Philosophy 4 April 12, 2016
Handout #4: Kant's Moral Philosophy Prof. Aaron Zimmerman

1. The Preface

(a) A Threefold Division:

(i) Physics: A description of the behavior of physical systems. As the observed behavior of the universe is (somewhat) uniform, these descriptions can be (more or less) general in form. This enables prediction. Kant thinks much of this uniformity comes from the way our minds are set up. Space is the "form" of visual perception and other modes of outer sense. Time is the "form" of both outer sense and our introspective awareness of our own thoughts and feelings.

(ii) Logic: Rules for Thought. Logic has nothing to do with actual objects. Its laws hold regardless of what exists. (Although classical logic assumes that at least one thing exists insofar as this is one of its axioms.)

(iii) Ethics: Rules for action. Just as logic is the development of laws proscribing how we ought to reason, ethics is the development of laws proscribing how we ought to act or **will**: i.e. choose rules to live by.

Kant claims that Physics and Ethics will have two parts to it: (1) an entirely formal part that can be known **a priori**; and (2) an **empirical** part that takes into account facts about human and physical nature and uses these facts to come up with more specific descriptions and generalizations.

(b) <u>A priori vs. A posteriori</u>

A priori: S knows P a priori if and only if S's justification for believing P (or her reason for believing P, or the evidence on which she believes it) does not concern or involve *sensory experience*.

A posteriori: S knows P a posteriori if and only if S's justification for believing P ultimately concerns or involves sensory experience.

<u>Analytic v. Synthetic</u>

Kant thinks a statement is analytic iff its falsehood entails a contradiction.
Kant thinks a statement is analytic iff the concept that delimits its subject is in some sense "contained" within the concept associated with its predicate.

According to Kant, when a statement's falsehood entails a contradiction, the truth of that statement can be demonstrated by producing an "analysis" which brings to light the structure of the concepts involved in the statement.

For subject/predicate sentences of the form "A is F" (so long as a determinate (descriptive) concept is associated with the subject term A and the predicate F): "A is F" is analytic just in case the concept associated with "F" is one among the *marks* (or sub-concepts) that together constitute the concept associated with "A."

Example: "Bachelors are unmarried"

The concept associated with the subject term "bachelors" has among its constituent marks the concept "unmarried." This is because the concept associated with the term "bachelors" is actually a complex concept itself composed of the concepts *unmarried* and *male*.

Synthetic: Negative definition: A synthetic truth is any truth which is not analytic. Positive definition: A synthetic truth is one in which the concept associated with the predicate is added to the concept that delimits the sentence's subject.

Kant thinks arithmetic truths are synthetic: i.e. not analytic; i.e. not demonstrable via definitions and logic alone. But we have a priori knowledge of these truths. **Fundamental moral laws are supposed to be the same as arithmetic truths in this regard**. Just as we know a priori the synthetic truth that everything that is equilateral must be equiangular, so too we know a priori certain synthetic (non-definitional) moral truths: e.g. that we can do what we know we ought even when this means acting against our own interests or even the interests of those we love.

Note: **Be careful not to confuse the a priori/a posteriori distinction with the analytic/synthetic distinction**. Kant claims that it is both a priori known and analytic that we only deserve credit or esteem when we act from duty. But Kant thinks it is non-analytic but nevertheless a priori knowable that we have the capacity of act from our sense of duty alone (i.e. without the aid of incentives or sentiments).

2. In Search of an Example of A Priori Moral Knowledge

In saying that the most fundamental laws of morality can be known a priori, Kant is saying that they can be known on some basis devoid of insight into anthropology and the study of human nature. He is also saying that they can be known on some basis other than experiences of pain/pleasure or misery/happiness insofar as these essentially involve sensations, and they can be known on some basis other than morally fraught emotions like love, sympathy, guilt and indignation insofar as these emotions essentially involve feelings. Again, according to Kant we have moral knowledge that resembles our knowledge of mathematical facts in its being grounded in (or based on) pure reflection or thinking alone. This is tied to his emphasis on fairness or "not making an exception of yourself" to a general rule or policy. Considerations of fairness of this sort are at least candidates for math-like knowledge. Considerations of happiness or utility aren't even in this running.

Potential Examples: (1a) It is wrong to act with the purpose of harming another. (1b) It is wrong to act with the purpose of burning another.

Both are true, but to know the truth of (1b) you need to know that *burning harms people*, which is a piece of a posteriori knowledge. You believe that burns harm people because you have **felt** the pain of a burn, or you have **seen** the damage that burning causes (or you were told as much from someone who observed the damage in question). Your justification for believing burns harm people is that you have, say, seen burns, and seen the damage they cause.

Someone might claim that you don't need to know anything about the empirical facts about how humans are constructed to know that (1a) is true. And such a theorist might argue, on this basis, that our knowledge of (1a) is wholly a priori. Some forests thrive after periodic electrical fires that thin them. "Controlled burns" protect the forest as a whole from conflagration. It is not wrong to burn trees in a strategic way to augment the overall health of the forest. This suggests that to know that it is wrong to burn an X one must know something about the nature of Xs. But, on the face of it, it would still be wrong to **harm** the forest or act with the aim of harming it; so perhaps our knowledge that it is wrong to harm something being does not rest on empirical investigation, but is instead a priori.

Questions: Can you come up with an example of a kind of thing it would **not** be wrong to intentionally harm? What about beings that deserve to be punished? **Does anyone really deserve to be harmed?**

In his 2nd Critique (5:59-5:63) Kant distinguishes between wellbeing and goodness and he similarly distinguishes between woe and evil. **When he draws these distinctions Kant seems to suggest that (1a) is not knowable a priori.** Instead, only the modified claim, "It is wrong to act with the purpose of harming another who does not deserve to be harmed" has this status. And to figure out who *deserves* to be harmed we must employ concepts like fairness and justice—the very concepts Kant is trying to articulate when he advances the various formulations of his first principle of morality: i.e. the categorical imperative. A person deserves praise or esteem when she acts dutifully; she deserves blame or demerit when she shirks her obligations.

Kant's example of an a priori knowable wholly categorical moral law: "'You ought not lie,' is valid not

merely for human beings, as though other rational beings did not have to heed it; and likewise all other genuinely moral laws; hence that the ground of obligation here is to be sought not in the nature of the human being or the circumstances of the world in which he is placed, but a priori solely in concepts of pure reason, and that every other precept grounded on principles of mere experience, and even a precept that is universal in a certain aspect, insofar as it is supported in the smallest part on empirical grounds, perhaps only as to its motive, can be called a **practical rule**, but never a **moral law**."

Questions: Is "You ought not lie," more like "Do not harm another," (or "Do not harm another who does not deserve it") or more like "Do not burn another"? Can you imagine an alien creature it would be morally okay to deceive? **Is "You ought not lie" really true or valid when interpreted in full generality, or are there cases in which we should lie**? Don't we need to know something about the "circumstances in which a man is placed" if we are to determine with any confidence whether he might permissibly lie?

Most think Kant is wrong when he argues in "On A Supposed Right to Lie Because of Philanthropic Concerns" that it's wrong or immoral to lie even when one has excellent evidence that doing so is necessary to save a life. There's a crazed man at the door asking whether his intended victim is home. Surely, you should do your best to convince him that the target of his evil plan isn't home. It probably isn't even a *bad thing* that the murder-bent man at the door is deceived by your ruse, as it makes it less likely that he will involve himself in great evil. Does this mean that "Don't lie" is nothing more than a hypothetical imperative?

