

**NOTE ON KANT'S *GROUNDWORK*, PP. 1-40
WINTER, 2009
Dick Arneson**

PHILOSOPHY 160

SECTION I.

Kant argues in this section to the conclusion that we believe that we are bound by the categorical imperative. That is, ordinary common sense includes beliefs that imply a commitment to the categorical imperative. This is the principle that Kant asserts is the supreme principle of morality. Kant does not think this shows that we really are bound by the categorical imperative: ordinary common sense, what we all believe as a matter of course, might in fact be mistaken. Hence to show that acceptance of the categorical imperative is warranted, a different type of argument is needed, an excursion into what Kant calls "the metaphysics of morals," in section two.

Kant's starting point is the claim, which he believes common sense accepts, that nothing is good unconditionally except a good will. Some things are good conditionally: They are good given the presence of certain conditions. For me, tuna fish is not unconditionally good, it's good only on the condition that mayonnaise is also available. Kant thinks that after reflection we will all agree that the only thing that is good under all circumstances, good no matter what other conditions obtain, is a good will. Happiness, says Kant, is not good unconditionally; the thought of an evil person gaining happiness is repulsive. Bravery is not unconditionally good, because "the coolness of a scoundrel" (p. 8) renders her more abominable than she would be if she were prone to fear. Even if the good will is entirely ineffectual and unable to bring about good consequences, "like a jewel, it would still shine by itself, as something that has its full worth in itself" (p. 8).

Kant's next step is to identify a good will as a will that is disposed to act from duty. Duty is how morality presents itself to beings like us, partly rational but partly animal and so prone to act from inclinations. Morality would not appear as duty to a perfectly rational being. For beings like us, a good will is a will that acts from duty. To explain the idea of acting "from duty," Kant distinguishes four cases. Case 1 involves acting contrary to duty. Such an act clearly lacks moral worth. Case 2 involves acting in accordance with duty but not from duty. In this case, the act done is what duty requires, but the agent has no immediate inclination to do the moral act, but sees that doing the moral act will serve his purposes and does the moral act as a means to satisfy his inclinations. For example, a shopkeeper might give correct change to customers because he figures that a reputation for honesty will bring him greater profit in the long run, and he is honest in order to maximize his long-run profit. This act also has no moral worth. Case 3 is an act that is also in accordance with duty, but here the agent is moved to do the moral deed because he has an immediate inclination to do so. In Kant's example, suppose the shopkeeper loves his customers and is honest in his dealings with them because he wants to do nice things for those he loves. More surprisingly perhaps, Kant says case 3 actions also lack moral worth. Case 4 situations are those in which all of the agent's inclinations go against doing one's duty, but the agent does her duty anyway. Kant's example concerns an

individual who is very glum about his life prospects and wants to kill himself, but reflects that his duty is not to commit suicide (Kant assumes here that suicide is wrong), and does his duty against inclination. Kant says that in this case "the action first has its genuine moral worth" (p. 12).

What exactly is Kant saying here? Why is doing your duty and loving it not as morally worthy as doing your duty and hating what you are doing? Is Kant recommending that to have moral worth we should become sourpusses? What's wrong with acting from inclinations? One concern Kant has is that inclination might not reliably lead you to do your duty. An inclination to be spontaneously kind will lead me to help a roadside accident victim but will also lead me to offer a crowbar to a burglar who needs help to complete his criminal plan successfully. But suppose some inclinations would reliably lead the person who had them to act in accordance with duty. Suppose a person has an inclination, a desire, to uphold truth and justice? A second concern here is that Kant evidently thinks that what confers moral worth should be equally possible for anyone to get. It cannot be a matter of sheer luck that one person acts with moral worth and another does not. But people are fortunate or unfortunate, lucky or unlucky, in the inclinations they happen to have. One person is born with a cheerful disposition, one is not. One person is well socialized by his parents, making him cheerful, another is not. Kant thinks that what renders us morally worthy or not must be something that lies within the individual's power to control. But we can't directly control our inclinations. Kant thinks that it always lies within our power, given that we have the capacity for rational agency, to recognize what is reasonable to do and do it just from this recognition of what is reasonable. We have the capacity to be moved by reason, Kant thinks. Acting from duty is doing one's duty and doing it because one recognizes that the act is required by duty, which means for Kant, it is the act that is required by reason. Whether or not one has inclinations that push one toward doing one's duty or pull one away from doing one's duty, one always has the capacity to recognize one's duty and do the right thing just because one sees this is what duty demands. This disposition of the will is what alone has moral worth, according to Kant.

