Philosophy 152A
Moral Status and Psychological Capacities
Lecture Notes

Roy Zhao, UCLA

Mar 2023

Moral Status Apr 3 2023

Two kinds of definitions:

1. Absolute:
   S has moral status = S deserves moral consideration, S can be wronged, S has interests that can compel.
2. Comparative:
   S_1 has greater moral status than S_2 = S_1 deserves greater moral consideration, can suffer greater wrongs, interests carry more weight.

which brings up questions immediately relevant to abortion, AI, etc. And answering these questions often appeals to psychological capacities.

Central claim / Determination claim: the moral status of an individual S is completely determined by S's actual, present psychological capacities.

Two objections to the claim:

1. present?
   so infants, disabilities, or Alzheimer's? Maybe replace it with typical or recurrent.
2. Problem of other minds
   Rock with immense fear of dying? Or some super AI? If we don't know other mind, why do we want to ground ethics on something epistemologically unknown?

Another claim: Human cancer carries more moral weight than mouse cancer.

1. Pain matters more because it belongs to human.
2. OR humans just have more pain.

Utilitarianism: Mill & Bentham Apr 10 2023

The overarching Pricniple for Mill's Utilitarianism is The Greatest Happiness Principle: An action is right to the extent that it results in greater total happiness than alternative actions.

Let's reverse-engineer this on how to build an ethical theory! Often times we come up with a Theory of Values:

Tn Mill's case, he is grounded in

- Hedonism
  Happiness is the only thing that's intrinsically valuable.
  Happiness = total pleasure - total pain. It's the internal state that matters.
- Impartiality / Universality:
  No one's interests matter more than anyone else's, morally speaking

Along with a Theory of Right Action:

- Consequentialism
  An action's moral worth is determined by how valuable its consequences are.
- Maximization
  Action that maximizes the consequences.

The ends justify the means: not matter if it's stealing or killing, the maximized happiness is the only thing that matters.

Now we can revise the two kinds of definition for moral status for utilitarianism.

1. Absolute:
   S has moral status = "S is sentient"
   They have to be able to feel happy/pain.
2. Comparative:
   Any S that is sentient has equal moral status.
   From the impartiality criterion

Bentham encourage to extend morality for animals: it matters that they can suffer.

Hedonism

Mill's Argument for Hedonism:

1. "The only proof that something is valuable is that subjects (people) actually value it."
2. The only thing that subjects (people) value intrinsically is happiness.

$\therefore$ Happiness is the only thing that is intrinsically valuable.

For a simple example: money. Mill argues that it's extrinsically valuable, since it is a mean for happiness. If one is not allowed to buy something, it's worthless! Now for a harder one: truth, also autonomy, or health?

Questions remain also for the epistemological question on how to measure happiness, or how different types of pleasure/pain are.

Continued Mill & Utilitarianism Apr 12 2023

Recall for Hedoism that happiness is the only thing that's intrinsically valuable.

Objections to Hedonism follow:

1. How do you quantify happiness?
   How do you know what feels like for a worm to be stepped on, what about comparing it to me who gets stepped on? Ultimately an epistemic problem. Could also be a metaphysical problem if different happiness is incommensurable.
   Mill's solution: which one would you rather do?

2. An Argument:
   P1) If hedonism is true, then the best life is the one that contains the most happiness.
   P2) A life of base pleasure contains the most happiness. (bodily pleasures)
   P3) A life of base pleasure is not the best life
   ∴ Hedonism is false.

   Mill's Response: Denies P2, higher pleasure brings more happiness than bodily pleasures.
   P1) If a competent judge prefers A to B, As is a more valuable pleasure
   P2) Competent judges prefer higher to lower pleasures
   ∴ Higher pleasures are more valuable.
   Here, a competent judge would be a person who has done both. The comparison tells the value. Hence the utilitarianist move from happiness to **preferences**, kinda like Economics nowadays.