Two kinds of hypothetical imperative: ***Desire Dependent***: "Do X so long as you want, prefer or need Y." ***Desire Independent*:** "Do X so long as conditions C obtain" where C does not involve the addressee's desires or preferences.

"Do not lie" must be interpreted as "Don't lie unless it's necessary to save a life" to escape counter-examples to its truth like the murderer at the door. So, despite Kant's claims to the contrary, it's really a hypothetical imperative. But for all that "Don't lie" might still be a **desire-independent hypothetical imperative**. This depends on whether the moral permissibility of lying ever depends on the desires of the person considering lying.

For instance, **is it ever okay to lie to escape embarrassment?** Kant argues that this is never permissible; that a rule permitting such an exception to a prohibition on lying could not pass the test on maxims imposed by the categorical imperative. Against this, critics might point to contexts in which it would be overly harsh to call a lie told to escape embarrassment "immoral." (Can you think of such a case?) But Kant can at least make an argument for the desire-independence of the (fairly obviously conditional) imperative "Don't lie."

3. Empirical Ethics

Kant says there are two things for which we need the empirical part of ethics: (a) applying laws to instances, and (b) overcoming inclinations.

(1) Example of (a): You might know a priori that it is wrong to harm another person (who doesn't deserve harm), but if you don't realize that you have mono and that kissing an innocent person will make her sick, you can't use your a priori knowledge of the wrongness of harm to stop you from kissing her. (Kant calls the faculty that applies rules to instances "the faculty of judgment.")

> * Note that on Kant's account the **faculty of judgment** accomplishes these tasks by applying rules or concepts to instances. You use your judgment to tell whether the concept "unfair" or "harmful" applies to the particular action under review (i.e. kissing S).

(2) Example of (b): Suppose alcoholism is partly genetic in origin. And suppose you know that it is true that it is okay to drink a moderate amount of alcohol but morally wrong to waste your life and talents by becoming an alcoholic. (Kant certainly thought this.) Still, if you don't know about your innate propensity

toward alcoholism, you may not be able to use your a priori moral knowledge to stay away from the drink altogether.

4. The Groundwork's First Section: Transition from Common Sense Morality to Genuine Moral Knowledge

(1) The Good Will Claim 1: "There is nothing it is possible to think of anywhere in the world, or indeed anything outside it, that can be held to be good without limitation, excepting only a good will."

(2) Three Questions: Question A: What is the will? Question B: Why aren't there other things that are good without limitation? Question C: What makes a will good?

Kant begins by telling us what the will is *not*: The will must be distinguished from "the talents of the mind": understanding, wit, and intelligence. The will must be distinguished from what Kant calls "temperament": courage, *resolve*, and heroism.

Since "resolve" is close in meaning to "will power" this is some indication that Kant's understanding of the will is non-standard.

Three definitions:

Definition 1: "Only a rational being has the capacity to act *in accordance with representation of laws*, that is, in accordance with principles, or has a *will*. Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason." GMM, 4:412.

Definition 2: "The will is thought as a capacity to determine itself to action in conformity with the representation of certain laws. And such a capacity can be found only in rational beings" (GMM, 4:428).

Definition 3: "Reason is concerned with the determining grounds of the will, which is a faculty either of producing objects corresponding to representations or of determining itself to effect such objects (whether the physical power is sufficient or not), that is of determining its causality" (2nd Critique, 5:15).

An answer to Question A: The will is the faculty we use to make plans and decisions, adopt rules and policies, and resolve to stick to these plans and policies. It is (in Kant's language) the faculty with which we **set ends** for ourselves.

Answer to Question B: Pleasant or admirable talents and temperaments can be used for evil, so too can fortune, wealth and honor. And when someone has these gifts but does not deserve them, this is unjust and bad.

But what about happiness? According to Kant, **happiness is not good without limitation** because: (i) it isn't good when a bad person is happy, and (ii) happiness can lead to arrogance and it isn't good when it does this.

Moreover, Kant argues, we can see that the end or *goal of human life* cannot be happiness with the following argument. (Notice that he disagrees with Mill about this.)

5. The Teleological Argument

(1) Everything has a purpose or proper function.
(2) The good for a kind of thing is given by its end or purpose: it is good for a thing to fulfill its purpose.
(3) The proper function of people cannot be living in happiness because we are poorly designed for this. We have reason and the kind of free will (autonomy) that depends for its exercise on reasoning about which ends to pursue and which policies to follow. But happiness is better achieved by inclination and instinct.
Therefore,
(4) There must be something good about reason and autonomy—any goal or end that does not necessarily

involve them is better reached by an automatic mechanism (e.g. inclination or instinct) set up to achieve it.
Therefore,
(5) The end or proper function of all people (indeed, all rational agents) is the development and use of reason and the kind of autonomy that requires the use of reason.
Therefore,
(6) The (autonomous) will is good in itself and happiness is not.

(See too the 2nd Critique reformulation of the argument 5:61-2; and the related discussion of God at 5:147-8.)

"Even if through the peculiar disfavor of fate, or through the meager endowment of a stepmotherly nature, this will were entirely lacking in the resources to carry out its aim, if with its greatest effort nothing of it were accomplished, and only the good will were left over (to be sure, not a mere wish but as the summoning up of all the means insofar as they are in our control): then it would shine like a jewel for itself, as something that has its full worth in itself" (GMM: 4:394). Virtue in rags is virtue nonetheless.

6. The Good Will

We now turn to Question C: What makes a good will good? How can we tell a good will from a bad one?

Claim 2: A good will is the will of someone who acts, chooses, or resolves to act *from* duty.

"We put before ourselves the concept of duty, which contains that of a good will, though under certain subjective limitations and hindrances which, however, far from concealing it and making it unrecognizable, rather elevate it by contrast and let it shine forth all the more brightly" (GMM, 4:397).

Four Kinds of Willing:
(1) Acts done contrary to duty.
(2) Acts done in accordance with duty but from an ulterior (non-moral) motive.
(3) Acts done in accordance with duty from benevolent inclination.
(4) Acts done from duty (i.e. respect for the moral law).

Willing of the first type obviously isn't good. Example of (2): **The Prudent Shopkeeper**. Example of (3): **The Instinctively Benevolent Shopkeeper**. Example of (4): **The Miserable Wretch** who continues his life from duty.

The important distinction: actions done **in conformity with duty** vs. actions done **from duty**.

"To be beneficent where one can is a duty, and besides this there are some souls so sympathetically attuned that, even without any other motive of vanity or utility to self, take an inner gratification in spreading joy around them, and can take delight in the contentment of others insofar as it is their own work. But I assert that in such a case the action, however it may conform to duty and however amiable it is, nevertheless has no true moral worth, but is on the same footing as other inclinations . . . Thus suppose the mind of that same friend of humanity were clouded over with his own grief, extinguishing all his sympathetic participation in the fate of others; he still has the resources to be beneficent to those suffering distress, but the distress of others does not touch him because he is sufficiently busy with his own; and now, where no inclination any longer stimulates him to it, he tears himself out of this deadly insensibility and does the action without any inclination, solely from duty; only then does it for the first time have its authentic moral worth" (GMM, 4:397-8)

"Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it

possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not; but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him" (2nd Critique, 5:30)

See too the extended discussion in the 2nd Critique of a man who refuses to give false witness against Anne Boleyn despite the suffering he and his family are forced to endure because he refuses 5:155-9

Four interpretations of what it is to deserve moral praise or esteem:

(i) ***Actual***: S's act A has moral worth (i.e. deserves praise and esteem) if and only if S recognizes that A is her duty and S does A in the absence of any inclination to A (and, perhaps, in the presence of an inclination to refrain from A).

