

Computer Ethics

Elia Ravella

October 25, 2021

Contents

1	Ethics	2
1.1	Ethical Theory	2
1.1.1	Utilitarianism	2
1.1.2	Duty Ethics	3
1.1.3	Virtue Ethics	3
1.2	Normative Argumentations	3
1.2.1	Argumentation by Analogy	3
1.2.2	Argumentation in a Utilitarian Framework	3
1.2.3	Argumentation in Kantian Reasoning	4
1.2.4	Argumentation in Virtue Ethics	4
1.2.5	Fallacies	4
2	Responsibility	5
2.1	Moral and Professional Responsibility	5
2.2	Passive / Backward Looking Responsibility	5
2.3	Active Responsibility	5
2.4	Ideals	5
2.5	Engineers vs Managers	6
2.5.1	Engineers vs ?	6
2.6	Technology Assessment	6
3	Moralizing Technologies	7
3.1	"Do Artifacts have Politics?"	7
3.2	Technological Mediation	7
3.2.1	Embedding Morals in the Environment	7
4	Computer Ethics	8
4.1	The Invisibility Factor	8
4.1.1	IT-configured Activities	8
4.1.2	IT-configured Domains of Life	8

1 Ethics

There is not a single unique definition for ethics, but all the most used ones usually falls around *the science that considers what is good and what is bad*. The study of ethics, so the determination of what is *objectively good*, is done in order to find a way to "live life well" and so improve the life of a single person, and the community around him. Obviously, these are not objective themes, as they're influenced by the single vision of what is good and what is bad, the values and limits and education and culture of the single. Ethics is the systematic reflection of what is moral; morality is the set of opinions, decisions, actions that are taken by people and that express what is good or right. So: ethics is not a manual with answers, rather than a **process** of searching for the right kind of morality.

1.1 Ethical Theory

We distinguish between two main ethical theories:

1. Descriptive ethics: this branch describes the *existing* morality
2. Normative ethics: this branch *judges* morality and proposes norms in order to behave better

This is not a clear division of theories: how much a norm is a strong position or only a suggestion, or viceversa: when does a description highlights "good" or "bad" behaviours, *neutrally*? Both branches carry the associated idea of "judgement": the act of *stating a fact* about something. Distinguishing between normative and descriptive judgement is again far from easy.

Values Values can be described as the lasting convictions or matters that people feel to be strived in general and not just for themselves, in order to leave a good life in a fair society. We can distinguish values in

- intrinsic values: the value held *naturally* by something
- instrumental values: value that exists because of the *values it generates*

The same thing can have both, at different eyes.

Virtues Virtues are *human qualities*. Virtues are "intrinsically positive" because they are necessary to live a good life.

Virtues can be moral (like justice, honesty, courage) other intellectual (as integrity). Most of virtues are *also* values, but the main difference is that virtues are tailored to the people that "practice" them, while values are more detached from the people themselves and are to be achieved.

Norms Norms are rules: set of regulations that are used to organize society. Norms are strongly related to values also because *some values can be directly translated into norms*. Norms can be seen as **means to realize values**.

1.1.1 Utilitarianism

Utilitarianism is a type of consequentialism based on the utility principle. Actions are judged only by the amount of pleasure and pain they *provoke*. Big problem: if actions are to be chosen only by the crude amount of value they bring to the majority of people how do we select that majority? What happens to the "condemned" minority?

A first appearance of utilitarianism can be seen in Hedonism, where pleasure is considered the *only intrinsic value*, while all others are just instrumental. Actions should be weighted basing on how much pleasure they bring to how many people, maximizing the ones that affects more people.

One of the most clear problem of this theory is: how do we assign a value, a measure to happiness, to pleasure? Every person reaches them in a different way. Utilitarianism "lacks of humanity". Utilitarianism also can lead to *exploitation*: the utilitarian motivations can be used to justify the overexploitation of groups of people or resources. Moreover, utilitarianism bases his judgement system on the assumption that every action has *clear and foreseeable consequences*: this is a bold assumption. What if something goes wrong, or a prediction just lacks of context?

1.1.2 Duty Ethics

Duty ethics focuses on the notion of *duty*, without caring for the consequences. The action are to be evaluated with respect to the motives behind them and the moral value of them, instead of basing on the consequences they provoke. The value or weight of an action is so relevant only when it's compared to a set of moral rules, that can be religious, social or even made up with sufficient argumentations.

One of the most famous duty ethics formulations is the Kantian one:

- Universality principle: every action should be taken only if the person taking it thinks that that way of doing should be what is mandatory to do, a universal right way to doing so
- Reciprocity principle: treat humans always as ends, never as means

1.1.3 Virtue Ethics

Virtues, as *human good qualities* such as integrity, courage, wisdom etc are seen as the only way to thrive, to reach happiness, in virtue ethics. This is different at the core from duty ethics: while for Kant the morality is a matter of "doing the right thing", virtue ethics (such as utilitarianism, in a way) is about going towards happiness. The greater good for virtue ethics is *a good life*, while for duty ethics is *the goodwill*.

