Phil 4: Intro Ethics — April 12, 2016
Handout #3: Critiques of Consequentialism — Prof. Aaron Z. Zimmerman

(1) An initial distinction: Intrinsic vs. Instrumental Value

> X has intrinsic value if and only if X has value and X would have value even if it had no valuable consequences or produced nothing valuable distinct from itself.
>
> X has merely instrumental value if and only if X has value but X would not have value if it neither had valuable consequences nor produced anything valuable distinct from itself.

Examples: Money (or rather having money) has merely instrumental value. Pleasure seems to have both intrinsic and instrumental value. It's good "in itself" and also good when (e.g.) the pleasure that comes from accomplishing something difficult or the pleasure of earning parental approval reinforces the disposition to work hard or perform prudent and virtuous actions. Is there anything that has only intrinsic value? (Mill and Aristotle say that only happiness has wholly intrinsic value—are they right? How must we understand 'happiness' to make them right?)

(2) Consequentialism I

'Consequentialism' is the name for a family of ethical views—utilitarianism is just one of them.

> Consequentialism (Ends Def.): The only things with intrinsic value are certain states of affairs; everything else that has value owes its value to its relation to intrinsically valuable states of affairs.

Question: What's the problem with this definition according to Williams?

(3) Different Versions

Different versions of consequentialism result from different views about what is intrinsically valuable. Williams argues that consequentialists must admit that certain actions are themselves intrinsically valuable if they are to escape gross Hedonism. If a consequentialist thinks, e.g., that a person's **happiness** is intrinsically valuable, he must allow that certain actions are intrinsically valuable, because (Williams argues) happiness is not a sensation produced by certain activities; it is instead a matter of **actually doing things** with a certain amount of enjoyment.

> "To say a man finds certain actions or activities pleasant, or that they make him happy, or that he finds happiness in them, is certainly not always to say that they induce certain sensations in him, and in the case of happiness it is doubtful whether that is ever what is meant. Rather it means such things (among others) as that he enjoys doing these things for their own sake."

Nozick makes a similar point when describing his experience machine. "Would you plug in? What else can matter to us, other than how our lives feel from the inside?" If 'happiness' is used to cover all (or most) of what matters to a person (even when she is thinking in a largely self-interested way) happiness must include more than mere experience. But doesn't understanding "happiness" so that it refers to something non-mental—something other than a state of mind—conflict without ordinary understanding of the term? So shouldn't we conclude from Nozick's experience machine that we place intrinsic value on things other than happiness? Moreover, as Nozick points out, the question remains, "Why do we want to do the activities rather than merely have the experience of doing them?"

(4) Consequentialism II

> **Consequentialism (Maximizing+Uniqueness)**: An action A performed by a person P at a time t is the morally right action for P at t if and only if P's performing A produces more intrinsic value than any other action available to P at t.
>
> **Consequentialism (Maximizing Only)**: An action A performed by a person P at a time t is a morally right action for P at t if and only if P's performing A produces as much intrinsic value as any other action available to P at t.

Williams says that these definitions make the moral rightness of an action *objective* rather than *subjective*.

> An action A is **available** to a person P at a time T if and only if P is physically capable of performing A at t.

Non-culpable ignorance: P may be physically capable of performing A at t but be *unaware* that he can A at t. If P is not to blame for being unaware that he can A at t, then even if Aing is the right thing for P to do at t, P is not to be blamed for failing to A at t.

Questions: Can you think of an example in which someone is excused from blame despite not doing what is right? How does Mill distinguish the narrow class of blamable actions from actions that just fail to be morally best? Recall the following passage:

> "We do not call anything wrong, unless we mean to imply that a person ought to punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience. This seems the real turning point of the distinction between morality and simple expediency.

(5) Rejecting Consequentialism

Williams first claims that non-consequentialists might want to simply talk about the rightness and wrongness of actions and refuse to evaluate the moral properties of the consequences or "states of affairs" that result. But these consequences will include the feelings and emotional responses of the people affected by the act, the quality of life that results for those people, and so on. So it is *absurd* to suppose that these items might be given no moral weight or importance.

But Williams goes on to suggest that non-consequentialists (e.g. Kantians) might just reject the maximizing conception of right action altogether.

> "A non-consequentialist can hold both that it is a better state of affairs in which more people keep their promises, and that the right thing for X to do is something which brings it about that fewer promises are kept."

Questions: Can you describe such a case? How can someone admit that X-ing would produce more value than not X-ing, that X-ing will in the long run lead to better results than not-Xing, and still think it not right to X? What about hurting one person to save many? Consider here Williams' examples of George and Jim. What are your intuitions about these cases?

(6) Limiting the Scope of Non-Consequentialism

What is it to say, "No action is absolutely right"? What is it to say that there are no absolute truths in morality?

> (1) **Brute Relativism**: A person P is right to A in community c at time t, while (Temporal) a relevantly similar person P' is wrong to A in community c at time t+n, or (Spatial) a relevantly similar person P'' is wrong to A in community c'≠c at t, where there are no morally relevant

differences between the situations faced by P and P' or P and P'' except for the times at which or places where their respective actions are performed.

Question: Can you think of an example that shows brute relativism to be true? How could brute relativism be true? Doesn't it conflict with the principle we called "Rationality as Conceptual Supervenience" when discussing Thomson and her arguments from analogy?

