The Alan Turing Institute

AI Ethics and Governance

Day 1: Introduction to Practical Ethics

21/11/2022

Professor David Leslie
Director of Ethics and Responsible Innovation

Welcome to AI Ethics and Governance!

OVERVIEW

Introduction to Practical Ethics

01

AI Harms & Values

02

AI Sustainability

03

Fairness, Bias Mitigation & Accountability

04

Explainability, CARE & Act principles

05

OVERVIEW

Introduction to Practical Ethics

AI Harms & Values

AI Sustainability

Fairness, Bias Mitigation & Accountability

Explainability, CARE & Act principles

01 02 03 04 05

CONTENTS

Why care about ethics?

Why care about ethics?

Who would you save and why?

Why care about ethics?

Who would you save and why?

- 30 middle-aged workers (~45 y/o) from socioeconomically disadvantaged backgrounds or 20 affluent college students (~20 y/o)?
- How did you reach that decision? What is the rationale behind your choice?

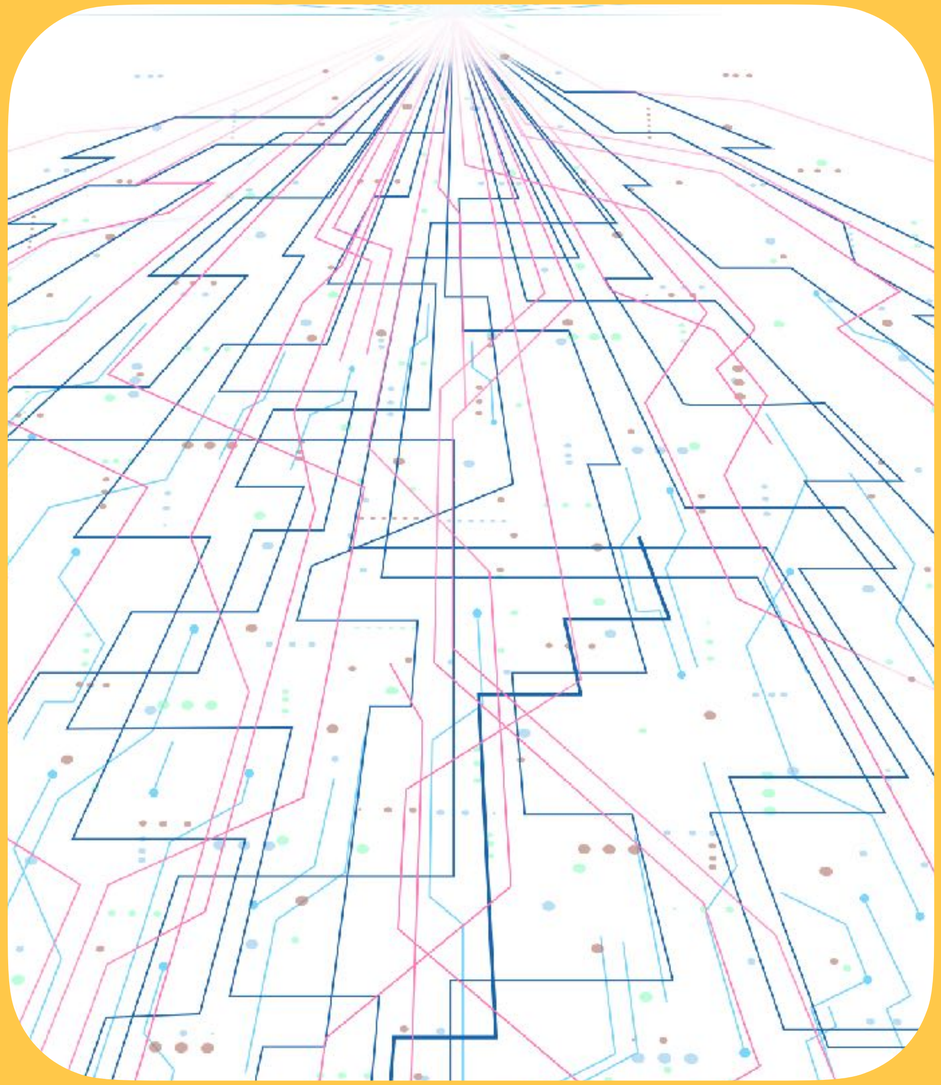
Some further ‘metaethical’ questions:

- How do you know if your decision is the correct one?
- Is there an objective ‘right’ decision? Is there a moral truth or fact of the matter? If there is, is there a way to determine it?

This is the topic of this morning’s lecture!

01

METAETHICS



Ethics is the philosophical study of morality [...]

Deigh, 2010

[...] of what are good and bad ends to pursue in life and what it is right and wrong to do in the conduct of life.

Deigh, 2010

It is therefore, above all, a practical discipline.

Deigh, 2010

Ethics is the philosophical study of morality [...] of what are good and bad ends to pursue in life and what it is right and wrong to do in the conduct of life. It is therefore, above all, a practical discipline.

Deigh, 2010

Ethics is the philosophical study of morality [...] of what are good and bad ends to pursue in life and what it is right and wrong to do in the conduct of life. It is therefore, above all, a practical discipline.

Deigh, 2010

Ethics is the philosophical study of morality [...] of what are good and bad ends to pursue in life and what it is right and wrong to do in the conduct of life. It is therefore, above all, a practical discipline.