1.2 Normative Argumentations

To argument is the act of justify or refute a statement. An argument is a consequential set of statements where the last follows the others logically. The argument wants to reach a **conclusion** from the **premises**. A **valid argument** is an argument whose conclusion follows with necessity from the premises.

Non Deductive Arguments Often, the premises aren't "stable"¹ and so we can change the conclusions. In non deductive arguments usually **the conclusion is logically stronger than the premises**. Non deductive reasoning is always uncertain: a critical question (one that aims to check the **degree of palusibility** of a statement) cannot always be answered positively.

1.2.1 Argumentation by Analogy

Used in technology to contextualize new technologies. Analogy argumentation wants to fill a judgement vacuum around something by providing a "similar" situation where the judgement is clear or already been taken.

Critical questions plays a huge role in the analogy analysis: questions like

- are the situations comparable?
- are there core differences?
- the importance and scale of the analized situation is the same?

When the differences are higy relevant, an analogy is said to be *false*.

1.2.2 Argumentation in a Utilitarian Framework

The classic formulation is the means - end argumentation: it's a type of non deductive argumentation in which *from the conclusions, the means are derived*, leading to the moral justification of some immoral actions.

Critical questions:

- Does an action *indeed realize* the given end?
- Can the considered action be *actually carried out*?
- What about side effects?
- is the end *acceptable*?

¹non monotonicity

1.2.3 Argumentation in Kantian Reasoning

The typical argumentation in a Kantian line of reasoning is proof by absurd: the usual proof tries to transform a statement in a general law and then tries to obtain a contradiction out of that. As an example:

1. "I will not keep my promise" is acceptable if you need money
2. transform into maxim: "if I'm in need of money, I can break my promise"
3. This *justifies* to break promises \Rightarrow promises do not make sense anymore
4. contradiction

1.2.4 Argumentation in Virtue Ethics

The argumentations usually follows the "is this action what a virtuous person would do?". The problems here are many:

- Which virtue are we pursuing?
- How do we define a virtuous person?

Engineers Virtues Which are the virtues that should characterize a software engineer? A morally responsible engineer that focuses on *engineering practice* should have (or strive to reach)

- Expertise
- Communication
- Cooperation
- Objectivity
- Being open to criticism
- Quality oriented

as theorized on Pritchard, 2001.

1.2.5 Fallacies

When analyzing an argumentation, there are some typical fallacies:

- Decontextualize
- Attack on person
- Confusion of law and ethics: "if it isn't illegal, it's ethical"
- Whisful thinking: the interpretation of facts is based on personal beliefs rather than actual/rational evidence
- Privacy fallacy: "I got nothing to hide"

also, very used in the field of acceptability of technological aspects, there are *risk fallacies*:

- The sheer size fallacy: "you should accept nuclear energy because it's less risky than driving a car"
- The naturalness fallacy: "if it's unnatural, should not be done"
- Ostrich's fallacy: burying the head in the sand to avoid danger
- The delay fallacy: "if we wait we will clarify our knowledge, so no action for now"
- Technocratic fallacy: "this is engineering stuff, engineers should norm this"
- Pricing fallacy: risks are considered only as economic risks

2 Responsibility

2.1 Moral and Professional Responsibility

If we define responsibility as "being held accountable for your actions and the effects of them" we can then analyze two kind of responsibility that we deal with everyday:

1. **Moral responsibility** is the one based on the person moral, so his duties, norms, obligations
2. **Professional responsibility** is the one connected to the *role* as a professional the person is in. This can be applied in both formal (work) spaces and informal (family, faith communities) ones

2.2 Passive / Backward Looking Responsibility

The usual model that is applied (the most natural one for human society) when "looking for a responsible" is to act only after something undesirable occurred. Only at that moment a person to be held accountant is looked for, to justify some actions towards others (the jury). Backward looking responsibility also carries the meaning of *blameworthiness*: being a target to blame for one's actions or effects of actions. In order for someone to be blameworthy, we usually look for

1. Wrong doing: acting against the norm, be it the law or a moral norm
2. Causal contribution: not acting against the wrong doing
3. Foreseeability: being aware that some consequences are undesirable
4. Freedom of action: nothing stops the blamed person to act against "wrong" actions

2.3 Active Responsibility

Active responsibility is "preventing the negative effects of a new technology but also realizing certain positive effects". This boils down to the classic *think about the consequences* problem. We should consider, in this process:

- the perception of the responsible actor of the norms he's threatening, if so
- the autonomy of the creator: its freedom of thought and speech
- the role obligations of the role he's working into, and the code he's subject to

2.4 Ideals

Why ideals? Ideals set the aim and the boundaries of engineers, understanding them means understanding the behaviour of them and the personal limits and motivations behind each one of them.