Whether an act is done from duty and whether we can know that it is done from duty are two different things. Kant thinks we can never in actual life be sure what our motivation for doing what we do really is. I may imagine I am doing what is right because it's my duty, but I may not notice a selfish inclination that is actually what induces me to act as I do. In fact, Kant thinks that for all we know, all of our acts may be caused by circumstances beyond our control. We may not have free will, in which case no act would ever have moral worth. What Kant thinks he is establishing here is that when an agent ought to conform to a moral duty, only doing one's duty in such a way that one is acting from duty has moral worth. IF we have free will, and if we can be moved to act by the recognition of what is most reasonable to do, THEN we all equally have this capacity, Kant thinks. In another of his publications Kant imagines the worst sewer rat of a person, condemned to the gallows for some terrible crime. Kant supposes the person as a rational agent will hold himself responsible for failing to do his duty. "I had a choice, I

didn't have to commit the terrible crime," Kant supposes the man will think on the morning he is to be led to the gallows. This means that whatever one's inclinations, in so far as one thinks of oneself as a rational agent, one thereby thinks one can recognize what there is most reason to choose and make that choice just because one recognizes what reason requires. There is then a strong democratic equality aspect to Kant's thinking on moral worth. Anyone who is a rational agent can equally well act from duty and thus manifest a good will.

Question: Isn't it a matter of luck, good or bad fortune, that enables one person correctly to perceive what duty requires while another misperceives the nature of duty? Should not Kant say that moral worth attaches to an agent in virtue of her sincerely trying to discover what is right and choosing to do what she takes to be right just because it is right? On this view, moral worth is manifested not just in acting from duty but in sincerely trying to act from duty, in other words, in acting conscientiously. It seems that the view that "good or bad luck should influence neither our moral judgment of a person and his actions, nor his moral assessment of himself" (quoted from Thomas Nagel, "Moral Luck"), consistently carried through, does require that when a moral duty applies to an agent and fixes what she really morally ought to do, acting in conformity with that duty is NOT required for one's choice of action to manifest moral worth. For it can be due merely to good or bad luck, good or bad fortune, that one agent discovers what is right to do and acts accordingly whereas another agent sincerely tries to discover what is right to do but innocently and nonculpably forms a mistaken judgment and acts on it. Kant does not actually take this line, however. He might hold that discovering what moral duty is and requires of an agent is always a simple matter which any rational agent who sincerely tries to discover will in fact discover. This is implausible, but might be Kant's actual view. Alternatively one might reason backwards here, and hold that Kant does not accept that a conscientious act not conforming to duty—the act of a conscientious Nazi slaughtering innocents, for example—has any moral worth, Kant must in the end not embrace the no-moral-luck position. In this connection see Thomas Nagel, "Moral Luck" (available from course web page).

One might consider whether moral motivation as it ought to be according to Kant crowds out other sorts of motivation,. Can a good Kantian agent be a good friend?

Some hold that the person who has desires that conflict with duty but overcomes those desires and behaves dutifully is less virtuous than the person whose desires are harmonious, who wants to do her duty, and who does behave virtuously. Does Kant's view of moral worth and desire conflict with this position? Yes, or so it seems. At least, if it is a matter of luck or good fortune beyond one's power to control that one is born with desires that reinforce one's resolve to do one's duty, then having good luck of this sort cannot be what makes one morally worthy. Of course, Kant can hold that one has a duty to instill pro-moral duties in oneself, to increase the likelihood that one will not act against one's duties, but nothing but doing one's duty from the recognition that what one is doing is one's duty confers moral worth on what one chooses to do.