Deigh, 2010

Ethics is the philosophical study of morality [...] of what are good and bad ends to pursue in life and what it is right and wrong to do in the conduct of life. It is therefore, above all, a practical discipline.

Deigh, 2010

Then... what is metaethics?

Then... what is metaethics?

- The study of what morality is: are there moral facts about the world?

Then... what is metaethics?

- The study of what morality is: are there moral facts about the world?
- If there are, how can we know them?

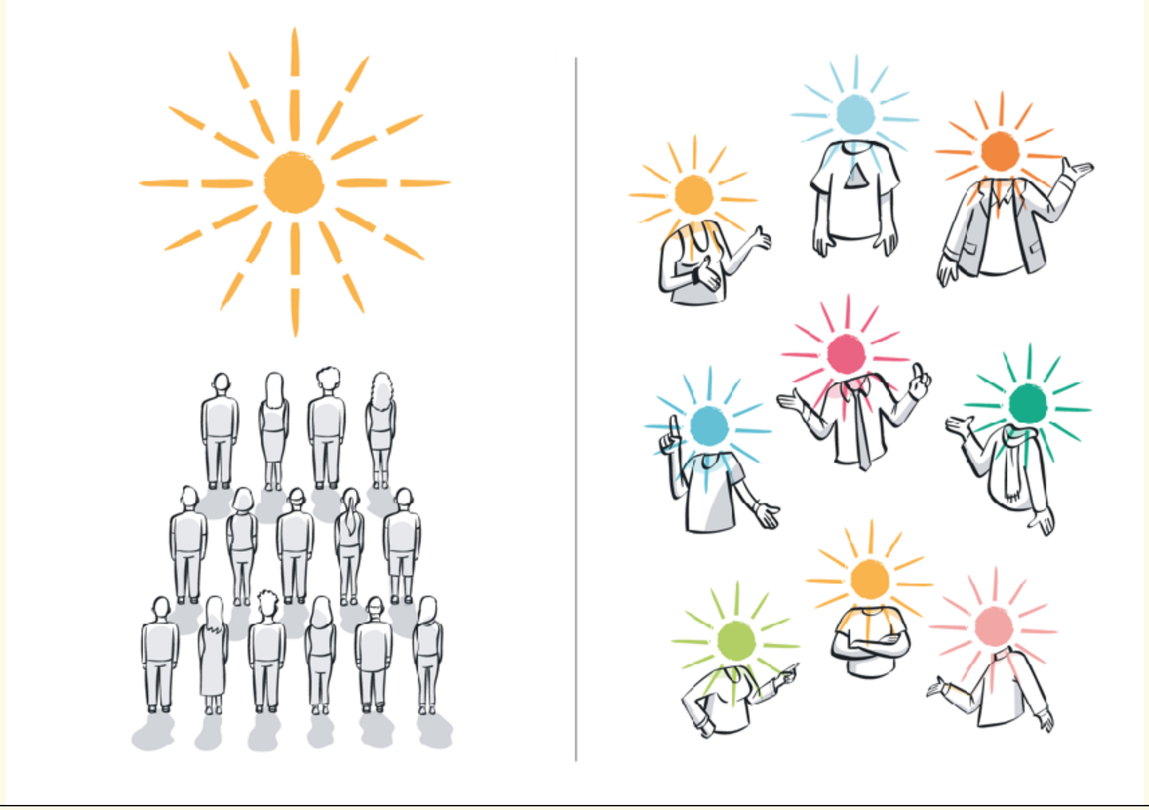
Then... what is metaethics?

- The study of what morality is: are there moral facts about the world?
- If there are, how can we know them?
- And if there are not, how can we decide whether our actions are right or wrong?

How have these questions been answered in contemporary metaethics?

How have these questions been answered in contemporary metaethics?

There have been two ends of the metaethical spectrum:

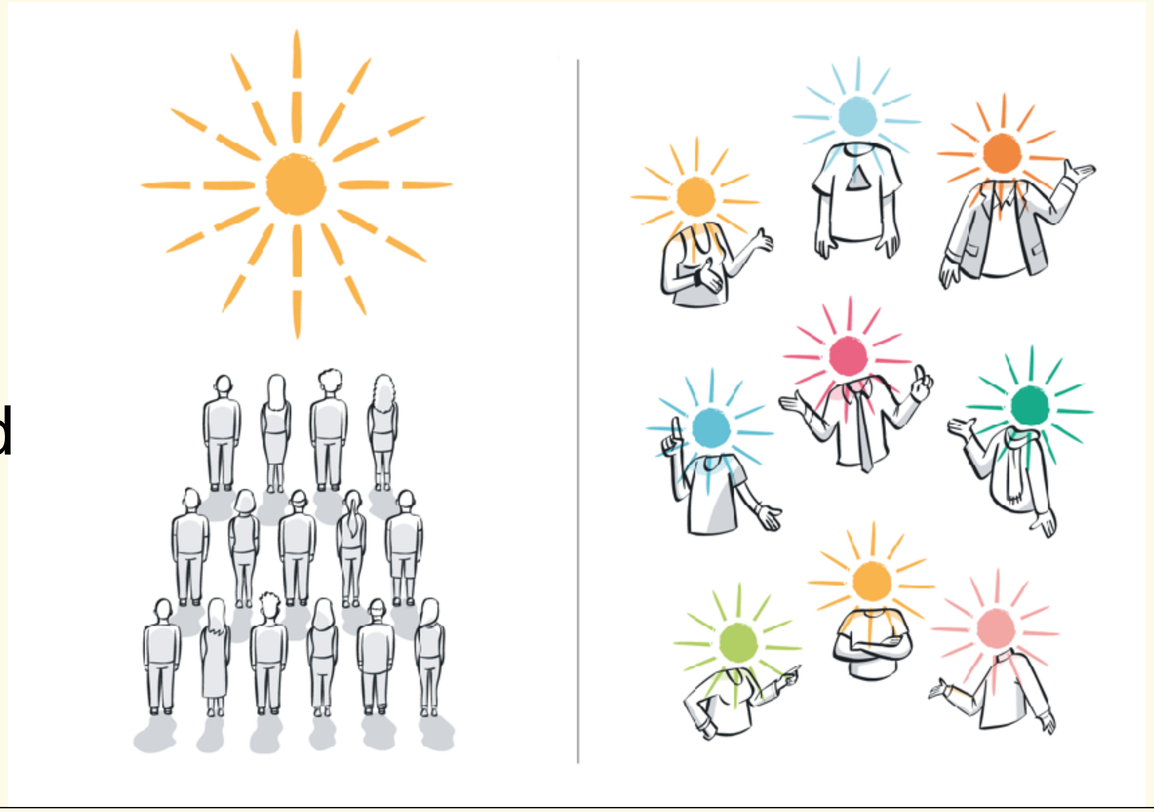


How have these questions been answered in contemporary metaethics?

There have been two ends of the metaethical spectrum:

Moral realism

The view that there are mind-independent moral facts about the world and that moral judgements can be objectively true.

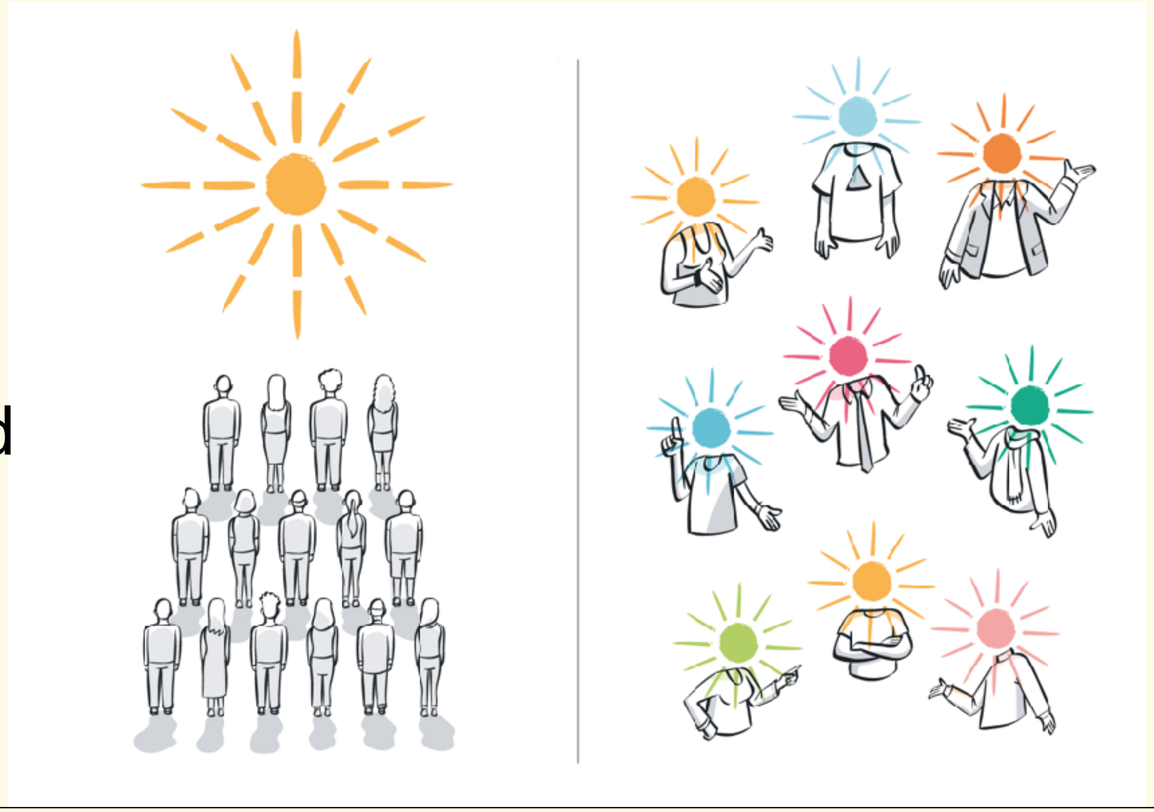


How have these questions been answered in contemporary metaethics?