(ii) ***Counterfactual***: S's act A has moral worth if and only S recognizes that A is her duty and either S has no inclination to A or S does have an inclination to A, but S would have performed A even if she had no such inclination.

(iii) ***Causal Actual***: S's act A has moral worth if and only S recognizes that A is her duty and it is this recognition alone that moves her to A, where any other desire, inclination or motive to A she has plays no role in motivating her action.

(iv) ***Causal Counterfactual***: S's act A has moral worth if and only S recognizes that A is her duty and this recognition moves her to A, where her A-ing may be also be motivated by other desires, inclinations or motives, but where S would still have been moved to A by her recognition of her duty even if these other motives hadn't been present. (In such a case, S's A-ing is *causally over-determined*.)

Tentative evidence for "actual" interpretation of some sort: 2nd Critique; 5:72.

Evidence for a causal interpretation: duty as "proper moving force" 2nd Critique 5:88; cf. 5:54-5:58.

Questions: How do the actual interpretations (i and iii) square with Kant's repeated claim that the cultivation of sympathy and other pro-social or pro-moral emotions and inclinations is a moral duty?

"Sympathetic joy and sadness (*sympathia moralis*) are sensible feelings of pleasure or displeasure (which are therefore called "aesthetic") at another's state of joy or pain (shared feelings, sympathetic feeling). Nature has already implanted in human beings receptivity to these feelings. But to use this as a means to promoting active and rational benevolence is still a particular, though only a conditional duty. It is called the duty of humanity (*humanitas*) because a human being is regarded here not merely as a rational being but also as an animal endowed with reason. Now, **humanity can be located whether in the capacity and the will to share in others feelings (*humanitas practica*)** or merely in the receptivity, given by nature itself, to the feelings of joy and sadness in common with others (*humanitas aesthetica*). **The first is free, and is therefore called sympathetic (*communio sentiendi liberalis*); it is based on practical reason**. The second is unfree (*communio sentiendi illiberalis, servilis*); it can be called communicable (since it is like receptivity to warmth or contagious diseases), and also compassion, since it spreads naturally among human beings living near one another. **There is obligation only to the first**." (MM, 6:456-7)

Question: Can you explain the distinction Kant draws between these two different forms of sympathy, only one of which is grounded in an exercise of our capacity for practical reasoning and is therefore "free" in the sense Kant has in mind?

Argument that Kant Embraces a Counterfactual Characterization of Meritorious Action: Might Kant think that we have a duty to rob ourselves of humanity (*humanitas practicas*)—so as to provide us with occasions on which we can act from duty—even though we will therein make it less probable that we will live virtuously? This would be like telling someone with a gambling problem to move to Vegas so she can multiply her opportunities to resist temptation and therein display the kind of virtue on which Kant is

focusing. I think this interpretation is exceedingly uncharitable, but it is necessary if we are to consistently interpret Kant as arguing that a desire to help someone is incompatible with exercising praiseworthy (or meritorious) benevolence toward that person. **The more charitable reading is therefore one of the counterfactual interpretations articulated above**.

7. The Importance of Reasoning or Reflection

Claim 3: Actions that are done from duty derive their moral worth from the **maxim** or policy that leads a person try to perform them; they do not derive their worth from (a) successful completion or (b) any desired end.

In the above passage from MM Kant makes the distinction between mere "emotional contagion" and the choice to keeps one's "heart open" to the feelings of others. The passage makes clear that this later choice is an (imperfect) duty and that the resulting actions will be virtuous (and presumably deserve praise) if they truly have their "ultimate source" in this choice—so long as the choice is itself grounded in respect for the moral law.

So suppose someone reasons her way to an initial decision to become a sympathetic person and then, on this basis, develops habits of caring for others and responding to their needs. But suppose that these habits of caring are then no longer dependent for their operation on the reasoned choice that made her indulge and cultivate her sympathetic tendencies. Suppose that what was done from choice is now done from habit. In such a case, Kant argues, virtue has been lost.

"Virtue is always in progress and yet always starts from the beginning. – It is always in progress because, considered objectively, it is an ideal and unattainable, while yet constant approximation to it is a duty. That it always starts from the beginning as a subjective basis in human nature, which is affected by inclinations because of which virtue can never settle down in peace and quiet with its maxims adopted once and for all but, if it is not rising, is unavoidably sinking. For moral maxims, unlike technical ones, cannot be based on habit (since this belongs to the natural constitution of the will's determination); on the contrary, **if the practice of virtue were to become a habit the subject would suffer loss of the freedom in adopting his maxims which distinguishes an action done from duty**." (MM, 6:409).

Questions: Suppose that Kant is right that acting from mere benevolent inclination or spontaneous love does not deserve praise (or at least does not deserve praise of the highest form). After all, non-human animals have pro-social inclinations and many of us resist thinking of them as virtuous or morally excellent despite these inclinations. (According to a common strain of thought, non-human animals can act neither morally nor immorally; they are proper objects of moral concern but not moral judgment.) So let us suppose, though it requires argument, that one must *reflect* on spontaneous motives, policies or entrained maxims and act from the conclusions (or outputs) of that reflective process if one is to truly deserve praise. This gets Kant the intermediate conclusion that one only deserves praise if one acts from reflective awareness of (or a self-conscious belief in) the value of one's actions, inclinations, or policies.

The Conscientious Utilitarian as a Counter-Example to Kant's Derivation of the Categorical Imperative from The Idea of Dutiful Action: But we might suppose that someone A is motivated or tempted to help someone else B out of love, and that when A reflects on the situation she concludes that love is not only a permissible motive but a good one (or that her loving action is not only a permissible action but a good one) because it promotes the happiness of both parties while detracting from no one's wellbeing. (We might even suppose—in accord with something like the counterfactual criteria—that A wouldn't have helped B if she had concluded that helping B wasn't good or permissible in a utilitarian sense.) Does the loving act of someone who has self-consciously examined and approved of her loving behavior in this way deserve esteem? Does our answer to this question depend on whether her belief in the permissibility of her actions is grounded in teleological (e.g. utilitarian) reasoning or deontic thought? Does it depend on the reasoning's having a wholly a priori element?

The utilitarian (or sentimentalist) might argue, against Kant, that helping from love is praiseworthy when it is buttressed by (and/or motivated by) sincere reflection on the value of acting from love that

culminates in self-conscious approval for this motive, but that acts of this kind needn't be driven by a sense of obligation or duty. (Cf. Kant's claim that God is a being who does not experience moral laws **as** obligations or duties because his will is **necessarily** good.)

Two Alternative Requirements for Meritorious Action:

(1) The Utilitarian Rationalist: S does not deserve praise or esteem for A if she performed the act from benevolent instinct alone. S only deserves praise or esteem for A if she deliberately performed A from her (sincere, deeply considered) positive assessment or evaluation of her acting from benevolent inclination in this circumstance.

(2) The Kantian Rationalist: S only deserves praise or esteem if she acted from a sense of moral duty or obligation.

Questions: It would seem that (1) does not entail (2) without auxiliary premises. So Kant must mount an independent argument against the sentimentalist or utilitarian who accepts (1) while rejecting (2). Does reflection on Kant's examples—of the shopkeeper who charges the fair price and the man who decides not to commit suicide—bridge the gap? How much of his view would Kant have to give up where he to allow that all actions performed from knowledge of their moral permissibility deserve praise whether or not this knowledge is accompanied by (or constitutes) a sense of duty or obligation?