Now, what does this all say to animals? From the competent judge's argument, as people who can experience both, we would not swap life with a pig. Therefore, human pleasures are more valuable than animal pleasures.

Again rejections: what are competent judges? Are human and pig pleasures same?

Singer: *All Animals are Equal*

Two questions:

1. Why do animals deserve equal consideration?

2. What does equal consideration entail?

Consider Singer's "Expanding Circle" formulation: we all begin by caring for ourselves, then we extend the circle to the family, then tribe, nation, etc. We should also extend the moral circle to animals, which we are irrationally leaving out. Therefore, the same mistake that sexism and racism make.

An often-heard argument: The Argument for the Equality of Races (Attempt 1):

P1) There are no inherent differences in intelligence, ability, etc among races.

P2) The truth of P1 is what makes racism wrong.

$\therefore$ Racism is wrong.

Continued Singer: *All Animals are Equal* Apr 17 2023

Define the (sex/race/species)ism: giving greater weight to the interests of members of group A than B.

A couple of questions:

Q1. why is racism wrong?

Q2. Does the answer shows that speciesism is wrong too?

Peter Singer thinks it's true.

Back to the first attempt for equality, Singer thinks it's wrong, it also allows you to be a specieist since there are clear differences in intelligence between animals and humans. He responds by:

1. Don't risk it on empirical claim
   essentially opening up possibility that sociology or etc could change the result. If we really find out that one race is better at math, etc, then that means we can be racist, which is wrong.
   The upshot: some differences are compatible with moral equality.
2. seems to allow hierarchies based on the ability
   looks like we can have the smartest group of people ruling and everyone else is like garbage

We do a tailored version! The Argument for the Equality of Races (Attempt 2):

P1) If speciesism is justified, then there must be some morally significant feature (MSF) that all and only humans possess.

P2) There is no such MSF

$\therefore$ Speciesism is not justified.

Singer's defence of his premises:

1. Defence of P1):
   Swapping test: hold capacities of A and B as equal. If still privilege A, due to "Xism." ONLY okay if something is morally significant about A-hood.

2. Defence of P2):
   Human has to be smarter than all animals, and that's what allows us to do that. However, the intelligence distribution of chimps and humans will overlap! There is some human that is less intelligent than chimps.

An argument could be that all & only humans are moral. Humans are MSF because moral creatures/agents deserve more consideration.

Also a possible objection to P1): we can measure status by "normal" or typical level of species. (taking the mean by quantifying everyone's intelligence level) If we have a human that has less IQ than a chimp, this person gets the mean value of IQ of all human.

A possible objection to P2): being a homo sapien itself is a MSF.

Continued Singer: *All Animals are Equal* Apr 19 2023

Principle of Equal Consideration: in moral deliberations, we ought to give equal weight to the like interests of all others affected by our actions.

The idea behind: we have a scale when doing moral deliberation of all interests. If the total hurt outweighs the total benefit, then that's a no.

Egalitarian Views: we need to consider people on an equal footing of

1. Consideration
2. Opportunity
3. Equal Capabilities/outcomes/treatment

Equal consideration $\neq$ equal treatment & equal consideration $\neq$ equal outcomes. Ex. A is crushed, B has minor cut, we only have two doses of morphine. Singer thinks we should give both to A. So having the equal outcome sometimes means doing different things. What if A is a doctor?

Singer says that the rule of maximization follows: choose action that will result in greatest satisfaction of interests.

Now the scenario: A: lost a leg already and would lose a toe, B: will lose a leg. We most likely choose B, violation of Equal Consideration!

Different interests generate different consideration. Ex. Women should have abortion rights, men don't.

Objection: intelligent creatures have more weighty interest than others, so we will end up heavily favouring humans over other animals.

Reply: nonintelligent have a special interest. Ex. taking animals or humans captive.

Singer's Argument on eating meat:

P1) Principle of Equal Consideration: in moral deliberations, we ought to give equal weight to the like interests of all others affected by our actions.

P2) Eating meat satisfies a minor bodily interest of humans and violates major bodily interest of animals.