The next step in Kant's section I argument occurs on p. 13. There Kant says that "an action done from duty has its moral worth. . .in the maxim in accordance with which it is decided upon." A maxim is a subjective principle of action, a personal policy. It is the reason that explains what we do. Often our maxims may be inexplicit, but by thinking they can be discovered. The general formula of a maxim is this: "I choose to ____ in order to ____." A maxim states what I do and what my purpose is in doing what I do. What Kant is saying then is that moral worth resides in the quality of the agent's will, the quality of the agent's intentions.

Next step, p. 13: "Duty is the necessity of an action from done respect for law." I must do my duty, one might say. The force of this "must" is that once I see where my duty lies, as a rational being I see that this is what is most reasonable to do, and as a rational agent I must do what there is most reason to do (the action supported by the strongest reasons). Notice that that acting out of respect for the law is not acting on inclination or desire.

Next step, bottom paragraph, page 13, top paragraph, page 14: Here Kant argues that acting from duty is acting for the sake of the law. In acting from duty one is not acting to satisfy any inclination. One is moved by the law as such (the thought: here lies my duty). Kant concludes: "Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law.*" This is a formulation of the categorical imperative. This is the conclusion of Section I: Acting from duty is acting in conformity with the categorical imperative.

Kant has argued that if we analyze the concept of a morally worthy action, we uncover the fundamental principle of ethics. If you ask what motivates the person doing an action that has moral worth, you discover the content of morality, a statement of what morality requires. The connection between moral motivation and the substance of moral requirements that Kant draws here is controversial. In fact it may seem that Kant at the end of section 1 is reasoning fallaciously. Kant has affirmed that an action that conforms to duty has moral worth only if the agent chooses to do that act on the ground that she perceives it to be right, morally required. This says nothing at all about what in fact renders it the case that something is morally right. For all that has been said so far, the supreme principle of morality might be (1) the norm to do whatever God commands or (2) the consequentialist norm to do whatever brings about the best possible outcome in one's circumstances or (3) anything else. This leap to a conclusion to which he is not entitled (if I am right here) at the end of section 1 may not be a big problem for Kant provided he comes up with a sound derivation of the Categorical Imperative principle later in the *Groundwork*.

Mill and in fact any consequentialist will hold that whether the motive of an action is good is one issue; whether the action itself is good is a separate and distinct issue. A nonconsequentialist might agree. One might hold that acting morally consists in respecting (not violating) anyone's rights, but hold also that the quality of motivation and assessment of it are irrelevant to the rightness or wrongness of an action, though relevant to the assessment of the character of the agent. Kant here takes an opposed position.

SECTION II.

Kant seeks to answer the question whether categorical imperatives are possible. What he means is this: Is the categorical imperative really a requirement of reason, a requirement that any person would comply with if the person was fully rational?

Hypothetical and categorical imperatives. An imperative is addressed to an agent and tells the agent that she ought to do something (must do something). Imperatives according to Kant come in two flavors, hypothetical and categorical. The hypothetical imperative in its general formulation asserts that if you will a goal, you should will the necessary means to that goal. "Willing" here is more than wishing. To will a goal is to resolve to attain it. An example of violating the hypothetical imperative would be willing a goal and taking some steps toward achieving the goal but declining to do some things that one knows must be done if the goal is to be attained. For example, I might resolve to lose ten pounds, buy a lot of cabbage and other healthy food and nibble on it hopefully, read how-to-do-it-books on weight loss, enroll in a Jenny Craig weight loss program, but continue to eat more calories per day than is consistent with any weight loss. You might say: Well, maybe you decided that eating ice cream sundaes is more important than weight loss. What's irrational about that? Answer: Nothing would be wrong with deciding that some other goal is more important than weight loss and giving up the goal of losing weight. The hypothetical imperative commands that one EITHER take the necessary steps to goals one wills OR decline to will the goal. The hypothetical imperative constrains an agent's choice of actions only if the agent wills some goal (which the agent qua rational need not will). In contrast, a categorical imperative is a requirement of reason that commands categorically, that is, regardless of whether or not one wills or does not will any particular goal (that is not itself rationally required). Kant thinks that if morality imposes any requirements of reason, these must be categorical imperatives. Either morality is bunk or we are bound to obey the categorical imperative. Kant will argue later in this section that if there are any categorical imperatives in the sense just noted at all, there is just one categorical imperative, and it must be the principle he calls the Categorical Imperative (which itself comes in different versions—the Formula of Universal law, the Formula of Humanity, the Formula of Autonomy, the Formula of the Realm of ends, etc.) To see the distinction between the claim that there are categorical imperatives and the claim that the supreme principle of morality is the Categorical Imperative, notice that one could claim that Rossian prima facie duties (for example, tell the truth, keep your promises) are categorical imperatives, or that utilitarianism or consequentialism is a categorical imperative. These claims make sense and could be true. But it would make no sense to claim that utilitarianism is the