There have been two ends of the metaethical spectrum:

Moral realism

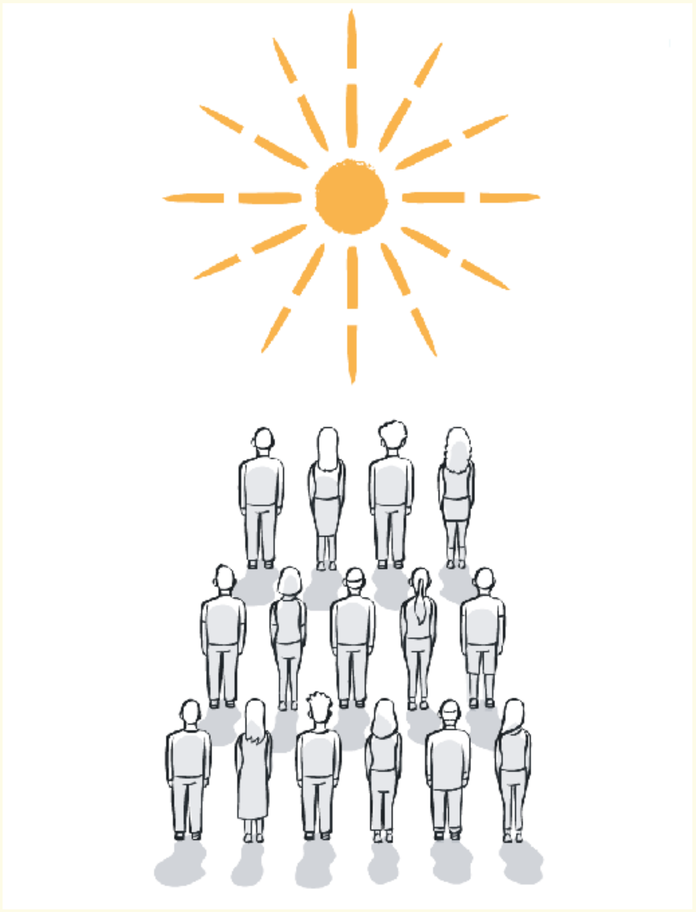
The view that there are mind-independent moral facts about the world and that moral judgements can be objectively true.



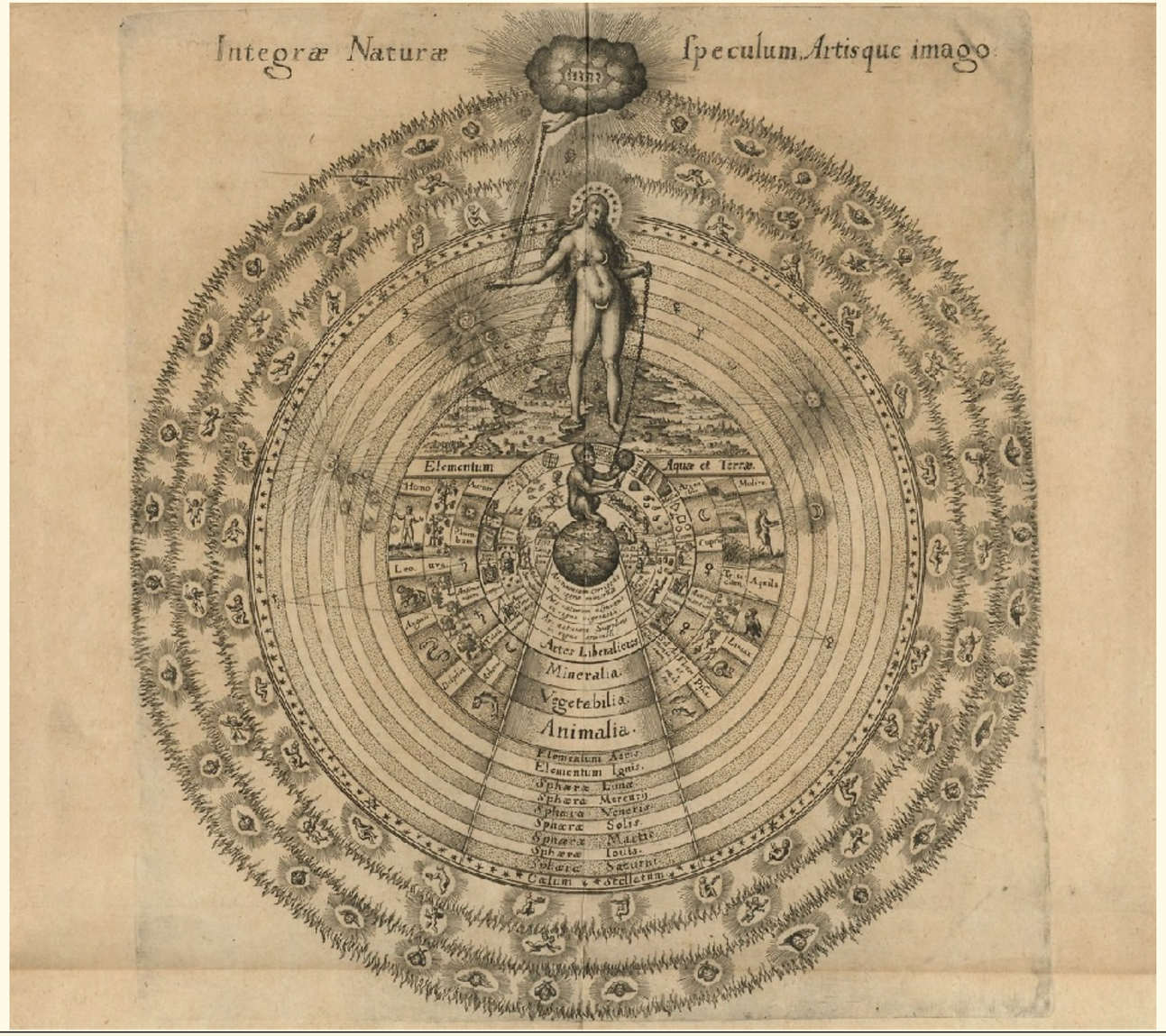
Moral subjectivism

The view that moral judgements are subjective expressions of feelings, attitudes, or preferences rather than objective truths.

