated-disinformation-beat-help-address-the-infodemic-in-canada/>

Code and Other Laws

Lawrence Lessig

<https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/>

The Facebook whistleblower says its algorithms are dangerous. Here's why.

Frances Haugen's testimony at the Senate hearing today raised serious questions about how Facebook's algorithms work—and echoes many findings from our previous investigation.

By Karen Hao October 5, 2021

Frances Haugen testifies during a Senate Committee

Facebook whistleblower Frances Haugen testifies during a Senate Committee October 5. Drew Angerer/Getty Images

During her testimony, Haugen particularly blamed Facebook's algorithm and platform design decisions for many of its issues. This is a notable shift from the existing focus of policymakers on Facebook's content policy and censorship—what does and doesn't belong on Facebook. Many experts believe that this narrow view leads to a whack-a-mole strategy that misses the bigger picture.

"I'm a strong advocate for non-content-based solutions, because those solutions will protect the most vulnerable people in the world," Haugen said, pointing to Facebook's uneven ability to enforce its content policy in languages other than English.

Haugen's testimony echoes many of the findings from [an MIT Technology Review investigation](#) published earlier this year, which drew upon dozens of interviews with Facebook executives, current and former employees, industry peers, and external experts. We pulled together the most relevant parts of our investigation and other reporting to give more context to Haugen's testimony.

[@bonstewart Oct 3](#)

<https://twitter.com/bonstewart/status/1444823495335190535>

the design of a tool shapes how it's used. if DeltaMath is designed so that teachers can conflate compliance with perfection, that feature will be used & abused on the regular until you get rid of it. onus is on YOU to have thought this through. some teachers don't. some won't.

Also: <https://twitter.com/MrDeltaMath/status/1444482304802533377>

The Joker Problem

“Some men just want to watch the world burn.” (*The Dark Knight*, Nolan & Nolan, 2008)

Ethics, as we have come to understand it, is based on what the reasonable person would do, given the knowledge and insight. But what if people are not reasonable?

We have a presumption that people act on comprehensible motives, and that these motives can provide an account not only of ethical behaviour, but also of the structure of systems and methods that define ethics, legality, and possibility.

Here's an example. In the world of decentralized networks, there's a problem called the Byzantine Generals Problem. **ref** The idea is that the Byzantine Emperor is receiving information from his generals, but that the generals can't be trusted, because they may want to overthrow the emperor. How then to build trust into a distributed system?

In blockchain networks, the analogous situation exists when a member of the system may be tempted to report extra income for himself, thereby cheating the system? How, again, can trust be built into the distributed system? The solution, as instantiated in blockchain networks, is to make it too expensive to cheat the system. The would-be cheater would have to spend more money to rebuild the blockchain than he would earn as a result of the deception. **ref**

This solution works in the presumed cultural, social and ethical context of financial networks, where the motive is profit. But as Alfred says in *The Dark Knight*, “some men aren't looking for

anything logical, like money. They can't be bought, bullied, reasoned or negotiated with. Some men just want to watch the world burn.”

Nothing prevents a person from spending the money just to see the blockchain burn. What would make such an action unethical? Is it because they don't like blockchains? Is it because they don't seek money? Is it because they are unreasonable? Given different *motives*, any action can seem reasonable. Any action can seem ethical. Maybe, from a certain point of view, the blockchain *should* burn.

Maybe Ethics Is a Technical Problem?

event, and then expand. A number of workshop attendees volunteered to kick this process off.

5.3 #EIAI2020 5.3.1 What is it? There is a need for a continued open discussion and knowledge-sharing relating to applied AI ethics. The success of the Ottawa and London events ('Ethics in AI 2019', #EIAI2019), which were both over-subscribed, suggests that a rerun in 2020 would be welcome, and would have the additional benefit of acting as a method to motivate and track progress against the aspiration of moving from principles to practice in AI ethics. 5.3.2 How would it work? The calendar is filled with events relating to AI ethics, but a focus on applied AI ethics is much less well-served, while the use of a networking event to build a consensus around actionable next steps currently seems to be unique. The events could be run independently again, and funded through sponsorship. Alternatively, to achieve a wider audience or to marry up complementary content, the event series could be co-located with other established academic or industry conferences, or run as a workshop stream inside one of them.