P3) From P2, eating meat violates equal consideration.

∴ Eating meat is wrong.

Some rejection: meat is not a minor interest. There are nutrients we need and etc etc. What if I have a cat who has to eat meat?

More on Animals' Consciousness Apr 24 2023

A quick philosophy of mind supplement: THE HARD problem - the problem of other consciouness

Qualia/Phenomenal Experience: what it feels like to be X? Culd be experience, consciousness, sentience. If I'm under anesthesia, then my qualia ceases.

We need to consider if animals have qualia or how their qualia is like, but how do we know?

Bertrand Russell's take: Argument from Analogy which is an inductive argument

P1) Some object Y is performing B

P2) When I'm performing B, I am in mental state M.

∴ Y is in M.

When is the argument weak or strong?

1. When X and Y area assimilar!
   Me rolling in mud and pigs rolling in mud?
   Rejection: But again, we have shared neurology - advil works on dogs as well, also animal experimenting!

2. when there are multiple Ms that can cause B.
   Rejection: what caused whimper if not pain? Maybe some other traits are vague, but pain isn't.

Carruther offers his Modified Argument on Consciousness:

P1) Some object Y is performing flexible and adaptive behaviours

P2) When I'm performing flexible and adaptive behaviours, I am conscious.

$\therefore$ Y is conscious.

but he denies P2, that we could have non-conscious experiences. We have experienced driving and becoming unconscious. The typical view is

$$\text{senses} \rightarrow \text{consciousness} \rightarrow \text{actionplanning}$$

but we could just have

$$\text{senses} \rightarrow \text{Consciousness}$$

and

$$\text{senses} \rightarrow \text{Actionplanning}$$

More Utilitarian Arguments Apr 26 2023

Also present the Future Happiness Argument:

P1) An action is right to the extent it promotes overall happiness

P2) Killing a creature deprives it of future happiness

P3) Killing a sentient creature reduces overall happiness

$\therefore$ Killing a sentient creature is wrong

Says it's wrong to kill a baby or be suicidal! However, the question remains to define "overall happiness." We have a few ways:

- Totalism: the total amount of happiness matters, including that of possible beings. Linked to longtermism.
- Prior Existence: total happiness of existing being that matters.

Kant May 1 2023

Very different way of answering why it is wrong to kill something...It **violates one's free and rational nature**. Therefore rationality stands central to Kantian morality. Not sentient or non-sentient, but rational or nor rational. Only rational creatures have moral status.

We first see Kant on animals:

P1) We only have direct moral duties to rational creatures.

P2) Non-human animals are not rational creatures.

$\therefore$ We have no direct moral duties to animals.

Kant himself is rather vague on whether we should believe this...also what is rationality and why it is importance. Also what is "direct"?

If you set fire to a dog, why is it wrong? You cannot wrong a dog by Kant since it has no moral status. Kant says it's wrong because it gets you used to cruel actions and become more likely to eventually apply to people.

Why art and flower are good is because he thinks we appreciate them for its own good but not to use them for use or for means...his **categorical imperative** that we should treat people as ends not means.

Kantian Ethics 101

Starts again from the building blocks of Ethical Theory: we summarise both traditions first

	theory of value	theory of right action
Utilitarian	Hedonism and impartiality	consequentialism and maximization
Kantian	good will of autonomous agents	confer value by rationally endorsing things

Kantian theory of value: the only thing that is good without qualification is the **good will of autonomous agents** (the intention on which you act). All good actions without the good will are worthless or even bad. A point of support is the pleasure of killing: utilitarians would say the killer's pleasure is good. Kant: dude no, not without the good will:)

Kantian theory of right action: because we value something it becomes valuable - we confer value on things by rationally endorsing them. Ex. What's the intrinsic value of climbing Everest? There are people who care and think it's worthwhile, they confer value on this action.

Kant's ultimate train of thought: we don't have control over the maximization process since it's global. We should focus on things we can control - our intentions.