A More Specific Formulation of this Question: What if a utilitarian helps someone in great need from his belief that it is his duty to do so, where this belief is inferred from his belief that he has duties to perform those actions which promote happiness? And what if he believes (contra Kant) that our having this obligation to promote happiness is self-evident or in some sense derivable from the fact that humans desire happiness (and only happiness) for its own sake? Does the utilitarian we have imagined deserve no credit for what he's done? Surely, Kantian ideology isn't necessary for moral merit. But is a sense of obligation grounded in thoughts of justice necessary? If we think in utilitarian (or axiological) terms and from this reasoning perform acts of great benevolence and charity, don't we deserve whatever merit might be given to those who achieve the same results after deontic reflection? Think of Singer's role in advancing the wellbeing of animals here.

8. The Categorical Imperative

The Categorical Imperative (1st Formulation): "I ought never to conduct myself except so that I could also will that my maxim become a universal law."

The Initial Derivation of the Categorical Imperative:
(1) The only unconditionally good thing is the good will.
(2) The good will is exhibited in acting from duty.
(3) Acting from duty consists in acting not from inclination or desire, but from respect for the moral law.
(4) With the exception of the categorical imperative, every practical policy (or maxim) is capable of being an object of desire or inclination. The categorical imperative is the only purely formal practical guide; it has no empirical content—no other possible end (or determination of the will) is like that.
Therefore,
(5) The only unconditionally good thing is acting in accordance with the categorical imperative.

What does (4) mean? "But what kind of law can it be, whose representation, without even taking account of the effect expected from it, must determine the will, so that it can be called good absolutely and without inclination? Since I have robbed the will of every impulse that could have arisen from the obedience to any [particular] law, there is nothing left over except the universal lawfulness of the action in general which alone is to serve the will as its principle. . ."

If one is acting from moral duty alone in choosing some policy, then the idea that the policy is a moral one has to be one's reason for choosing to apply the policy. Now if one has no further aim or end when choosing the policy, the moral value of the policy must be an intrinsic, non-instrumental or "formal"

feature of it.

What "formal" moral value could a policy have? Since the feature has to be entirely formal, it won't be anything about the goodness, or pleasure or happiness brought about by adopting the policy. These features are extrinsic, instrumental and empirically substantive. Kant reasons that the value a policy must have to be morally permissible is its **generality** or **universality**—this is the summum bonum of principles, and the very feature codified by the first version of the categorical imperative.

10. Hypothetical vs. Categorical Imperatives ". . . all imperatives are formulas of the determination of action, which is necessary in accordance with the principle of a will which is good in some way. Now if the action were good merely as a means to something else, then the imperative is hypothetical; if it is represented as good in itself, hence necessary, as the principle of the will, in a will that in itself accords with reason, then it is categorical."

First definitions: (1) A **hypothetical imperative** is an imperative of the form "Do X so as to achieve Y." (2) A **categorical imperative** is an imperative of the form "Do X."

Examples of hypothetical imperatives: "Exercise so as to maintain your health." "Invest your money so as to increase your wealth." (Is this an example? "Lie cunningly so as not to get caught.")

Second definitions: (1') A **hypothetical imperative** is an imperative of the form "Do X if you want to Y." (2') A **categorical imperative** is an imperative of the form "Do X (regardless of what you want…because reason demands it)."

Examples of hypothetical imperatives (so defined): "Exercise if you want to maintain your health." "Invest your money if you want to increase your wealth." Example? "Lie cunningly if you want to get away with it."

Remember the amendment to this distinction we introduced above: There are two kinds of hypothetical imperative. There are *Desire Dependent* hypothetical imperatives like "Do X so long as you want, prefer or need Y" and *Desire Independent* hypothetical imperatives like: "Do X so long as conditions C obtain" where C does not involve the addressee's desires or preferences.

Remember why we introduced this distinction: "Do not lie" must be interpreted as "Don't lie unless it's necessary to save a life" to escape counter-examples to its truth like the murderer at the door. So, despite Kant's claims to the contrary, it's really a hypothetical imperative. But for all that "Don't lie" might still be a **desire-independent hypothetical imperative**. Contemporary advocates of Kantian ethics might argue that moral rules are desire-independent hypothetical imperatives: these rules have exceptions but not exceptions for cases in which we just don't want to follow them.

11. Three Classes of Imperative

Rules of Skill: Hypothetical imperatives that cite an end other that happiness.

Examples: "Exercise so as to maintain your health." (Example? "Take cyanide so as to kill yourself.")

Counsels of Prudence: Hypothetical imperatives that have happiness as their end.

Examples: "Make friends so as to be happy." "Engage in pleasurable activities so as to be happy."

Kant claims that we (humans) necessarily have happiness as our aim, so every human is "bound by" counsels of prudence. That is, counsels of prudence apply to us all; we should all follow them given that we have happiness as our end.

Questions: Are we really all trying to be happy? Don't many of us follow rules or policies (regarding eating, sex, etc.) that actually undermine our long-term happiness? Do we follow rules of policies we know

to be imprudent? Consider your average smoker one month before New Year's Eve and compare his mind then to his mind on New Year's Day after he's resolved not to smoke. In both instances he knows he shouldn't smoke, but it's only upon forming his resolution that he has rejected the maxim "Smoke for pleasure" or "Smoke to appease your desire for the activity and its effects." Kant chalks this up to the smoker's not knowing whether smoking will make his life less happy overall than it would be were he to quit smoking. The lack of certainty allows a kind of self-deception where one hopes (against the odds) that one's smoking won't actually make one's life less happy overall and acts on that hope. (Kant's example is someone drinking despite his gout.)

In the teleological argument, Kant allows that we do not automatically make ourselves happy by inclination. (We have inclinations for things we know are not good for us.) Under what conditions will someone do what he must to be happy?

According to Kant, categorical imperatives don't just "bind" every human being in the way that counsels of prudence do; they bind every rational creature. And they don't just bind every rational creature because all rational creatures happen to share a common aim, end or goal like happiness. Instead, **categorical imperatives apply to all rational creatures regardless of what their particular ends happen to be.**

"There is one imperative that, without being grounded on any other aim to be achieved through a certain course of conduct as its condition, commands this conduct immediately. The imperative is **categorical**. It has to do not with the matter of the action and what is to result from it, but with the form and the principle from which it results; and what is essentially good about it consists in the disposition, whatever its results may be. This imperative may be called that **of morality**."

Commands of Morality: Categorical Imperatives which do not specify a means to an end, but instead represent a way of willing or deciding what to do—where deciding what to do in this way (or being disposed to decide in this particular way) **is identical with** having a good will.

12. Kant's question: How are all these imperatives possible?

What does this question mean? One possibility is that Kant is asking whether we are capable of acting from duty alone. Is it possible (and can we know that it is possible) for humankind to act morally from considerations of fairness or justice rather than self-interest, love, or benevolent emotion. This question is taken up in the final section of the Groundwork.

Another question that Kant might have in mind here (or elsewhere in the Groundwork) is, "What **makes it true** that we ought to do what each of these imperatives tells us to do?" What *makes it true* that we ought to seize upon opportunities to achieve our goals? What *makes it true* that we ought to deliberate about what to do in a moral way?

(a) Imperatives of Skill: It is *analytically* true that we ought to follow these. "Whoever wills the end, also wills the means (insofar as has decisive influence on his actions) the means that are indispensably necessary to it that are in his control. As far as volition is concerned, this proposition is analytic. . ." (p. 34).