Technological Enthusiasm The behaviour of *overseeing* political and security issues in order to provide what is considered a breakthrough technology is called technological enthusiasm. The engineer just consider his personal vision of the technology he's developing and *overlooks* social/environmental/security issues because he only focuses on the positive aspects of it. This is often called technosolutionism, or thinking that every problem can be solved through technological advancements.

Effectiveness and Efficiency Effectiveness and efficiency are usually depicted as objective (and even measurable) aims and ideals. These are NOT in fact neutral ideals: Taylorism (the movement which aim was to shape society as an industry) is a efficiency-driven philosophy.

Human Welfare "Human welfare" as an ideal is the one that drives engineers to create and contribute to technologies that augment human conditions.

The problem here is straightforward: which is the "right way" to augment human welfare? This can depend from a multitude of aspects, such as the engineer specialization, his background, his other ideals and the formation he received... In fact, this is NOT MORALLY NEUTRAL.

2.5 Engineers vs Managers

Big problem in corporations and agencies (NASA challenger disaster). The engineer figure has a *professional responsibility* towards his role and his company and a *moral responsibility* towards himself when dealing with critical missions/projects. There are several way to address the kind of tension that arises when these two morals are in conflict:

- Separatism: scientists and engineers are considered only as the "tech people" and acts only driven by the morals of the company. "Value decisions" should be taken from appropriate management organs. This way of handling the problem is based on the so-called Tripartite model², that states: the engineers can only be held accountant for design issues of the product, not for eventual larger social effects.
- Technocracy: "govern by experts". Only science people are allowed to govern in a certain field. Obviously this is *undemocratic and paternalistic*, because it proposes a model where only a tight minorance is allowed to take decisions that affects a multitude of people.
- Whistle blowing: disclosing to the public some information (often critical information) in order to remedy about some abuses or warn the public of such abuses or injustices. Whistle blowing is considere **morally required** (De George 1990) in case of
 - Harm to the public
 - Internal abuse ignored after reporting
 - Useless internal procedures
 - Evidence is present to warn about a threath
 - Revealing the threath prevents the harm

2.5.1 Engineers vs ?

Engineers are not the only ones responsible for the development and consequences of a technology.

- Developers and producers of a certain technology (among them engineers)
- Users: a lot of technologies are developed to fill a user need, that most of the time is *requested* by the users themeselves
- Regulators as political and governing organizations, but also standard enforcers

in this "responsibility graph" we could also insert trade unions, worker unions, educational institutes, professional associations...

2.6 Technology Assessment

TA is a systematic method for exploring future technologies and assessing their potential societal consequences. TA must be carried out before and during the development of a new pece of technology, in order to maintain contact with the reality in which such technology will operate adn the direction towards which the development itself is going.

TA and Contructive TA (so the "spiral" TA, where discussions and studies that are TA-like are carried out parallel to the design and the results are *fed back* to the dev and design process) are needed in modern industry; these should, however, live along the Collingridge dilemma:

- it is not possible to **fully** predict the effects and consequences the introduction of a new technology will generate
- when a negative effect of a technology materializes, it becomes more and more difficult to handle it if the dev process is gone far already

So we *need* to think about consequences in design phase, cooperating with actors with different POVs of the changes we're introducing³ and also be aware of the fact that the ones that "slips through" the dev process will be very hard to manage.

²Tripartite because we have a rigid division between the citizen, the engineer and the politician figures.

³also bringing to the table managers, political figures, health and environmental workers, and users

3 Moralizing Technologies

Technologies and infrastructures have an impact on society: they can be a facilitation utility, a barrier, a social disruptive infrastructure...

A famous example are Robert Moses' overpasses in the USA: Moses' was a very influential urban planner and the planning of roads and transportation infrastructure was up to him. He was *profoundly racist*, and this can be seen in his design: all the overpasses, bridges, parkways, roads are built to discourage the transit of public transportation means, because at the time they were used mostly by the low income portion of the population.

3.1 "Do Artifacts have Politics?"

In the famous paper "Do Artifacts have Politics", Winner suggests the idea that the artifacts *should* be politically and morally charged, because they in cases take decisions for people; as an example, a speed bump (a *piece of road technology*) can be seen as a decision-taker for drivers.

3.2 Technological Mediation

For "technological mediation" is intended the phenomenon of a piece of technology not only fulfilling his purpose, but also shaping the decisions and behaviours of the people interacting with it. This concept encapsulates a lot of different phenomena: incorporating technologies as glasses are *used* to perceive reality, we do not perceive them; representing technology instead are artifacts (as a digital thermometer) that are used to specify a single characteristic of reality itself.