The intentions are done from moral duty, which duties are done from the conception of the moral law - the **Categorical Imperative**.

In all, Kantian Ethics is good intentions that correspond to the categorical imperative.

Though it looks like Kant has some moral laws from nowhere, he is proposing somewhat of a democracy: there are no moral laws *a priori*, but we all come up with moral laws together from CI. We introduce two version of CI:

CI Ver. 1. Universaliability Principle:
"I am never to act otherwise than so that I could will that my maxim should become a universal law."

Consider an action: borrow money without intention to repay.
Kant: if I think this is good, then I'm endorsing the **maxim** that people should make promises with intent to not keep them. Will that maxim be universal to everyone? (would you be okay with everyone doing this?) But I cannot will this maxim. Such world will have no promises.

Kant 101 Continued May 3 2023

Back to CI:

CI Ver. 1. Universaliability Principle:
"I am never to act otherwise than so that I could will that my maxim should become a universal law."

Consider an action again: eat animal when hungry. Let's see if it fails! A test fails if

1. Leads to logical contradiction
2. Korsgaard: If everyone did it, would decrease the effectiveness of your action
3. I would object to others doing it...(not so Kantian though)

Looks like eating animal is okay. However, these two don't rule out lots of important stuff, especially when it's "natural" and "conventional" actions. This universal test doesn't show that natural action is wrong. It also fails to show that murdering human is wrong- so not a good test. We need a better one:

CI Ver. 2. Principle of Humanity:
"Act so that you treat humanity always as an end and never as mere means."

Consider scenario: we have a coffee machine and a barista. They are a means to me to get coffee. Any difference between how we have to treat them? We respect the barista since she has other needs, also cannot hurt Barista if no coffee is made. The barista does not just exist for coffee - there is a life that matters - an end but not a mere mean. We need to respect people as their **rational autonomous nature**. This rules our lying, slavery, non-consent (central), etc.

Serial grandma killer case: if a grandma killer hops by and ask if your grandma is upstairs. You cannot lie since it will disrespect the rational agency of the killer, by violating their rational decision-making process. Although you not lying would get your grandma killed, you should not be responsible for what bad people do taking advantage of your goodness.

The mistake made against the Principle of Humanity is against rationality. If I deem a serial killer irrational, I'm making a clear rational decision.

What about being rational or what is rationality? Suppose a situation, you performed an action. The relationship between the situation and rationality is **principle**. A rational agent is one that is "conscious of principles on which we act" and evaluate whether the principle is rational or good reason to act, and whether is supports the action you're considering. For irrational, their action is just instincts.

Therefore, we define **rationality** as the capacity to conceptualize reasons for acting, endorse or reject them.

Kant & Korsgaard May 8 2023

Let's conclude Kant: due to the second CI, we cannot treat people as mere means, but we could do that for animals. Humanity matters because of rationality, animals fail this standard because they aren't rational.

but the key question is: why does the Principle of Humanity only apply to rational creatures? We take a few tries:

1. Treating rational persons as means is factual error.
   Rational things act because we choose to, not programmed to do. Treating people as mere tool is wrong and for animal is okay
   So sexual assault is wrong because the criminal treats the victim as mere means - but sounds insufficient and the pain of the victim should also matter
2. can't violate consent if it can't consent.
   bad for people if we substitute sex with a child - total wrong
   Does consent for animal matter? Are they like sleeping people or a water bottle?

One try is in the form of The Value Argument, which gets us closer to the heart of the problem. Kant says there was no value before human, we perceive things and decide whether it is valuable. Therefore, **value is a human creation, made possible by rationality**.

1. All value is created and conferred by (legislative) actions of rational creatures.
2. Animals are not rational creatures.
3. Therefore, animals cannot confer value on their own lives.

$\therefore$ Animals are not ends-in-themselves.