Categorical Imperative. They are substantially different principles, with different content. If the supreme principle of morality is utilitarianism or consequentialism or allegiance to Ross's list of prima facie duties, then the supreme principle of morality is not the categorical Imperative.

A priori and a posteriori; analytic and synthetic. That the hypothetical imperative is a genuine requirement of reason is easily demonstrated, Kant believes. He holds that it is an analytic a priori truth that a rational person wills the necessary means to the goals that she wills. This jargon needs to be explained. According to Kant, a claim is known a posteriori if it is known from experience (from empirical observation), and a claim is known a priori if it is known independently of experience (empirical observation). An analytic claim is an expression in which the subject term "contains" the predicate term. Example: "A bachelor is an unmarried male" is analytic according to Kant, because the concept of bachelor already contains the idea of unmarried male. You wouldn't know the meaning of the term "bachelor" unless you knew a bachelor is an unmarried male. This claim is a priori, we can know that it is true independently of experience. I don't discover the claim is true by looking around with a lantern for bachelors who are married males and failing to find any. Once I know the meanings of the terms I know there is no point in looking around for married males who are somehow still bachelors. Nothing I observe could count as a married male bachelor I could not learn by observation that two plus two is not really four. What kind of an observation would shake one's confidence that two plus two equals four?. (If I give my cat two pieces of kibble and then a minute later two more pieces and I see that there are just two pieces of kibble there, I conclude that he ate two pieces of kibble, not that two plus two equals two.) Truths of arithmetic are known a priori, not from experience. A synthetic claim is one that is not analytic; its predicate term does not "contain" its subject term. According to Kant, that we are subject to the Categorical Imperative (that the Categorical Imperative is a requirement of reason) is a synthetic a priori claim. If it is true, its truth cannot be established just by understanding the meanings of the terms involved in the claim, and if it is knowable at all, it must be known independently of experience--observation cannot confirm that the claim is true.

The universal law formula; applying the categorical imperative; Kant's four examples.

Kant states three formulations of the Categorical Imperative principle. The first is the universal law formula: "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (p. 31). The Categorical Imperative seems to be regarded by Kant as a test for the moral permissibility of maxims. See the discussion on pages 30-32 and the discussion on pages 36-37. That is, the Categorical Imperative test is a way of determining whether the maxim you are proposing to act on is morally OK or morally forbidden. In this regard, the Categorical Imperative is on all fours with act-utilitarianism, which is a principle that in theory determines what act an agent should do in whatever circumstances she finds herself in. Kant's test is a formal test that works on proposed maxims--statements of what an individual is planning to do. Notice

that the result of the test is not to show that a type of action is permissible or not, but that a type of action done for a certain reason is right or wrong.

Kant tells us how to operate the test only by giving us examples of applying it. The examples are hard to interpret. It is not entirely clear how Kant thinks the test is supposed to work. Look and see how the examples seem to work. Here is one attempt at interpreting the categorical imperative test. (I should note that the suggestion I am going to describe does not fit Kant's actual procedure in applying the universal law formula to some of the examples he discusses. To test one's maxims by the CI test, do the following:

1. State your maxim. The maxim will follow the formula, "I choose to ____ in order to ____." If you describe your maxim inaccurately, the test procedure will not give morally correct results.
2. Universalize the maxim. Do this by supposing that everybody does the same. Everybody chooses to ____ in order to ____, where the blanks are filled in by the content of your own maxim.
3. The contradiction in conception test: Ask yourself: Can I consistently conceive that I adopt and act on this maxim I am considering and that everybody else adopts and acts on the same maxim? If the answer is No, the maxim is morally impermissible, and violates what Kant calls a perfect duty (examples of perfect duties: Truth-telling, keeping promises, refraining from murder). If the answer is Yes, the maxim and its universalization are together conceivable, this doesn't settle the issue of its permissibility. There's another test.

