

Why Ethics?

Module 3 of a course on *Ethical Issues in AI*

Prepared by

John Hooker

Emeritus Professor, Carnegie Mellon University

Chautauqua, June 2024

Ethics as a negotiation tool

- Ethics is a method of reaching **rational consensus** on how we are to live together.
 - *It judges **actions**, not people.*
- Ethics does not **impose values** on others.
 - *Just the **opposite!***
 - *It seeks **consensus!***



Ethics and religion

- Isn't ethics based on **religion**?
 - *If so, how can we reach **consensus**?*

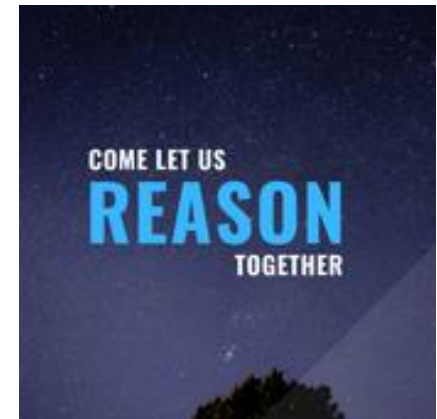


Ethics and religion

- **This much is true...**
 - *Religions have made fundamental contributions to ethics.*

Ethics and religion

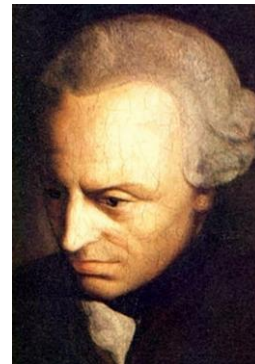
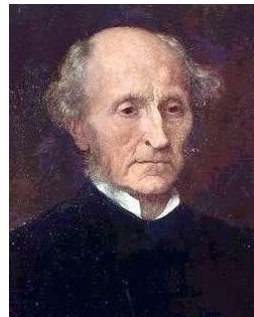
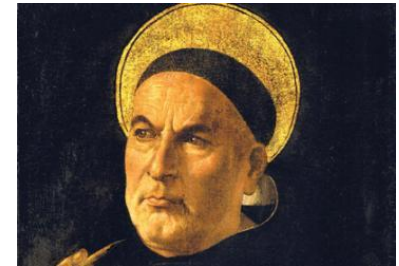
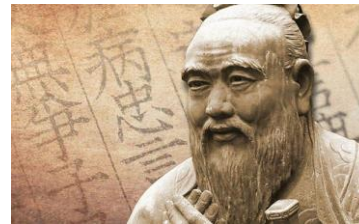
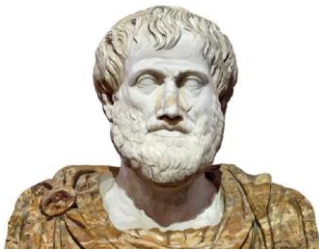
- **This much is true...**
 - *Religions have made fundamental contributions to ethics.*
- **But...**
 - *Religions have a long tradition of **rational discourse** on ethics, etc.*
 - Judaism: Talmudic scholarship
 - Islam: Ulema (community of scholars)
 - Christianity: Thomist philosophy
 - *Religions made fundamental contributions to **mathematics**.*
 - Yet we pursue the field without religious assumptions.



Isaiah
8th c. BCE

Ethics and rationality

- Is **rational analysis** possible in ethics?
 - *People have been doing it for **centuries**.*
 - Including some of the smartest people who ever lived.
 - We can distill the best ideas from their thought.

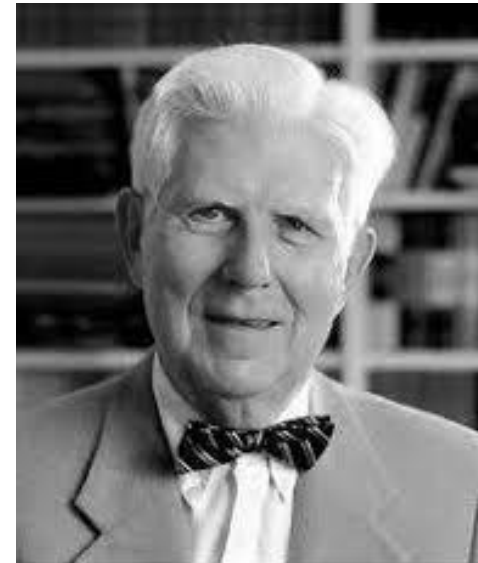


Ethics and rationality

- But aren't humans are driven by **emotion** and subconscious motives, **not reason**?
 - *If so, ethical reasoning has no influence on behavior.*
 - At best, it only justifies what we already want to do.