There is an important historical story behind the relationship of moral realism and moral subjectivism



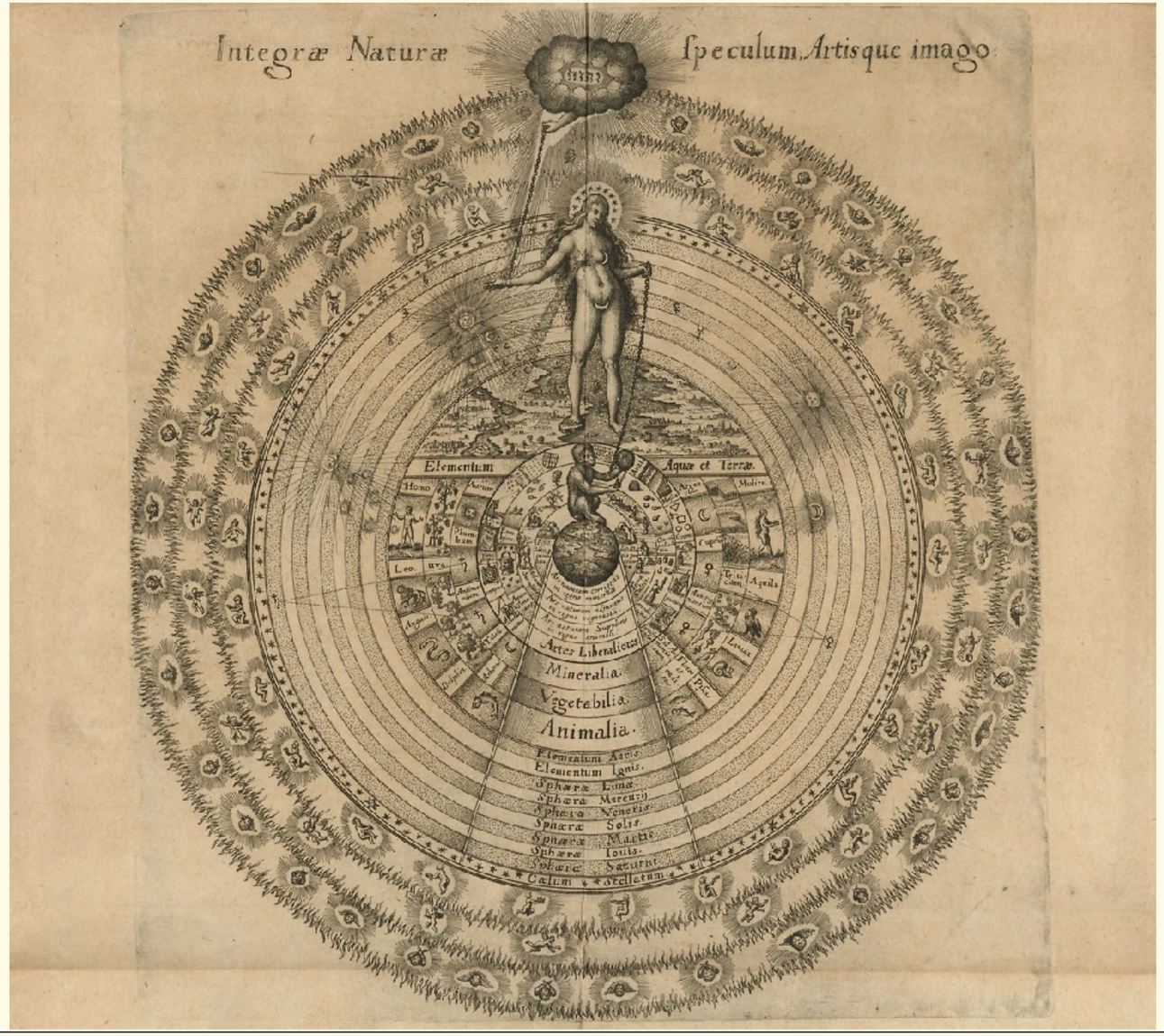
The story begins in the Renaissance period



Robert Fludd, *Urtriusque cosmi maioris scilicet et minoris metaphysica, physica atque technica historia, in duo volumnia secundum cosmi differntiam diuisa* or "The Great Chain of Being" (1617)

'The Mirror of the Whole of Nature and the Image of Art'

The story begins in the Renaissance period



Robert Fludd, *Urtriusque cosmi maioris scilicet et minoris metaphysica, physica atque technica historia, in duo volumnia secundum cosmi differntiam diuisa* or "The Great Chain of Being" (1617)

Pre-modern cosmic order is continuous with the moral universe.

Ideas of the "good life" and virtuous action were anchored in this theologically pre-structured or teleologically pre-patterned universe.