What would you say to someone who said, "I've decided to become healthier and I know that I must exercise to do this, but I've decided not to exercise"? Kant says that *this person probably hasn't really decided to become healthier* as this would itself involve deciding to exercise (so long as we assume she knows or believes she must exercise to become a healthier person). If she has **really** decided to become a healthier person, and she **really knows** that exercising is necessary for this, then she **must** decide to exercise. This follows from the very concepts we associate with 'decision', 'means', and 'end.'

(b) Imperatives of Prudence: It also analytically true that we ought to follow these imperatives when they are determinate, but it is hard to come up with a precise imperative of prudence, because it is hard to figure out how to best achieve happiness. Imperatives of Prudence therefore reduce, in practice, to imperatives of skill.

(c) The Categorical Imperative: It is *synthetically* true that we are bound by this. But it can be known a priori. (Kant thinks the truths of geometry are like this too—they are synthetic a priori.) It is analytically true that the good will acts on the categorical imperative. It is synthetically true (but a priori knowable) that we ought to display good wills by acting on the categorical imperative. It is synthetically true (but a priori knowable) that, since we ought to respect Kant's categorical imperative, we can do it.

13. First Formulations of the Supreme Principle of Morality

The Categorical Imperative (**Universal Law Formulation**): "Act only in accordance with that maxim through which you can at the same time will that it become universal law."

The Categorical Imperative (**Natural Law Formulation**): "So act as if the maxim of your action were to become through your will a universal law of nature."

Kant makes it clear that he is not arguing for the claim that this categorical imperative (CI) is the supreme principle of morality by showing that acting on the CI leads to behaviors common sense judges to be morally good. According to Kant, **we shouldn't need empirical proof of this kind that we ought to act on the Categorical Imperative**. (Again, our knowledge of this fact is supposed to be non-inferential and a priori.) But he does think we can get a better grasp of the meaning of the categorical imperative by looking at particular examples.

Second example: Lying Promises.

(a) "Would I be content with things if my maxim (of getting myself out of embarrassment through an untruthful promise) should be valid as universal law (for myself as well as others), and would I be able to say to myself that anyone may make an untruthful promise when he finds himself in embarrassment which he cannot get out of in any other way? Then I soon become aware that I can will the lie but not at all a universal law to lie; for in accordance with such a law there would properly be no promises, because it would be pointless to avow my will in regard to future actions to those who would not believe this avowal, or, if they rashly did so, would pay me back in the same coin; hence my maxim, as soon as it were made into a universal law, would destroy itself."

(b) "For the universality of a law that everyone who believes himself to be in distress could promise whatever occurred to him with the intention of not keeping it would make impossible the promise and the end one might have in making it, since no one would believe anything has been promised him, but rather would laugh about every such utterance as vain pretense."

First Example: Suicide

Consider the maxim, "From selflove, I make it my principle to shorten my life when in the longer term it threatens more ill than it promises agreeableness. . . .One soon sees that a nature whose law it was to destroy life through the same feeling whose vocation is to impel the furtherance of life would contradict itself, and thus could not subsist as nature; hence the maxim could not possibly obtain as a universal law of nature, and consequently it entirely contradicts the supreme principle of all duty."

Third Example: Laziness

"Although a nature could still subsist in accordance with a universal law [which says 'indulge in gratification rather than trouble yourself with the expansion and improvement of your talents'], though then the human being (like the South Sea Islanders) would think only of letting his talents rust and applying his life to mere idleness, amusement, procreation, in a word, enjoyment; yet it is impossible for him to will that this should become a universal law of nature, or that it should be implanted as a natural instinct. For as a rational being he necessarily wills that all the faculties in him should be developed, because they are serviceable and given to him for all kinds of possible aims."

Fourth Example: Indifference

"Suppose one decides on the following policy, 'Let each be as happy as heaven wills, or as he can make himself, I will not take anything from him or even envy him; only I do not want to contribute to his welfare or to his assistance in distress!' A will that resolved on this would conflict with itself, since the case would sometimes arise in which he needs the love and sympathetic participation of others, and where, through such a natural law arising from his own will, he would rob himself of all hope of assistance that he wishes for himself."

Question: Apply the universal law formulation of the categorical imperative to each of these four examples. How plausible is Kant's claim that the universal law formulation of the categorical imperative indicates: (a) that suicide to prevent future suffering is impermissible, (b) that lying to escape embarrassment or distress is impermissible, (c) that a general policy of laziness is impermissible, and (d) that indifference to the suffering of others is impermissible?

I'm not asking whether you think suicide, lying, laziness, and indifference are **actually** morally impermissible, though you should think about that too. I'm asking whether policies licensing suicide, lying, laziness and indifference to others fail to be universalizable as natural law and, therefore, fail Kant's test for moral permissibility.

14. The Three Formulations of the Categorical Imperative

(a) **The Formula of Universal Law**: Only act in accordance with a maxim if you can at the same time will it as universal law.

(b) **The Formula of Humanity**: Only act so that you use humanity, whether in yourself or another, as an end (in itself), and never merely as a means.

(c) **The Formula of Autonomy**: Only act in accordance with maxims that would be enacted into law by a legislator and member [i.e. citizen] of a realm of ends.

14. The Universalizability Test

Maxim 1: Make false promises to get money.
Maxim 2: Kill everyone who creates a nuisance.

One might argue that Maxim 1 *cannot* be willed as universal law because a world in which everyone obeys maxim 1 is a world in which *everyone* issues false promises to get money. If I'm in this world and I try to issue a false promise, I won't succeed. Kant would have you suppose that you are the one responsible for the fact that everyone follows maxim 1. (Perhaps you are the creator of this world and you make it the case that everyone in it must follow maxim 1.) Suppose that while you're responsible for everyone following maxim 1, you're also trying to get money by making a false promise.

Two questions: (1) Is what you're doing **incoherent**? (2) Does the incoherence in question justify Kant's claim that maxims 1 and 2 **cannot** be willed as universal laws?

(a) One possibility is that it is *literally impossible* to issue a false promise in the world that you've created, because there are no promises in this world. To say that there are no promises in this world is to say that when the world contains a certain crucial amount of deceit, there can be no promises. If you're the one who is keeping Maxim 1 in force in this world (if you're *willing it as universal law*) **and** you're also trying to get money by lying, then you're trying to achieve some goal **while** making that goal impossible to achieve; and that's incoherent in a fairly straightforward sense.

(b) Another possibility is that it's not impossible to issue a promise in this world because some people would still be gullible. Still, by keeping Maxim 1 in force in this world you **undercut** your own attempt to make money off a lie. Remember that Kant's already concluded that insofar as you are rational you

couldn't will an end without willing the (known) means to that end. So you would be irrational to attempt to get money by lying without doing what you can to make this **likely**. You can only make it **likely** that people will buy your false promises by getting rid of maxim 1. So you cannot rationally will that Maxim 1 be a law of nature and (at the same time) will that you get money by making the false promise.

*For critical discussion see J. Sobel, "Kant's Compass"

Kant's Reconstruction of the Distinction Between Perfect and Imperfect Duties

Recall the distinction between perfect and imperfect duties we introduced when discussing Thomson's view of abortion. At 4:424 Kant argues that we have a perfect duty not to (1) deceive others or (2) commit suicide. This is reflected in the outcome when we employ the test imposed by the Universal Law Formulation of the CI. It is, Kant suggests, literally impossible for deception and suicide to exist as the general rule rather than the exception to it. (This corresponds to interpretation (a) above.) But Kant thinks we only have an imperfect duty to develop our talents and help others. It is not literally impossible for us to be lazy or indifferent as the rule rather than the exception to it, but, Kant says, we cannot "will" that these attitudes be the rule rather than the exception to it. (This corresponds to interpretation (b) above.)