5.4 Competition to stimulate innovation 5.4.1 What is it? There is a need to incentivise innovation in applied AI ethics. Competition is one way to achieve this, and there have already been a number of successful examples targeted at evaluation of the ethical aspects of AI systems, such as:

- The Gendered Pronoun Resolution competition hosted on Kaggle, on behalf of the competition sponsor Google AI. This competition offered \$25,000 in prizes for solutions addressing gender-bias in pronoun resolution . 34
- The NeuRIPS 2018 Adversarial Vision Challenge (Robust Model Track) pitted machine vision models against adversarial attacks. The prize was \$15,000 of Paperspace cloud compute credits, while the top 20 teams in each track received \$250. The competition attracted 382 participants and 1,995 submissions . 35
- The NeuRIPS 2018 Inclusive Images Challenge involved developing models that address distributional skew and increase global inclusion. This challenge was hosted on Kaggle and sponsored by Google (in partnership with NeurIPS) with a prize pot of \$25,000, and attracted 109 teams . 36

ender and race bias was incorporated as an additional evaluation criterion 37 alongside the regular performance-oriented SemEval emotion prediction 38 competition. Competition has historically had a great impact on bringing the attention of the research

community to a particular problem and in attracting new researchers to the areas. Created datasets are also a valuable resource for data-hungry machine learning fields.

34 <https://www.kaggle.com/c/gendered-pronoun-resolution>

35 <https://www.crowdai.org/challenges/adversarial-vision-challenge>

36 <https://www.kaggle.com/c/inclusive-images-challenge/>

37 Kiritchenko, S. and Mohammed, M, S. Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems, 2018 arXiv e-prints arXiv:1805.04508

38 <https://competitions.codalab.org/competitions/17751>

(deBruijn, et.al., 2020)

Or a Commercial Market Waiting to Develop?

Applied AI Ethics Supplier presentations:

- Accenture Fairness Tool: Caryn Tan, Digital Strategy Consultant from Accenture, 39demonstrated the Accenture Fairness Tool, which enables a user to identify and fix some of the problems that result in unfair outcomes by analysing both training data and models. It also enables users to understand the trade-offs they may be making between model accuracy and fairness.
- Ellpha: Stephanie Creff, CEO, spoke about the Ellpha Bias Detection EngineTM40that identifies gender-biased language by analysing text, providing feedback and making suggestions for more neutral language. Ellpha is currently focusing on deploying their product in the HR environment, and is undertaking research to detect bias in datasets.
- Hazy: Alice Piterova, Head of Privacy, presented the company's work in synthetic 41data innovation. She addressed the new challenges and scenarios that organisations are facing in an increasingly data-reliant age. Hazy generates synthetic data that provides companies with the ability to enhance privacy, address class imbalance, act as a benchmarking tool and enable forecasting and simulation of future events.
- OpenMined: Andrew Trask presented the work of OpenMined, a community 42focused on researching, developing and elevating open-source tools for secure, privacy-preserving and value-aligned artificial intelligence. Andrew described three techniques that can be combined to

keep both data and models private: federated learning, secure multi-party computation, and differential privacy - all of which are implemented in OpenMined's PySyft library .

39 <https://www.accenture.com/gb-en/blogs/blogs-cogx-tackling-challenge-ethics-ai>

40 <https://www.ellpha.com/>

41 <https://hazy.com/>

42 <https://www.openmined.org/>

43 <https://github.com/OpenMined/PySyft>

(deBruijn, et.al., 2020)

Complexity

Simple universal principles are not sufficient to address complex problems. As Jones (2011) writes, "Implementation must deal with interdependent problems, navigating nonlinear and often unpredictable change processes, involving a diverse range of stakeholders."

First, he writes, "the capacities to tackle complex problems are often distributed among actors." While we may desire to ascribe agency and assign responsibility, many problems inhabit what are in fact complex systems. Who is responsible for the current inflation rate? Where do we assign responsibility for global poverty? What factors account for a class average score of 72?

Second, the outcomes of complex systems are unpredictable. "Where causality is not well understood success may rely on adaptation and flexibility to emerging insights, rather than trying to completely fix the shape of policy responses in advance." Douglas Rushkoff described this capacity as akin to riding a surfboard, where you cannot plan in advance, but rather must adapt to rapidly changing circumstances.

"The ocean is a chaotic dynamical system with sensitive dependencies on initial conditions. Macro info keeps being folded in while micro info keeps being excavated." This process also gives a sense of the stance we need to take. "The only real weapon against the fearful vision of

a cold siber-Cyberia is joy. Appreciation of the space gives the surfer his bearings and balance in Cyberia.” (Rushkoff, 1994:180-182).

