

CS 3001 Notes

CS Ethics Notes Based On Lectures/Lecture Slides

Krish Katariya

Last updated: **February 24, 2024**

Contents

1. Therac-25	1
1.1. Software Errors	1
1.2. Design Flaws	1
1.3. Moral Responsibility	2
2. Utilitarianism	2
3. Kantianism / Deontology	2
3.1. Duties	2
3.2. Imperatives	3
4. Social Contract	3
5. Virtue Ethics	3
6. Internal Realism	4
7. Evidentialism	4
8. Reliabilism	4

1. Therac-25

Definition 1.1

The **therac-25** was a linear accelerator created to treat cancer with radiation therapy

The therac-25 was an all new version of earlier therac-6 and therac-20 machines, but unlike them did not have manual overrides.

The therac-25 had malfunctions and would give too much, or too little (which could also be dangerous) radiation causing lots of harm.

1.1. Software Errors

Two of the therac-25's main software errors were **race conditions**.

Definition 1.1.1

Race Conditions is when a system tries to perform two operations at very similar times but because of the nature of the system, the operations have to be done in specific manner for it to work correctly.

One race condition was with the system acting on old information that had been changed to new information while the computer wasn't checking the values.

1.2. Design Flaws

Design process flaws by reusing older code designed to work with a system that had manual overrides and locks.

- So the flows were system flaws, not only software errors

The system was not designed with being fail safe in mind

1.3. Moral Responsibility

In order for a moral agent to be responsible for a harmful event, two conditions must hold:

- **Casual Condition:** The actions (or inactions) of agents must have caused the harm
- **Mental Condition:** The actions (or inactions) of agents must have been intended or willed by the agent*

*This also includes unintended harm if it came from carelessness or negligence

2. Utilitarianism

Theorem 2.1

In utilitarianism, an action is good if the benefits exceed its harms, and an action is bad if its harms exceed its benefits

Definition 2.1

Act Utilitarianism is the ethical theory that an action is good if the net effect (over anyone and everyone who is affected) is to produce more happiness than unhappiness

Case For:

- It focuses on happiness
- It is practical
- It is generally comprehensive

Case Against:

- Where to draw line in calculations? How bad is too bad if the end justifies the means?
- Takes a lot of time for each moral calculation
- Ignores innate sense of

Definition 2.2

Rule Utilitarianism is the ethical theory that holds that we ought to adopt those moral rules that, if followed by everyone, lead to the greatest increase in total happiness over all affected parties

A *rule utilitarian* chooses to follow a moral rule because its universal adoption would result in the greatest net increase in happiness. *This differs from kantianism because a rule utilitarian is still concerned with the consequences and not the motive.*

Case For:

- Not every moral decision requires performing utilitarian calculus
- Exceptional situations don't overthrow moral rules
- Reduces problem of **bias**(?)
 - idk what this means

Case Against:

- It forces us to use a single scale or measure to evaluate completely different kinds of consequences
- It ignores the problem of an unjust distribution of good consequences

3. Kantianism / Deontology

Unlike utilitarianism, Kantianism is not concerned with outcomes. Instead, the motive for why you did certain actions is more important.

3.1. Duties

Definition 3.1.1

Perfect duties are duties we are obligated to fulfill in every instance. **Imperfect duties** we are obligated to fulfill in general but not in every instance.

3.2. Imperatives

An **imperative** is a way in which reason commands will.

Definition 3.2.1

A **hypothetical imperative** is a conditional rule of the form “If you want X then do Y” A **categorical imperative** is an unconditional rule: a rule that always applies, regardless of circumstance. For Kant, only a categorical imperative can be a moral imperative.

Both of the *main two categorical imperatives* should be held for an act to be considered morally right.

1. Act only from moral rules that you can at the same time will to be universal laws
2. Act so that you always treat both yourself and other people as ends in themselves, and never only as a means to an end

The Case For:

- It treats all person as moral equals
- It gives all persons moral worth by considering them rational, autonomous beings
- Everyone is held to the same standard
- Kantianism produces universal moral guidelines

The Case Against:

- Sometimes no single rule can fully characterize an action
- Sometimes there is no way to resolve a conflict between rules
- Kantianism allows no exceptions to perfect duties

4. Social Contract

5. Virtue Ethics

Virtue ethics can address the questions of how one should live, what kind of person one should become, and even what one should do in certain scenarios. Would a good person perform this action under this virtue?

Definition 5.1

Virtues: behaviors showing high standards. The qualities or states of character that find expression in morally good actions and morally good purposes.

Definition 5.2

Vices: the opposite from virtues. A vice is a character trait that prevents a human being from flourishing or being truly happy

Usually, there are two vices for every virtue

Note 5.1

Virtue ethics pays particular attention to the agent as well as the action and the consequences of the action. This touches on social contract theory, utilitarianism, and Kantianism. *A good person does the right thing at the right time for the right reason.*

Pros

- In many situations it makes more sense to focus on virtues than on obligations or rights
- Personal relationships can be morally relevant to decision making
- There are no irresolvable moral dilemmas
- Recognizes the importance of having a say in living a moral life

Cons

- Different people may have quite different conceptions of what human flourishing is
- Virtue ethics cannot be used to guide government policy
-

6. Internal Realism

- How do we know there is a desk here?
 - We are all trapped in our subjective perceptions of the world around us, but our subjective perceptions oddly agree

7. Evidentialism

Definition 7.1

Evidentialism: all beliefs must be supported by evidence

- Some beliefs are basic and justified, such as things you learn through a reliable sense

8. Reliabilism

Definition 8.1

Reliabilism: a belief is reliable if it is the result of a reliable cognitive process.

- A *reliable* process is one that generally leads to the correct outcomes