The example that perhaps best illustrates what Kant has in mind by this test is his discussion of the person whose proposed maxim is to borrow money by a lying promise. ("I choose to borrow money, promising to repay but planning not to, in order to get money I need.")

4. The contradiction in the will test: Ask yourself: Can I consistently will that I adopt and act on the maxim I am considering and that everybody else adopts and acts on the same maxim? If the answer is no, action on the maxim violates an imperfect duty. (Charity is an example of an imperfect duty. The duty commands one to adopt a goal of being charitable to some people sometimes but leaves one wide latitude as to when one will choose to be charitable and to whom. If an imperfect duty is a duty to others, no particular person or persons have an entitlement to be aided.)

[In lecture, various objections to the claims that Kant makes on behalf of the universal law formula were advanced. One worry is that Kant thinks of immorality as wanting to make an exception in your own case to cooperative rules you want others to continue to follow. Immorality of this sort is captured by a universalizability test of the sort Kant offers. But not all immorality may be of this sort. Some immorality may consist in affirming and willing to act on bad principles. On the basis of pure hatred of Jews, a person may embrace the principle that all Jews should be eliminated from the world. This person may be perfectly able consistently to will that everyone act on the same maxim he proposes to act on. Consider also the W.C.Fields view of suckers, "There's one born every

minute." If Field's maxim is to exploit those who are exploitable, because those who are exploitable deserve to be mistreated, again it is far from clear that a universalizability test rules out acting from this maxim as impermissible. These are alleged to be cases of false positives--maxims that pass the universalizability test and are permissible according to it but still strike many as immoral. There are also alleged to be cases that are false negatives--maxims that flunk the universalizability test and are rated as impermissible according to it but intuitively strike many people as morally permissible anyway. Examples of this type of case are innocent coordination maxims, such as the maxim to go to the beach only when it is uncrowded or the maxim to shop only at special sales when prices are exceptionally cheap. In each case the agent proposes to behave in a way that will achieve its purpose only given that most other people do not do the same, but this reliance on the behavior of others in doing what others do not do seems not to involve unfair taking advantage of other people as in false promising. But then unfairly taking advantage of others seems to involve something beyond merely acting on maxims such that one cannot consistently will that others do the same. See also Mill's objection against Kant on this issue, in Utilitarianism, chapter 1. According to Mill, Kant fails "to show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral rules of conduct."]

The humanity formula. In section II Kant introduces three formulations of the Categorical Imperative principle (universal law, humanity, and autonomy). He asserts the three formulations are equivalent in their implications for conduct and may be viewed as alternate expressions of a single underlying principle. This is a striking claim because the three formulations appear to be different in content.

Recall that a categorical imperative is a requirement of reason that binds unconditionally. It binds one independently of any aim that one might or might not adopt. One route to the Categorical Imperative is by arguing that just in virtue of being a rational agent, hence committed to acting on reasons that are general, one is committed to the categorical imperative. This gets to the universal law formula. Another route is to investigate whether there is an end that as a rational being one cannot disavow--an objective end, in Kant's jargon. If there is, it could be the basis of a categorical imperative. On pages 36-37, Kant presents an argument (which I won't try to rehearse here) to the conclusion that there is indeed such an end. If we make a choice, we have made an evaluation, and evaluating something as good requires an exercise of rational agency. If we must value what is necessary for our valuing, we must value our rational agency. Rational agency is an end we cannot then disavow. This yields a version of the Categorical Imperative, the humanity formula: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (p. 38). In other words, act always in such a way that you treat rational agency capacity, yours or that of any other person, never merely as a means.