15. Problems w/ the Universal Law Formulation

(a) Too weak: It has seemed to some critics that maxims 3 & 4 would pass the test imposed by the first formula of the categorical imperatives. But to act on either one of these maxims would obviously be immoral.

Maxim 3 (The Cautious Con Artist): Lie to those who are too stupid to realize it whenever you can make money by doing so.
Maxim 4 (The Cautious Nazi): Kill the Gypsies, Jews and Communists as long as you are powerful enough to get away with it.

(b) Too strong: Some critics allege that Maxims like 5 & 6—acted on by the Proud Gaucho—would not pass the categorical imperative test, but it seems fine to act on them.

Maxim 5 (Proud Gaucho): Go to every UCSB basketball game to cheer for your team.
Maxim 6 (Proud Gaucho): Eat at the UCEN everyday for lunch.

Our Question: Can Kant show how the first formulation of the categorical imperative when properly interpreted: (a) rules out (as immoral) acting on Maxims 3 and 4, and (b) allows (as morally permissible) acting on maxims 5 & 6?

A promising strategy for answering part (b) of the above question would involve introducing conditions into Maxims 5 and 6 to make them genuinely universalizable. Perhaps it is immoral to act on maxims 5 and 6 if they are not "hedged." One should only go to the game or the UCEN when doing so won't risk injury or harm to oneself or another (e.g. Covid). The idea, here, is that a maxim should include all of the normatively relevant components of the agent's ends and plans when acting.

> A maxim is a personal or subjective plan of action, incorporating the agent's reasons for acting as well as a sufficient indication of what act the reasons call for. When we are fully rational, we act, knowing our circumstances, in order to obtain a definite end, and aware that under some conditions we are prepared to alter our plans… A full maxim... makes all this explicit. (Schneewind 1992, pp. 318-9)
>
> When we use the categorical imperative ... we suppose that we are examining a maxim embodying the agent's genuine reasons for proposing the action, rather than [what are in the agent's view] irrelevancies ... that might let it get by the categorical imperative [or, I add, that might let the categorical imperative stop it]. (Schneewind, 1992, p. 321.)

Though Kant uses "Lie to escape embarrassment" and similarly short sentences as examples of maxims, perhaps he intends these as short-hand versions of the real maxims as Schneewind describes them. Taking this more comprehensive understanding of maxims to heart, we can reformulate Maxims 5 and 6 thusly.

Maxim 5': Go to every UCSB basketball game to cheer for your team, so long as there is sufficient space.
Maxim 6': Eat at the UCEN everyday for lunch so long as there is sufficient room.

Questions: What would the world look like if everyone adopted maxims 5' and 6'? Isn't this still a weird thought experiment? Isn't it still strange (and possibly incoherent) to imagine a world in which everyone roots for UCSB or wants to eat at UCSB?

This suggests a further restriction on the content of maxims if they are to be meaningfully evaluated for universalizability. Perhaps we should insist that maxims contain only "general" terms in some sense. This would rule out the use of "UCSB" and "UCEN," giving us:

Maxim 5": Go to every one of your favorite team's games to cheer for them, so long as there is sufficient space.
Maxim 6": Go to your favorite dining establishment everyday for lunch so long as there is sufficient room.

It does make more sense to evaluate the adoption of these maxims as universal laws, and perhaps we do get the intended result that their adoption is morally permissible.

This demand for generality in the statement of a candidate maxim also provides a promising strategy for solving problem (a). We can concede to the objector that we can imagine a world in which Maxims 3 and 4 are "willed" as the rule and "adopted" as such. The world in which Maxim 3 is in this sense willed as a universal law is one in which everyone accepts a form of "Social Darwinism" and most (if not all) people think that they're too smart to be taken advantage of by the others. The world in which Maxim 4 is willed and adopted as a universal law is one in which the Jews etc. agree that they are "undesirables" and agree that their extermination is best for the species. (This is a world in which the Holocaust was an even greater success than it was in the actual world.)

Question: Might a Kantian say that the Con Man or Nazi cannot really will that Maxim 3 or Maxim 4 be universal law because this would amount to treating humanity as a mere means rather than an end in itself? Would this imply that the Formula of Humanity is needed to give sufficient substance or content to the Formula of Universal Law?

Even if we can imagine worlds in which everyone preys on those who lack the power to defend themselves and we can imagine the powerful willing them as universal laws, perhaps we cannot imagine a world at which a more fully general policy of deceit or murder is universally legislated and adopted as the rule. Perhaps (though this is not obvious) we should agree with Kant when he considers—and rejects as incoherent—the universal adoption of a general policy of committing suicide whenever life is painful and there are no easily foreseen prospects for improvement.

Maxim 3' (The Incautious Con Artist): Lie whenever you can make money by doing so.
Maxim 4' (The Indiscriminate Murderer): Kill whoever you are powerful enough to kill.

Questions**:** Is Maxim 3' something you can will as universal law? Is Maxim 4' universalizable? Why or why not?

Further Questions**:** Suppose the discriminating Nazi allows that 4' is not universalizable but he insists that he acts on Maxim 4, not Maxim 4'. Or suppose that the cautious Con Man insists that he is acting on Maxim 3, not Maxim 3'. And suppose that the man in question insists that because of this, the maxim on which he is acting is indeed universalizable and hence morally acceptable. Can Kant avoid agreeing with the Nazi or Con Artist on this matter? **Must we appeal to something besides the universal law formulation of the categorical imperative to argue that a person's maxims must be stated in fully**

general terms if their universality is to provide an accurate test of the moral permissibility of adopting those maxims? How would this admission affect Kant's derivation of the categorical imperative from the idea of meritorious action from duty? (See section 9 above for that derivation.)

A Related Set of Questions: Suppose the Con Man explains that he is lying to A and not B because B is too clever to be fooled and A is not. Suppose the Nazi explains that he is killing the Jews and not the Catholics because the Catholics are too powerful to be exterminated and the Jews are not. What kind of mistake (if any) is the Con Man or Nazi making here? Is the problem that he cannot will the policy on which he is acting as a universal law? Or does the policy draw a distinction between groups of people that is itself an objectionable distinction to draw on some other (more substantive) grounds? **Is it really possible to provide a "formal" test for the moral acceptability of a rule or policy?**

Famously, Elizabeth Anscombe criticized Kant on precisely these grounds:

> Kant introduces the idea of "legislating for oneself," which is as absurd as if in these days, when majority votes command great respect, one were to call each reflective decision a man made a vote resulting in a majority, which as a matter of proportion is overwhelming, for it is always I-0. The concept of legislation requires superior power in the legislator. His own rigoristic convictions on the subject of lying were so intense that it never occurred to him that a lie could be relevantly described as anything but just a lie (e.g. as "a lie in such-and-such circumstances"). His rule about universalizable maxims is useless without stipulations as to what shall count as a relevant description of an action with a view to constructing a maxim about it. (GEM Anscombe, "Modern Moral Philosophy, Philosophy (Jan 1958, 33, 124, pp. 1-19; quote on p. 2)

16. A Difficult Distinction: Acting <u>from inclination</u> vs. Acting <u>from respect</u> for the moral law.

(1) You consider whether to adopt a policy: e.g. "Lie to escape embarrassment"
(2) You consider whether the policy is fair. On one reading of the FUL, this is to consider whether you could rationally expect to achieve the represented aim (escaping embarrassment) in a world in which (you have made it the case that) everyone has the policy of lying to escape embarrassment.
(3) You believe, with Kant, that this is a good characterization of what it is for a policy to be morally unacceptable.
(4) You conclude that the policy is immoral.
(5) You reject the policy or resolve not to act on it.