A further example can be obstetrical ultrasound: it's not simply a "observing" technology, but *mediates the relationship between the unborn child and the parents*. This has a series of effects, like

- it separates the fetus from the mother's body: new ontological status of the fetus
- it places the fetus in a context of medical norms: the unborn child is now part of the medical process that is pregnancy

This makes this technology profoundly ambivalent: it may both encourage abortion or discourage it.

Artifacts mediates actions also by a **script**, that is the prescription on how to use it (the speed bumps example): this is a **invitation-inhibition** mediation, where the design choices of a particular artifacts are made in order to facilitate or discourage some behaviours. Actions and decisions and interpretations of the world are co-shaped by technologies.

This aspect has two sides: people should stop moralizing *each other* applying morality only to people but starting to develop things *morally* embedding ethics and morality in the technology we use, the environment we live in.

3.2.1 Embedding Morals in the Environment

Following the reasoning that suggests to *embody moral decisions into our tools and places* we can easily find the prime argument against it: behaviour-steering technologies are perceived as a restriction of human freedom. Using technologic solutions to *explicitly avoid* that someones take part in some dangerous/"wrong" actions means (in this argument) exchanging democracy with technocracy.

What's more, it complicates the charging process: who's to charge/blame when such a piece of technology/artifact is involved in an accident? Technologies are different from **laws** when it comes to limiting human freedom. Why? Because they're not part of a **democratic process**: and that's what we should look for, a *democratic way to moralize technology*.

This is also done by designers, that need to anticipate the mediating role of their own technologies; the design and creative process should be changed (in the collective imaginario) from *creating a functional product* to a way to **materialize morality**. The ethics of engineering design should take into account the moral charge of technological products.

4 Computer Ethics

4.1 The Invisibility Factor

Moor, in his document "What is computer ethics", highlights some of the characteristics that make computer technology so disruptive and yet it's so difficult to formulate right policies for them and moralize / weight them. One of this, set aside the *logical malleability*, it's the so-called **invisibility factor**: computer operations are invisible, as the users are not aware of the internal low level operation of the computer. Moor analyzes three main "invisibility factors":

1. Invisibility of **abuse**: exploiting the invisibility factor in order to perform unethical operations
2. Invisibility of **programming values**: biases can be embedded into programming values, such as a recommender system
3. Invisibility through **complexity**: the simple fact that an operation is hard to grasp in its wholeness (sheer size of the data, calculations and output performed) renders it super difficult to judge

These are new problems introduced by the IT, but the computer technology also forces us to do new conceptual analysis on old problems that are changed by the simple introduction of IT. In fact, *computer ethics is not applied ethics*.

4.1.1 IT-configured Activities

The fact that computers are *tools to build stuff with* and not simple buildings or static artifacts makes the design choices that go into them much more significative and impactful than other pieces of technology. In fact, IT is a socio-technical system, as it allows to achieve social function and eases the contitution and maintainance of institutions.

The ethical issues introduced by IT are hard to pin down as a whole; however, three are the main features of such issues:

- **Global, many to many scope**: the internet (and all technologies built on top of that) shares a major difference with the other global-scale information systems (as the television and the radio): it enables **many to many** communications, where radio and television are **one to many** communications; this means that's far easier to organize and put in place networks of people, because there's no need to pass through a central organ to do so. Moreover, *virtually anybody* is connected with *virtually anybody*, *virtually everywhere*: a tool of this granularity and size is incredibly difficult to control and manage.
- **Distinctive identity conditions**: internet communication is *mediated*, so there's a third party between the sender and the receiver of a message⁴. This means that the *identity options* that are available depends on the context: the two parties can be anonymous one another, or they can be partially known, or they can be even verified to be the physical person they're presenting. This is enabled by a complex socio-technical system built on the internet communication service.
- **Reproducibility**: electronic information has a particular property: **it can be stolen without causing the original owner to lose value**. Just think of a piece of code: copying it does not harm in any way the original writer of the code, that can even be unaware of the "steal". If we transfer the example from code to **words**, the scenario is a little more daunting: this system weakens the link between *who said something* and the actual content of his message, so there's less control of written words by those who writes them.

4.1.2 IT-configured Domains of Life

We can see the biggest ethical impact of IT in three major domains of everyday life:

⁴usually more than one. Could be the ISP, the company owner of the network, the owner of the *app* you're using...

- Social relationships: can a relation that's *fully internet mediated* be considered a "real" relationship? This is a problem that's also connected to the concept of online identity, that is something we have more or less total control over. Moreover, this part can be also analyzed by the *reproducibility* point of view: how a relation can evolve when each word or action is memorized?
- Education: again, a problem related to reproducibility. How to reconcile the **norms of traditional education** and the possibility offered by the information system that's internet? When is plagiarism actually plagiarism? Should all the knowledge free to access for everyone?
- Internet democracy...