Ethics and rationality

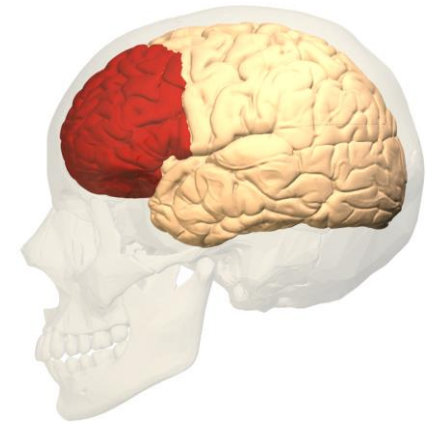
- But aren't humans are driven by **emotion** and subconscious motives, **not reason**?
 - *If so, ethical reasoning has no influence on behavior.*
 - At best, it only justifies what we already want to do.
- At best, this is **warmed-over Freud**.
 - *Today, cognitive psychology recognizes reason as an equal partner with emotion.*
 - Cognitive therapy is used to treat depression, PTSD, etc.



Aaron T. Beck
1921- 2021 (!)

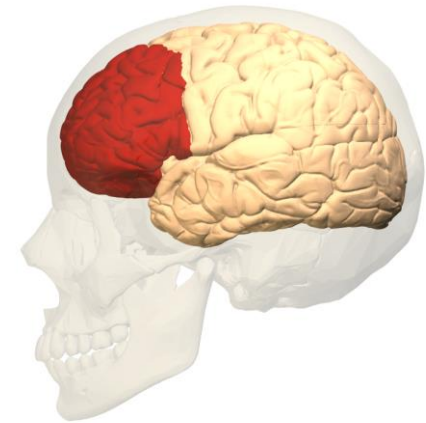
Ethics and rationality

- The **prefrontal cortex** evolved due its survival value in east Africa.
 - We are *thinking creatures* by nature.
 - People almost always (try to) **justify** their actions.



Ethics and rationality

- The **prefrontal cortex** evolved due its survival value in east Africa.
 - We are *thinking creatures* by nature.
 - People almost always (try to) **justify** their actions.
- But people believe in crazy conspiracy theories...
 - This proves the point. They want a **rationale**.
 - But we must **learn** to think rationally, as with any other skill.
 - Whence courses in **ethics!**



Facts and values

- **Facts \neq values**
 - *The distinction will factor into **every issue we discuss.***

Facts and values

- **Facts \neq values**
 - *The distinction will factor into **every issue we discuss.***
- **Why?**
 - *Facts are important in ethical arguments, but...*
 - *One can't deduce an ethical conclusion **solely** from facts.*
 - There must be an **ethical premise** at some point.
 - *This is the **is-ought** gap.*



David Hume, *Treatise on Human Nature* (1739)

Facts and values

- Conclusion doesn't follow:
 - *Everyone is cheating on the exam.*
 - *Therefore, it's ethical to cheat on the exam.*

Facts and values

- Conclusion doesn't follow:
 - *Everyone is cheating on the exam.*
 - *Therefore, it's ethical to cheat on the exam.*
- Conclusion follows:
 - *Everyone is cheating on the exam.*
 - ***Something that everyone does is ethical.*** ← *Ethical premise*
 - *Therefore, it's ethical to cheat on the exam.*

Facts and values

- Conclusion doesn't follow:
 - *Everyone is cheating on the exam.*
 - *Therefore, it's ethical to cheat on the exam.*
- Conclusion follows:
 - *Everyone is cheating on the exam.*
 - ***Something that everyone does is ethical.***
 - *Therefore, it's ethical to cheat on the exam.*

*Ethical
premise*



...but are the premises **true**?