Authority of Scripture and divine commandments and laws fixed by an intrinsically meaningful cosmic order secured the objectivity of moral truths

Violent disruptions of the pre-modern Western order

Expansion of the Ottoman Turks into Eastern Europe

Waves of Bubonic Plague (Black Death)

Mass displacement of serfs leads to establishment of free cities and the end of feudalism

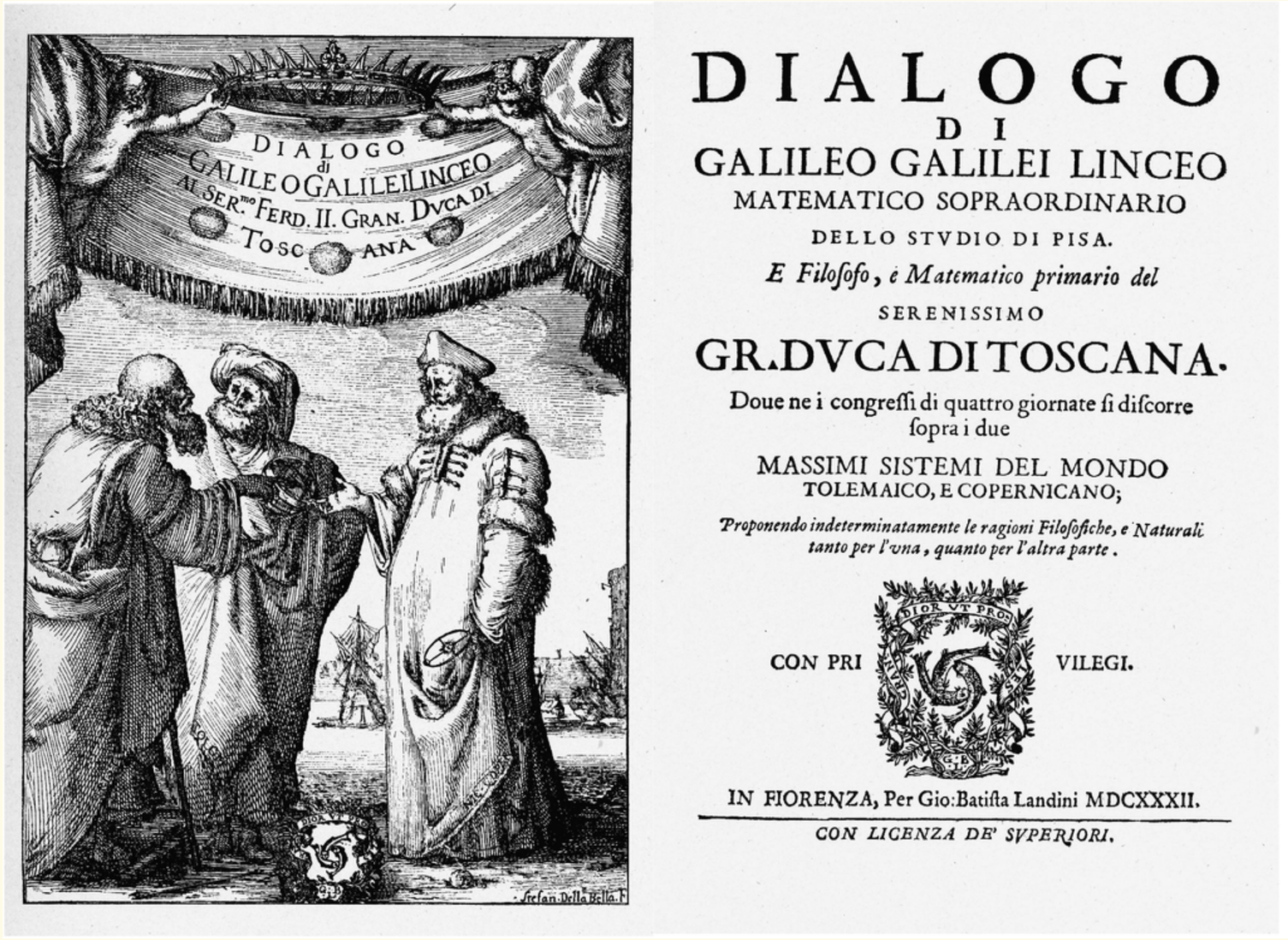
‘Discovery of the new world’ introduces a plurality of cultural histories and challenges single origin story.

Protestant Reformation ends the unquestioned and presumptive authority of the Catholic Church

New astronomical tools and observations contest Church cosmology and support Copernican Revolution

Towards a new era of creativity and consensus

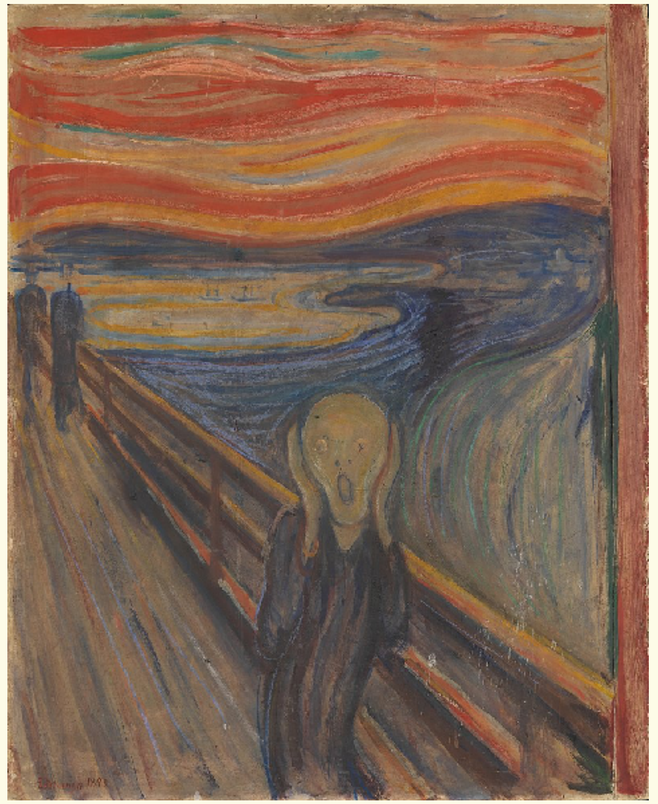
This is the frontispiece of Galileo Galilei, *Dialogue Concerning the Two Chief World Systems (1633)*



Two downstream effects of modernisation on metaethical self-understandings

(1) The rise of moral subjectivism, cultural relativism, and existential crisis

An increasing awareness of the moral uncertainty and the epistemic fragility and finitude of the human condition leads to skepticism about the very possibility of moral truth and ethical meaning, more generally.