Let's not worry about the *validity* of the test imposed by the categorical imperative. Let's assume, that is, that you come to *know* that the policy of lying to escape embarrassment is unfair and so immoral in the manner depicted by (1)-(4).

The internalism question: Does your knowledge of the immorality of a policy of lying to escape embarrassment, when it is acquired in the manner here described, provide a sufficient explanation of your rejecting that policy? Do we need to add something to (1)-(4) to explain (5)?

> <u>Strong Kantian Internalism (SKI)</u>: The conclusion is logically or conceptually sufficient for the decision. (I.e. It is impossible that you draw the conclusion in the manner depicted without therein rejecting the maxim.)
>
> <u>Weak Kantian Internalism (WKI)</u>: The conclusion is causally sufficient for the decision and so will move you to reject the maxim unless it is overrun or swamped by contrary desires or inclinations.

Note that SKI is consistent with moral weakness. You can reject the maxim on hand by *resolving* to tell the truth even when lying is necessary to escape embarrassment, but then, when the time comes, fail to follow through on your resolution. What SKI rules out (or asserts to be impossible) is someone's *accepting* the

test imposed by the CI as a test of the fairness or morality of policies, *admitting* that the policy of lying to escape embarrassment fails this test, while retaining the policy and so habitually lying when it's necessary to escape embarrassment.

One substantive hurdle to interpreting Kant as endorsing SKI: The relation between the moral law and the will when the law determines the will is regularly described as a causal one; albeit a causal relation unlike any other, a causal relation that cannot be fruitfully investigated.

"But man as subject of moral legislation proceeding from the concept of freedom, in which he is subject to a law he gives himself (*homo noumenon*) is to be regarded as different from the sensible man endowed with reason (*specie diversus*); but different only from a practical point of view, **for there is no theory concerning the causal relation of the intelligible to the sensible**." (MM: 439fn)

Evidence in favor of interpreting Kant as endorsing SKI: "Respect for the law, which in its subjective aspect is referred to as moral feeling, is one and the same with the consciousness of one's duty" (MM: 464); "It cannot be said man has a duty of self-respect, for he must have respect for the law within himself in order to be able to conceive of duty at all" (MM: 429).

We'll see below that respect for the law is equated with certain emotional affects of the law upon us. If this is itself consciousness of one's duty, then Kant is thinking (in the above passages at least) that this emotional motive is itself an aspect of one's recognition that the policy is unfair.

17. Respect for the Moral Law

"**How a law can be of itself and immediately a determining ground of the will (though this is what is essential in all morality) is for human reason an insoluble problem and identical with how a free will is possible**. What we shall have to show a priori is, therefore, not the ground from which the moral law in itself supplies an incentive but rather what it effects (or, to put it better, must effect) in the mind insofar as it is an incentive." (2nd Critique, 5:72).

"As a law we are subject to [the moral law] without asking permission of self-love [i.e. self-interested desire]; as laid upon us by ourselves, it is a consequence of our will, and has from the first point of view an analogy with fear, and from the second with inclination" (GMM, 4:401).

(1) You know that committing adultery is necessary for your happiness. You consider, on these grounds, a policy of cheating to be happy.
(2) You determine that the maxim cannot be universalized.
(3) You realize that you *can* reject the policy on this basis alone.
(4) This realization causes you pain because it infringes upon your self-love. (negative affective result of the moral law)
(5) You are inspired by your capacity to infringe upon your self-love in this fashion. (positive affective result of the moral law)

Question: Kant says the law "must effect" these results if it is to serve as an "incentive." If Kant thinks it needn't achieve these results, and so needn't serve as an incentive, wouldn't this make him a motivational externalist?

One more piece of tantalizing evidence:

"There are such moral qualities that if one does not possess them, there can be no duty to acquire them. These are moral feeling, conscience, love of one's neighbor, and respect for oneself (self esteem). There is no obligation to have these, because they are **subjective conditions of susceptibility to the concept of duty** and are not objective conditions of morality. They are all sensitive [ästhetisch] and antecedent but natural predispositions [praedisposito] of being affected by concepts of duty. **Though it cannot be regarded as a duty to have these predispositions, yet every man has them, and it is by means of them that he can be obligated**. The consciousness of them is not of empirical origin but can only follow upon

the consciousness of a moral law—upon its effect on the mind" (MM: 399).

18. The Formula of Humanity: "Only act so that you use humanity, whether in yourself or another, as an end (in itself), and never merely as a means." What does this mean? According to Kant, humanity is one of three principle characteristics of human beings. (The others are **animality** and **personality**.)

Humanity is the ability to set goals, utilize the means to achieve these goals and organize these means and goals into a coherent whole. So to treat humanity merely as a means is to undercut one's own or someone else's ability to set goals, utilize the means towards achieving these goals or organize their goals into a coherent whole.

Task: **Distinguish treating humanity *merely* as a means from treating it as *both* a means and an end**.

Note that the injunction grounds two sets of duties: (a) duties to respect the person and (justly acquired and retained) property of another person, (b) duties to respect the personality of another by refraining from arrogance, defamation, and ridicule.

Question: How does lying to someone constitute treating her as a mere means to one's ends?

Further questions: Does denying someone political representation constitute treating her as a mere means to one's ends? Does enslaving someone (including oneself) constitute treating humanity as a mere means? What about attending to a minimal number of an employee's needs (i.e. those necessary to keep them alive and working) while doing whatever one can to "extract" the rest of her value/labor? Is a person's **consent** to your treatment of her necessary and/or sufficient for your treating her as end in herself rather than a mere means?

> Kant thought that democratic political institutions were uniquely well suited to treating people as ends rather than mere means. Isolated in Königsberg, East Prussia he developed a horribly mistaken view about the differences between races (reflected in his South Sea Islanders example described above), which lead him to believe that Africans and Native Americans didn't have the capacity for autonomy or the desire for self rule necessary for inclusion within the "Kingdom of Ends" he identifies with a truly moral community. But when he gained a better sense of the capacities of Africans and Native Americans, and came to see race as superficial (and so utterly unlike the difference between species), he changed his mind and condemned both slavery and colonialism.
>
> * For discussion of Kant's evolving view on race see the essay by Kleingeld posted to the external website.
>
> * For a scholarly account of what Kant had in mind by "treating humanity as a mere means" and the notion of moral/political that rights this involves see the essay by Pallikkathayil posted to the external website.

Task: Apply the formulation of humanity version of the categorical imperative to the four examples Kant discusses. Does the formula of humanity better capture your moral sensibilities than the principle of utility? Think here of Thomson's arguments and the distinction between obligatory action and supererogatory action on which they depend. Who has an easier time recovering this distinction, Mill or Kant?

The Kantian's Problem w Environmental Ethics: If the formula of humanity is really a sufficient a priori basis for generating all of our moral rights and duties, how can it account for our duties to other animals? More generally, how can someone who thinks that morality is grounded in intuitions of the dignity of humanity and what could serve as a universal law for beings endowed with humanity, explicate or make sense of our obligations toward other animals?

> * For discussion see the essays by Wood and O'Neil posted to the external website.

19. The Formula of Autonomy "Only act in accordance with maxims that would be enacted into law by a legislator and member [i.e. citizen] of a realm of ends."