And third, complex problems often involve conflicting goals. We don't all seek “truth, justice and the American way.” Some of us are happy with “peace, order and good government.”

Change

Things don't stay the same. Fundamental values can be revised and change over time. Indeed, some ethical codes stipulate that review and revision ought to be built into the standard.

A good example of how things change can be found in the discussion of privacy. What does it mean, really? Ware, et.al. (1973: 39-40) make the case:

“Once we recognize these characteristics of records, we must formulate a concept of privacy that is consistent with records. Many noteworthy attempts to address this need have been made. The significant elements common to these formulations are (1) that there will be some disclosure of data, and (2) that the data subject should decide the nature and extent of such disclosure.... None accommodates the observation that records of personal data usually reflect and mediate relationships in which both individuals and institutions have an interest, and are usually made for purposes that are shared by institutions and individuals.”

Similarly,

“it is equally out of keeping with the mutuality of record-generating relationships to assign the institution a unilateral role in making decisions about the content and use of its records about individuals. Yet it is our observation that organizations maintaining records about people commonly behave as if they had been given such a unilateral role to play.” (Ibid: 40)

It may perhaps be an exaggeration to say that the ethics of privacy come in and out of fashion over time, but it is not an exaggeration that they are continually subject to change, and what held in the distant past may not hold today or in the future.

The End of Ethics

At a certain point, it becomes necessary to rethink the purpose of ethics as an activity in and of itself. What is accomplished by ethical theories and codes of ethics in the context of a diverse, changing and complex world?

The tone and attitude of both the theories and sets of principles is a tone of preventing wrong-doing and of the stipulation of what constitutes right and wrong practices. The purpose of the principles and rules is to set a bounds of acceptable action, a limit beyond which an ethical person shall not transgress.

“People tend... to think in terms of rules, and that may well be the reason why they shy away from ethics,” says James O’Toole. It’s easy to stipulate a set of rules, but ultimately, ethics isn’t about rules. As O’Toole argues, “it’s a way of thinking about how you conduct your life and do your job.” (Feffer. (2017)

Ethics are not abstract. Though we can derive universal principles based on abstract values or calculations, we shouldn’t. Ethics are not based on reason or universal principles, they are based on real people in concrete situations. The person matters, the other people involved matter, and the ethical context matters.

So - what is that ‘way of thinking’? That is the subject of the next section of this article.

<https://philpapers.org/archive/PORMAP-3.pdf>

Morality and Practical Reasons

“We caution readers against inferring that, in any individual principles document, broader coverage of the key themes is necessarily better. Context matters. Principles should be understood in their cultural, linguistic, geographic, and organizational context, and some themes will be more relevant to a particular context and audience than others. Moreover, principles are a starting place for governance, not an end. On its own, a set of principles is unlikely to be more than gently persuasive. Its impact is likely to depend on how it is embedded in a larger governance ecosystem, including for instance relevant policies (e.g. AI national plans), laws, regulations, but also professional practices and everyday routines” (Fjeld, et.al., 2020).

“Action theory and ethics tend to be distinct academic industries today. But that used not to be true. Major works on ethics once combined an account of ethics and of ethical standards with detailed theories of action and the various psychological capacities and powers connected with it.”

“Power and the metaphysical questions it raises is thankfully becoming of greater interest to contemporary philosophers of science and nature. But it needs to be attended to by moral philosophers as well.”

“One such power might be self-determination, as a power of freedom or free will leaving it up to us how we act. But there could be other ethically crucial forms of power – such as the power of reason to move us. Moral philosophers today tend to treat normativity as identical with reason. Ethical standards are taken to provide reasons for being moral, and in so far as we are rational, it is assumed that we can be moved by the reasons they provide. As Christine Korsgaard stated, ‘Thus it seems to be a requirement on practical reasons, that they be capable of motivating us’.

Now if reasons can move or guide us, that presupposes some sort of power that they exercise over us. They must have the capacity to produce in us as outcomes the attitudes and the actions that they justify. But you no more find contemporary moral philosophers investigating the nature of this power than you find them arguing about the metaphysics of self-determination.”

“Modern ethics tends on the whole to avoid overt commitment to free will. But it is riddled with metaphysically unexplained claims about reasons and our responsiveness to them – about, in other words, reasons as exercising a power to move.”