Two downstream effects of modernisation on metaethical self-understandings

(1) The rise of moral subjectivism, cultural relativism, and existential crisis

Two downstream effects of modernisation on metaethical self-understandings

Two downstream effects of modernisation on metaethical self-understandings

(2) The rise of procedural approaches to moral reasoning and ethical deliberation

In light of the plurality of values and cultural openness of modern life, procedural approaches stress:

-the importance of inclusive, balanced, and power-aware dialogue

-the rational exchange and assessment of ideas and beliefs

-fulfilment of preconditions of meaningful deliberation

Two downstream effects of modernisation on metaethical self-understandings

(2) The rise of procedural approaches to moral reasoning and ethical deliberation

In this course we will stress a **procedural approach**

In this course we will stress a **procedural approach**

What do we mean by this?

In this course we will stress a **procedural approach**

What do we mean by this?

Ethics as a form of **regulated discourse** aimed at reaching a **consensus on ethical values ethical values and moral principles**

In this course we will stress a **procedural approach**

What do we mean by this?

Ethics as a form of **regulated discourse** aimed at reaching a **consensus on ethical values ethical values and moral principles**

Moral validity is bound by the practices of **giving and asking for reasons**

In this course we will stress a **procedural approach**

What do we mean by this?

Ethics as a form of **regulated discourse** aimed at reaching a **consensus on ethical values ethical values and moral principles**

Moral validity is bound by the practices of **giving and asking for reasons**

Some relevant questions:

In this course we will stress a **procedural approach**

What do we mean by this?

Ethics as a form of **regulated discourse** aimed at reaching a **consensus on ethical values ethical values and moral principles**

Moral validity is bound by the practices of **giving and asking for reasons**

Some relevant questions:

What are the **requirements of rationality?**

In this course we will stress a **procedural approach**

What do we mean by this?

Ethics as a form of **regulated discourse** aimed at reaching a **consensus on ethical values ethical values and moral principles**

Moral validity is bound by the practices of **giving and asking for reasons**

Some relevant questions:

What are the **requirements of rationality?**

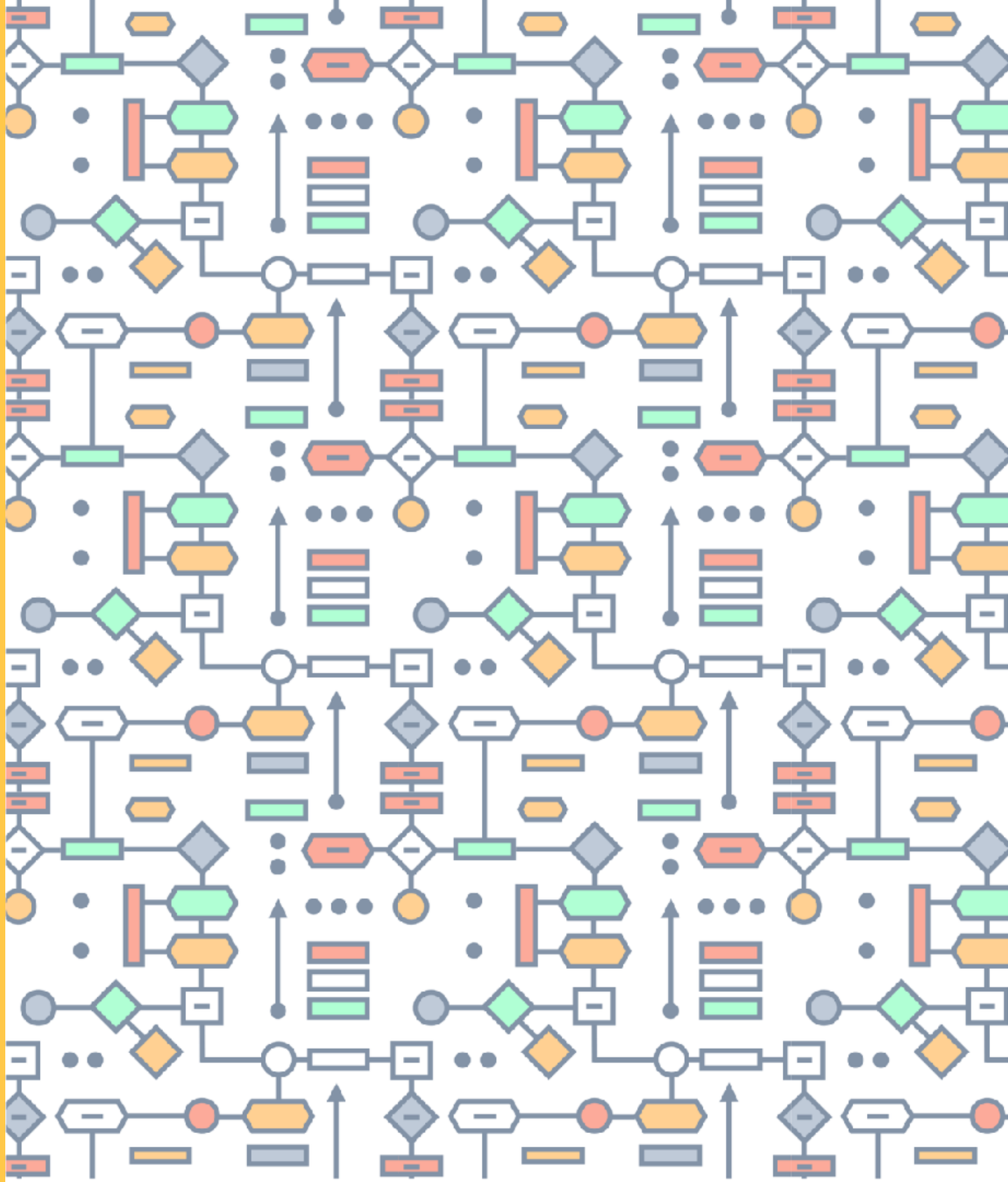
What are the **enabling conditions** to come to a defensible moral judgement?