This is the most transparently political of the three formulations of the categorical imperative and the formulation that ties it most closely with the ideals of democratic governance articulated in the U.S. Constitution. There are three aspects to this formulation:

The First Aspect of the Formula of Autonomy: The idea of moral maxims or laws that must be obeyed by those who enact them. This is captured by Kant's idea that we must regard ourselves as both legislator for and member of the realm of ends. This idea answers to the universalizability condition imposed by the first formulation of the categorical imperative.

Questions: Why is it important that legislators live under the laws they enact and that these laws do not contain exceptions for those who make them? When is it wrong to mention specific people or specific classes of people when articulating laws or policies?

The Second Aspect of the Formula of Autonomy: The idea of a realm of **ends**. This is captured by Kant's idea that the legislators must regard themselves as making and living under laws suitable for beings that are intrinsically and unconditionally valuable in virtue of possessing humanity (where, if you recall, humanity is the pre-condition for all value according to Kant).

What does it mean to say that humanity is the pre-condition for all value?

Argument
(1) The only thing with unconditional value is the good will.
(2) One cannot have a good will without having a will.
(3) One cannot have a will without possessing humanity.
Therefore,
(4) There is no unconditional value without humanity.
Corollary: (5) Everything with conditional value **receives** its value from something with unconditional value.
Therefore, (6) If there is nothing with unconditional value, then there is nothing with value.
Therefore, (7) Without humanity, nothing has value.

On Premise (3): Remember that humanity is the capacity to set goals and work towards them. Kant can plausibly argue that without this capacity one cannot decide to do things and perform actions in the "fullest sense" of these terms: i.e. actions for which we can be properly held responsible by others. That is, Kant might distinguish the kinds of actions for which people can be properly blamed or credited from reflexes, habits, and the instinctive fulfillment of drives by saying that human actions arise from **will**, while behavior of these other kinds arises from **mere desire**. (Kant would allow that some willed actions will serve desire—indeed, he suggests that for all we know *every* human action that has actually been performed has served desire. But these willed actions, which serve desire, are different from brute behavior because they don't arise **simply** from desire. They are instead mediated by reflection or reasoning of some form.

A *theoretical regress stopper* is a reason R for believing something P where one doesn't need a reason to believe R distinct from R itself.

A *practical regress stopper* is a reason R for making a decision D where one does not need a reason for valuing R or deciding on R distinct from R itself.

On premise (5): Kant regards humanity as a regress stopper. He, unlike Mill, thinks happiness cannot be a regress stopper because though it has *intrinsic value* it does not have *unconditioned value*. We only have a reason to promote or enjoy happiness insofar as it is deserved. Still, one might ask whether, say, the deserved pleasure is only "truly valuable" when one has reflected on it and in some sense "chosen" it as one among one's ends.

The Question of Existentialism: Do we discover than certain things are intrinsically valuable by (as Mill says) realizing that we want them for their own sakes? Or do we create value (as Kant says) by endorsing certain ends or choosing to think of them as components of our happiness? Mightn't there be both kinds of value? (See Kant's distinctions between fancy price and true worth below.)

Note: **The Kantian's problem w Environmental Ethic**s resurrects itself here. Aren't other animals intrinsically valuable? Doesn't plant life also have intrinsic value?

Task: In section I want you to discuss a core thought experiment in Environmental Ethics: Suppose that you were the last human alive and you had access to nuclear weapons you could use to destroy the Earth's environment. If you set the timer, immediately upon your death all the rest of the animals on Earth would perish immediately after you - you being the last human. Would it be wrong to set the timer and destroy the rest of nature? Why is this wrong, if value depends on an exercise of humanity: i.e. a choice of ends?

A further question: Suppose that you were the last animal alive and you had access to nuclear weapons you could use to destroy the Earth's environment. If you set the timer, immediately upon your death all the plants would perish immediately after you, you being the last animal. Would it be wrong to set the timer and destroy the rest of nature? Why is this wrong, if value depends on happiness, which requires consciousness or sentience of the sort plants presumably lack?

20. Unconditioned Value

This brings us back to the second formulation of the categorical imperative. Kant expands on his idea that humanity is of unconditioned value and so is always to be treated as en end. In developing the third formulation of the categorical imperative he does a better job of explaining what "unconditioned value" is supposed to mean.

Dignity vs. **Price**: two different kinds of value. Things that have a price either have a **market price** or a **fancy price**.

Market price: X has market price if and only if X has merely instrumental value or merely instrumental value and fancy price.

Fancy price: X has fancy price if and only if X has intrinsic value, but X is only intrinsically valuable because of our desires, inclinations or what Kant calls our "sensible nature." (This is that part of our nature we share with other animals.)

Dignity: X has **dignity** if and only if X has intrinsic value and X would have intrinsic value no matter how different our sensible natures happened to be.

(Something has dignity if it has intrinsic value, and it would still have had this value even if we had evolved and developed culturally in as different a fashion as can be coherently imagined.)

What follows from the fact that creatures with humanity have something that instills dignity rather than some lesser form of value? Kant thinks that this means that we cannot interfere with others or hinder their legitimate projects for the sake of happiness, even if we have their happiness in mind. Why not? Because humanity has more value than happiness—humanity has value of a completely different (and better) kind (i.e. dignity). So we can only interfere with humanity to preserve humanity.

Questions: What would Kant say about Jim the botanist? Would Jim's killing the one native to save the others amount to treating humanity in another merely as a means? Why or why not? What is the significance of Kant's distinction for the evaluation of slavery? What is its significance for the evaluation of colonialism?

The Third Aspect of the Formula of Autonomy: This is the aspect of the third version of the categorical

imperative that makes its first appearance here: it's the idea of our **legislating** or bringing morality into existence by making and living under laws. This is why the law is called "the principle of autonomy." 'Autonomy' means self-governance, self-rule, self-control, and independence.

Kant has us giving ourselves the moral law because he wants to argue that it is through being moral that we can be best realize or achieve our freedom. The capacity to act morally distinguishes us from other physical things—things that are determined to behave as they do by the laws of nature in a more or less deterministic fashion. We needn't act in a pre-determined way because we can do what we know to be right even when the past would otherwise determine that we fail to do what is right.

Questions: Is Kant right in connecting morality and free will in this manner? Are we free to act immorally—e.g. to kill ourselves or kill other people? How, if at all, does this sort of freedom differ from the kind of freedom disclosed to us when we realize we can do the right thing even when we are strongly inclined to do what we know to be wrong?

Further questions: How much do we value autonomy? Note that many people don't exercise their right to vote. (Even in a presidential election where the drama is greatest, only around 60% of eligible U.S. citizens vote.) As an anecdotal matter: friends of mine who grew up in Communist countries were overwhelmed by the consumer choices they faced after immigrating into this country. (There are too many brands of cereal: how should I choose?) Some behavioral economists study the conditions under which our choices become so numerous that we "shut down" and make an arbitrary decision rather than one grounded in an evaluation of the pros and cons of the various options. As we discussed when evaluating moral relativism, some religiously conservative people even reject the right to choose their own husband, wife or partner, trusting in their parents (or the matchmaker) to make a better choice. And is it better to have to remain faithful and keep your family together—or remain sober and pursue a career that requires this—without the "coercion" provided by the reputational judgments of others? How many of us would succeed at our ends without social pressure? (Surely, exercising autonomy is hard!)

Is happiness more important to you than autonomy? Is autonomy more important? Are they co-equal in your thinking? Are utilitarian and Kantian moralities just some among a plurality of moral perspectives worth pursuing and/or protecting?