(Pink, 2020)

The Structure of Moral Revolutions

Reviewed in:

Moral Revolution, Moral Reform and Moral Drift

<https://philosophicaldisquisitions.blogspot.com/2020/01/moral-revolution-moral-reform-and-moral.html>

Can we even eliminate bias in AI? --

<http://www.internetactu.net/2020/01/20/peut-on-corriger-les-biais/> --

<https://translate.google.com/translate?sl=auto&tl=en&u=http%3A%2F%2Fwww.internetactu.net%2F2020%2F01%2F20%2Fpeut-on-corriger-les-biais%2F>

"It is much easier to repair a camera that does not record dark skin than to repair a photographer who does not see people with dark skin"

"Do we need to develop these data-intensive technologies, the efficiency of which is questionable, while investment and development in these technos come at the expense of other solutions?"

"Ethics, on its own, won't make tech more responsible."

From <http://www.internetactu.net/2018/11/27/lethique-ne-suffira-pas/> : (trans):

<https://translate.google.com/translate?sl=auto&tl=en&u=http%3A%2F%2Fwww.internetactu.net%2F2018%2F11%2F27%2Flethique-ne-suffira-pas%2F>

A remark that echoes, in particular, that held by [David Banks](#) ([@da_banks](#)) in [the insolent The Baffler](#) . Banks emphasized how the world of engineering is linked to authoritarianism. The reason is certainly to be sought from the side of history. The first engineers were of military origin and designed siege weapons, he quickly recalls. They are also always trained to "connect with the decision-making structures of the chain of command" . Unlike doctors or lawyers, engineers have no ethical authority to supervise them or an oath to respect. "This is why engineers excel in outsourcing reproaches" : if it does not work, the fault lies with all the others: users, managers ... "They are taught from the start that the most moral thing they can do is build

what they are told to build to the best of their ability, so that the user's will is carried out with precision and fidelity. "

As Rachel Coldicutt writes, "A more transformative change both for boards and for product teams will be changing who makes decisions and how those decisions are made. After all, responsibility is a team sport and "ethics" are not a checklist. In the short term, this means product teams need help to embrace distributed decision making and make room for social-science expertise. To put it bluntly: computer scientists and software engineers need to not only learn the basics of ethical decision-making, they also need to work in environments that encourage trust and respect in non-technical or less-technical opinions. It's not just who's in the team that counts; it's who makes the decisions." (Coldicutt, 2018).

COVID-19 and the Impossibility of Morality --

<https://philosophicaldisquisitions.blogspot.com/2020/03/covid-19-and-impossibility-of-morality.html>

"We are now in the midst of a global health crisis of historic proportions. People are dying on a daily basis because of this rapidly spreading disease. We are forced to make choices we don't want to make; to trade lives and values off against each other. The speed at which we are forced to decide is what makes the tragedy so apparent. Whenever we confront clearly limited resources, both temporal and physical, we can perceive moral conflicts in a starker way. Ordinarily, we live with the illusion that we can postpone hard choices or reverse/correct them if we make a mistake. In the midst of a crisis, this illusion is unsustainable."

Weinberger, 2021

<https://www.hyperorg.com/blogger/2021/01/11/parler-and-the-failure-of-moral-frameworks/>

"Utilitarians might argue that the existence of Parler as an amplifier of hate threatens to bring down the overall happiness of the world. Of course, the right-wing extremists on Parler would argue exactly the opposite, and would point to the detrimental consequences of giving the monopoly platforms this power. I don't see how either side convinces the other on this basis.

"Deontologists might argue that the de-platforming violates the rights of the users and readers of Parler. the rights threatened by fascismOther deontologists might talk about the rights threatened by the consequences of the growth of fascism enabled by Parler. Or they might simply make the utilitarian argument. Again, I don't see how these frameworks lead to convincing the other side."

My view: Parler is a radical social network website sufficiently offensive to have been banned from most platforms and hosting services. David Weinberger's point in this post is to argue that traditional ethical frameworks do not give us a good argument for deplatforming Parler. The suggestion is, if you take on the perspective of 'the other side', there isn't a good argument that would convince them that they are wrong. But (in my view) this sort of approach is what Robert Nozick called 'coercive philosophy': "arguments are *powerful*, and best when they are *knockdown*, arguments *force* you to a conclusion." This just feeds into the ethos of sites like Parler; it draws us into engagement and battle with the beast. No. We as ordinary citizens living