If not, the argument **fails to establish its conclusion.**

Facts and values

- Also known as the **naturalistic fallacy**.
 - *G. E. Moore: Even when we know the facts, we can sensibly ask, “Is this ethical?”.*
 - Bombing a train station kills innocent people.
 - Yet one can still **ask** if it is wrong.
 - Some terrorists insist it is ethical.



G. E. Moore
1873-1958

Facts and values

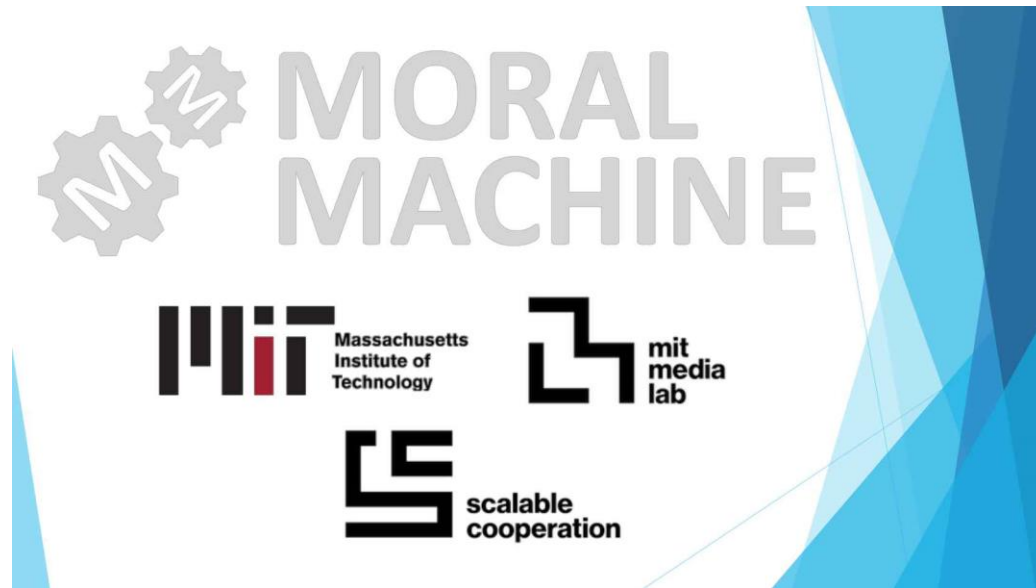
- Also known as the **naturalistic fallacy**.
 - *G. E. Moore: Even when we know the facts, we can sensibly ask, “Is this ethical?”.*
 - Bombing a train station kills innocent people.
 - Yet one can still **ask** if it is wrong.
 - Some terrorists insist it is ethical.
 - *So, it’s a **fallacy** to draw ethical conclusions from facts alone.*



G. E. Moore
1873-1958

Facts and values

- Recent example: Value alignment in AI
 - *To be discussed.*



Facts and values

- Many of the controversies in AI ethics are over the **facts**.
 - *Is personal data harvesting harmful or create risks?*
 - *Are Internet users misled about the extent of surveillance?*
 - *Does social media use lead to addiction*
 - *Do services like Alexa record private conversations?*
 - *Do online hate speech and conspiracy theories incite violence?*



Facts and values

- **Ethics** cannot tell us the **facts**.
 - *We must go out and investigate, gather **data**.*
 - *No amount of talk will settle issues of fact.*

Facts and values

- **Ethics** cannot tell us the **facts**.
 - *We must go out and investigate, gather **data**.*
 - *No amount of talk will settle issues of fact.*
- **Facts** cannot tell us what is **ethical**.
 - *We need ethical principles.*
 - *These are established by **analysis**.*

Facts and values

- **Ethics** cannot tell us the **facts**.
 - *We must go out and investigate, gather **data**.*
 - *No amount of talk will settle issues of fact.*
- **Facts** cannot tell us what is **ethical**.
 - *We need ethical principles.*
 - *These are established by **analysis**.*
- Ethics tells us which facts **matter**.
 - *This can save us a lot of effort.*
 - *Only chase down the facts that matter.*