Questions?

Activity 1: Using moral concepts

02

NORMATIVE THEORIES



NORMATIVE ETHICAL THEORIES

NORMATIVE ETHICAL THEORIES

Normative ethics focuses on the actual **content of ethical theories**

NORMATIVE ETHICAL THEORIES

Normative ethics focuses on the actual **content of ethical theories**

They attempt to answer the question of what makes **someone ethical**, or what makes **certain actions morally permissible**

NORMATIVE ETHICAL THEORIES

Normative ethics focuses on the actual **content of ethical theories**

They attempt to answer the question of what makes **someone ethical**, or what makes **certain actions morally permissible**

Normative ethical theories come in many different flavours

Consequence-based ethics

CONSEQUENCE-BASED ETHICS

CONSEQUENCE-BASED ETHICS

An action is right if its **consequences are better than**, or at least as good as, the consequences of any other action that the agent could have done

CONSEQUENCE-BASED ETHICS

An action is right if its **consequences are better than**, or at least as good as, the consequences of any other action that the agent could have done

What matters is not the *process* by which an outcome is produced, but rather the **outcome itself**.

Consequence-based ethics

Principles-based ethics

PRINCIPLES-BASED ETHICS OR DEONTOLOGY

PRINCIPLES-BASED ETHICS OR DEONTOLOGY

Rightness of an action is determined by an agent's application of **some universal standard, rule or maxim**

PRINCIPLES-BASED ETHICS OR DEONTOLOGY

- Rightness of an action is determined by an agent's application of **some universal standard, rule or maxim**
- This rule should be followed **irrespective of consequences**

PRINCIPLES-BASED ETHICS OR DEONTOLOGY

- Rightness of an action is determined by an agent's application of **some universal standard, rule or maxim**
- This rule should be followed **irrespective of consequences**
- Kant's moral theory: humans are free to the extent that they can **use their reason alone to judge** the validity of and bind themselves by **universalisable moral maxims**

- Consequence-based ethics
- Principles-based ethics
- Virtue ethics
- Biocentric ethics

Virtue ethics

Virtue ethics

Shift from actions to **character traits**

Virtue ethics

Shift from actions to **character traits**

Instead of starting with ‘what should I do’, it asks ‘**what sort of person should I strive to become’** in order to live an ethical life

Virtue ethics

Shift from actions to **character traits**

Instead of starting with ‘what should I do’, it asks ‘**what sort of person should I strive to become’** in order to live an ethical life

The **development of moral virtues** through **practice, discipline, and repetition** is the purpose of a human ethical life.

Biocentric ethics

Biocentric ethics

Justifies moral responsibility and the rightness of actions in a **non human-centred way**

Biocentric ethics

Justifies moral responsibility and the rightness of actions in a **non human-centred way**

Attempts to address the **failures of anthropocentric** moral theories

Biocentric ethics

Justifies moral responsibility and the rightness of actions in a **non human-centred way**

Attempts to address the **failures of anthropocentric** moral theories

Nature is not as an instrument or resource available for indiscriminate human use and exploitation

Biocentric ethics

Justifies moral responsibility and the rightness of actions in a **non human-centred way**

Attempts to address the **failures of anthropocentric** moral theories

Nature is not as an instrument or resource available for indiscriminate human use and exploitation

Intrinsic value of all life forms and ecologies

Virtue ethics

Shift from actions to **character traits**

Instead of starting with ‘what should I do’, it asks ‘**what sort of person should I strive to become in order to live an ethical life**’

The development of moral virtues through practice, discipline, and repetition is the purpose of a human ethical life.

Biocentric ethics

Justifies moral responsibility and the rightness of actions in a **non human-centred way**

Attempts to address the **failures of anthropocentric** moral theories

Nature is not as an instrument or resource available for indiscriminate human use and exploitation

Instead, claims the **intrinsic value** of all life forms and ecologies

Consequence-based ethics

Principles-based ethics

Virtue ethics

Biocentric ethics

Questions?

Activity 2: Using moral concepts II


Thank you!
See you tomorrow for Day 2: AI harms and values