Some objections:

1. To 1.: things could have natural good - not a Kantian response.
2. To 2. and 3.: animals act for their own good, have conception of self, could be rational enough to confer values.
3. Korsgaard: Objection to Validity

Objection to Validity by Korsgaard. We confer value in animal lives - we make them as ends.
We first confer value on animal nature - I want to get fed, being comfy etc. So now animals getting the same thing becomes important.

Kant & Korsgaard & Nussbaum May 10 2023

Recall Kant's Argument again

1. All value is created and conferred by (legislative) actions of rational creatures.

2. Animals are not rational creatures.

3. Therefore, animals cannot confer value on their own lives.

$\therefore$ Animals are not ends-in-themselves.

Some responses by Korsgaard:

1. Argument is invalid:
   a) "same"
   b) because human confer value on our animal nature (ex. eating), aspects of animal nature are valuable
   c) Animals share aspects of our animal nature
   $\therefore$ animal lives have value

Korsgaard distinguishes between active (can vote, active, etc) and passive (children) citizens. Active citizens make laws that govern passive citizens.

Nausbaum's objections #1 to Kant & Korsgaard:

1. too indirect.
   Doesn't seem to get why aniamls are really valuable. It is wrong thtat animals only get value from us. If humans never existed, there will be no value in the world. If we were androids, no animal values either.

2. too anthropocentric.
   We may not value all the right things - we only confer value on things we share. This excludes goods for animals we lack. Mantis eating each other, etc. We should interfere in a way so that the animals flourish.

Korsgaard's response #2, which Naussabaum likes a lot more: animals confer value on their own lives as ends-in-themselves, they do it in a way not how we do it (not rationally). Ex. animals strive to survive, perceive things as good or bad for them, their actions presuppose they matter.

Nasusabum's BIG objection to Korsgaard: "Our moral capacities are themselves animal capacities." Our rational nature is a part of our animal nature, not separated or exalted. Rationality should be considered only as one evolutionary path. Also accompanied by another claim that components of rationality can be found in animals - rationality is like a continuum, not Kantian picture that you either be rational or not at all. Ex. Kant thinks what's amazing about us is universalizability or to think in other perspectives, but that also exist in animals. Also the ability to conform to norms and morality.

Nussbaum's Capabilities Approach May 15 2023

Kant's view on psychological capacities and moral status, you can have moral status when you can

1. make active decisions

2. sense of self $\rightarrow$ "end-in-self"

3. universalize and affirm actions as norms for all
4. have good will & care about others

In Kant's favour, animals (ex. chimps) have these things:

1. decisions not impulse
2. sense of "I" that persists through time
   "mirror test" - maybe passing that is the necessary and sufficient condition for animals to have a concept of self?
3. social norms

Nussabaum's capabilities approach: developed originally as human good
A political system is just only if it secures to each individual a minimum threshold amount of central capacities (substantial freedoms, opportunities for choice and action in areas individuals deem valuable).

Come back to the two components: theory of value (what's valuable) and theory of right action (what to do *vis a vis* those valuable things)

Theory of Value	Theory of Right Action
List of species: specific capacities ability to choose & act & achieve	

We need to provide a way to animals in a way so that they flourish in their own ways.

The Capabilities Approach Against Utilitarianism: it cannot be aggregated over individuals and things aren't just pleasure and pain.

The Capabilities Approach Against Kant: it is not just autonomy but material outcomes.

Continued Nussbaum's Capabilities Approach May 22 2023

A few questions to consider:

Q1. Why does it extend to animals?
value: significant striving

A1. S has moral status iff S has the "standard animal package":
sentience & emotion & cognitive awareness & movement toward good and away from bad

Q2. What would the capabilities approach look like extended to animals?
THe "form of life" capabilities list, which starts off by being species-specific.