> **(2) No Simple Answers**: There is no type of action such that it is always and everywhere right to perform that type of action.

Problem: What about the following type of action: the right action? If we sort actions into the right type and the wrong type we can say that it is always and everywhere right to perform the right type of action and always and everywhere wrong to perform the wrong type of action.

> **(2') No Simple Non-Trivial Answers**: There is no type of action, other than those that are semantically (or linguistically) guaranteed to be always and everywhere right, such that it is always and everywhere right to perform that type of action.

Examples: It is right to help someone, but not if you know that your helping them will just enable them to harm someone else. (Don't lend your friend money if he's going to buy a gun to kill an innocent person.) It is wrong to kill someone, but not if you know that that person is going to kill you for selfish reasons and you have no other way to stop her from doing so.

Problem 2: But what if we build all the exceptions to a rule into our specification of an action type? We can say that it is always and everywhere wrong to kill someone for any reason unless it is necessary for self-defense, or one is a combatant in a justly fought war, **etc**. We can then say that the type of action picked out by 'a killing not done in self-defense, not committed against a combatant in just conditions of war, etc.' is always and everywhere wrong. We can say it is always and everywhere wrong to kill someone when that person is not trying to kill you, or that person is not fighting against you in a war you are justly pursuing, and so on.

> Potential Problems: (1) these highly "disjunctive" action types are (perhaps) not really *kinds* of action at all, or (2) these actions types cannot be easily denoted in non-moral vocabulary so as to be cited in sufficiently simple rules that we might *use* to guide our actions. (The 'etc' in our rule cannot be eliminated.)

I think (1) is a bogus concern. We already have simple vocabulary to pick out the relevant kinds or properties (i.e. with predicates like 'right' and 'wrong' that are already in our moral vocabulary), and it is rare that we have two distinct vocabularies that line up perfectly. We should expect that moral properties (like the property expressed by 'right action') would only be picked out by highly disjunctive predicates if these predicates were not part of our ethical vocabulary. But the second worry is a real one and may place limitations on the usefulness of moral theories or sets of moral rules that can be articulated.

> **(2'') No Simple General Non-Trivial Answers**: There is no type of action, other than (a) those that are semantically (or linguistically) guaranteed to be always and everywhere right, and (b) those that are picked out with highly disjunctive and hard to deploy descriptions, such that it is always and everywhere right to perform that type of action.

(2'') is probably right, but (one might say) big whoop. Is (2'') a surprising thesis? It does point out that moral decision-making is difficult precisely because the moral facts are extremely hard to describe in non-moral vocabulary.

There is, however, a more subtle way in which context can affect moral wrongness or (at least) some related property such as blameworthiness. For example, Thomas Jefferson had slaves. (Indeed, unlike George Washington, Jefferson didn't leave his slaves their freedom in his will.) But compare Jefferson with Wolfgang Priklopil, an Austrian man who kidnapped a 10-year-old girl and kept her enslaved in his

basement from 1998-2006. I take it that most people judge Jefferson and other slavers of his day less harshly than they judge Prikopil. Some of this is just ignorance: we judge some of the 18th and 19th century slavers less harshly than we would were we to learn the details of their crimes. But mightn't one justify adopting a less harsh attitude toward slavers of the past than slavers of the present by judging each **in relation** to his or her contemporaries?

> **Relativism About Character**: A person P is immoral (equivalently: P has a bad character) just in case P's motives and actions are significantly worse than those of others in her social context.

So, the relativist might say that Prikopil (or a contemporary European man who engages in human trafficking by importing young women from Southeast Asia and forcing them to work as prostitutes without just compensation) is immoral because significantly worse than his cohort in his treatment of people (where his cohort is taken to be other similarly situated European males). And yet the relativist might resist saying that Thomas Jefferson was (straightforwardly) immoral even though he too kept slaves if it is taken for granted that Jefferson's behavior towards his slaves and slavery in general, though worse than an abolitionists, was not significantly worse than the other members of his cohort (where the relativist might take this to be Southern property owners).

Questions: Is relativism about character true? To judge a person immoral do you compare him with others who are similarly situated in society? Or do you have some kind of absolute standard in mind? Mightn't our evaluations of others combine both? Might we have some minimal standards we think anyone must respect on pain of gross depravity no matter how challenging the moral environment they face whereas our overall assessment of a person also takes into account how well they coped with the challenges presented by their society in comparison to their cohorts?

Further questions: Mightn't one be a relativist about the goodness or badness of people (or their characters) and yet be a universalist about the goodness or badness of actions or motives or institutions? For instance, mightn't one argue that the desire to enslave others, or the act of enslaving another, or the institution of slavery is bad or wrong or something that shouldn't have been indulged, permitted or allowed, and that this is true of every such desire to enslave, act of enslavement or practice of slavery known to history? And mightn't one combine this view toward slavery (on which it is always bad) with the more nuanced, partially relativist evaluation of slave owners described above?

> **Moral Universalism About Certain Motives or Actions or Institutions**: There are certain motives, actions or institutions that have been, are and always will be bad or wrong to indulge or permit.

A further sense of "everything is relative":

(3) **Strong Act Consequentialism**: Whether an action is right is entirely determined by that action's consequences. So unless a rule proscribes or requires a type of action that is picked out by reference to its consequences—e.g., "Do the kind of thing that will lead to the best consequences"— it cannot be a good or useful moral rule.

Question: What would Mill say about (3)?