Continued Nussbaum's Capabilities Approach May 24 2023

Oftentimes we have conflicts between animals and humans, the CA is an useless theory if we don't have these conflicts outlined. We pull a list:

1. Sustenance/Territory: elephants destroying crops
2. Charitable Efforts: pet vs. human rescue
3. Religion/Culture: Whale hunt, Halal
4. Economics Cost for Implementing CA
5. Safety

Nussbaum calls many of these "tragic conflicts": when multiple things of value can't all obtain (conflict) usually through no one's fault. Some strategies for those are:

1. Analyze: does the resolution necessarily push one party below minimum threshold of flourishing
2. if tragic, minimize loss in this case
3. "Sublate dilemma" we should seek to change the structure of the world so that they don't arise in the future

So the argument now turns to if human should intervene in nature. We first have an Argument Against Intervention

1. There is value in wildness
2. Intervening to protect wild animals would distrupt wildness

$\therefore$ We should not intervene to protect wild animals

A suggestion for value is evolution.

Moral Status of AI May 31 2023

Question: what kind of systems have moral status that we need to take care of?

Master Argument of AI's Moral Status:

1. S has moral status iff S has psychological property/capacity P.
2. For AI system S that has P.

$\therefore$ AI system S has moral status.

We can also run an argument against this by saying that AI systems lack P, so AI systems lack moral status.

Let's step back and examine the properties. Currently, we have three different theories on what P is

1. Utilitarians: Sentience, qualia, ability to feel pain/pleasure

2. Kant: Rationality and autonomous decision-making

3. Capabilities Approach: Sentients and things can be good or bad from the point of view of S.

There seems like a problem with our ability to verify #2 premise in the master argument. Instead, Andrews and Burge think should go for the normal Marker Approach to Sentience.

1. When humans do B, they are in pain.

2. Animal S is doing B

∴ S is in pain.

This sounds like a good case for animals. This is justified as an inference to the best explanation. In terms of AI, this argument turns bad. If we write it out:

1. When humans do B, they are in pain.

2. AI system S is doing B

∴ S is in pain.

since now #2 is no longer the best explanation. A better explanation would be that AI is not in pain but has been programmed to behave like us. So either Siri is saying sorry because it actually is, or it's just a program.

One explanation for it not being the best explanation is the Problem of Gaming (def. non-sentient systems using human-generated training data to mimic human behaviours likely to persuade humans of their sentience.) Because AI can game us, we cannot treat their behaviour like sentience.

Continued Moral Status for AI Jun 5 2023

Recall the argument

1. S has moral status iff S has psychological property/capacity P.

2. For AI system S that has P.

∴ AI system S has moral status.

But a couple problems follow: How do we know if 2) is true?

1. Epistemelogical Problems:
   What is P? We assume consciousness is necessary human-like consciousness is sufficient.
   The Gaming Problem: AI seem conscious
   The N=1 problem: all conscious creatures that we observe are like us and came from the same lineage. If Ai is conscious, its consciousness may be very different.
   Currently, there is no consensus on theories of consciousness

Now the second question: what should we do if we don't know?

1. It is likely that we will soon have AIs of debatable personhood, once it's reasonable to believe it is a person and it is also reasonable not to.
2. A couple possibilities: full persons + full rights, a mistake could be persons + no rights.
3. OR not persons + full rights also a mistake, and not persons + no rights to be correct

Arised is the Full Rights Dilemma for AI:

1. Either we give or grant AIs of debatable personhood and rights or do not.
2. If we don't give them rights, we risk moral catastrophe. (full personhood but no rights)
3. If we do give them rights, we risk moral catastrophe. (could give full rights to not a person)

$\therefore$ AIs of debatable personhood necessarily carry the risk of moral catastrophe.

Continued Moral Status of AI Jun 7 2023

We continue on the Full Rights Dilemma for AI, focusing on premise 2.

- If we don't give rights, but they are persons/conscious.
  Machine revolt, moral catastrophe on us by creating a race of disposable slaves, murder, # of AIs >> # of humans, devalues conscious life

now what about premise 3? A claim: sentient AI will be in existential conflict with humans.

- enabling our own takeover
- save robots over people
- right to vote

What about giving partial rights? Maybe just don't make AIs of questionable personhood.

Morality might have to grow, just like Aristotle's physics to General Relativity.