Kant's humanity formula has seemed to many a ringing declaration that we must always treat persons with respect in virtue of their rational nature. What does this involve? I briefly indicate three construals of the meaning of the humanity formula.

1. Treating rational nature not merely as a means but also as an end is treating persons always according to principles that qua rational (if they were fully rational) they would accept. On this construal, I treat persons with respect just in case I treat them according to the principles that they would accept if they were fully rational. If someone is confused, ignorant, or malicious, but I treat him according to principles he could accept if he were rational, I do not treat him with disrespect.

Example: If the only way I can startle a small child in danger and save his life is by shattering your expensive vase, I do not treat you as a mere means if I am treating you according to principles you could if fully rational accept. The question then is whether "accept minor inconvenience to save a life" is a rational principle. Kant might hold that this issue is settled by seeing whether the principle is universalizable according to the universal law test. On this construal, it is easy to see why Kant believes that the two principles are aspects of the same basic idea.

2. Treating rational nature not merely as a means is refraining from treating persons in ways to which they could not possibly give consent. This line is suggested by Kant's comment, applying the humanity formula to the false promising example, that "he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving toward him, and so himself contain the end of this action" (p. 38). Some interpreters suggest that when judged by this standard, deception and coercion will be revealed to be always impermissible. When I deceive someone, I treat that person in a way to which she could not possibly consent, because she does not know what is occurring, and so it would be logically impossible for her to consent to it. In a similar way, so it is claimed, if I am coerced, I cannot voluntarily consent to what is occurring.

Quibble: Does the test on this construal really rule out deception and coercion? It seems I can possibly consent to being deceived, as when I tell my doctor to keep me from learning the truth for as long as possible if he diagnoses me with cancer. I think I can also imagine a case of consenting now to be deceived now and then actually being deceived. Similar examples suggest one can consent to being coerced. This quibble does not challenge the suggested test, but only disputes what would be the results of applying it. Objection to the test: The could-not-possibly-consent test seems too weak. I could possibly consent to being assaulted even when I do not actually do so and have no good reason to consent. The mere fact that it is possible that I could consent seems insufficient to render the act of assaulting me morally permissible. (For a contrary view, see the Christine Korsgaard essay, "Kant on the Right to Lie.")

3. The third construal is offered by Thomas Hill. According to Hill, treating rational nature not merely as a means is treating people always in ways that respect the unconditional and incomparable dignity of their rational nature.

Interpreters who follow this third line take their cue from Kant's distinction between dignity and price. Kant distinguishes what has price from what has dignity. What has dignity is above price—is priceless, as we might say. What has price can reasonably be traded for any other good that commands a price, as a certain quantity of hay might be traded for a smaller pile of diamonds. What has dignity is beyond price. Kant says that what has dignity has “an unconditional and incomparable worth.” On this basis Kant is interpreted as maintaining both that what has dignity is incomparably more valuable than anything with price, so that one should not accept the tiniest loss in dignity value in exchange for the greatest gain in what has price, and also that what has dignity has a special nonaggregative value, so one cannot reasonably quantify dignity values.

Consider the maxim of committing suicide to obtain relief from pain. According to the incomparable-dignity construal of the humanity formula, it forbids suicide done for this reason, which amounts to giving up rational nature (continued life in which rational agency capacity can be exercised) for greater pleasure or lesser displeasure. But this is in effect to exchange what has dignity for what has merely price value. Mutilating oneself in a way that lessens one’s rational agency capacity to gain some conditional value is also forbidden by the humanity formula in the same way. Hill gives other examples of the application of the humanity formula. He writes, “since the exercise of rationality is something to be cherished, in trying to influence others one should appeal to their reason rather than try to manipulate them by nonrational techniques.”

Comment: The idea that the categorical imperative commands each of us always to respect persons in virtue of their possession of rational capacity and never to act in ways that express wrongfully low valuation of rational agency is attractive and appealing. But does rational agency have a value incomparably greater than the values labelled goods that have only price? An alternative view is that rational agency has large but not incomparably greater value. Consider a choice to engage in a risky sport for fun even though the activity increases one’s chances of suffering an injury that destroys or damages one’s rational agency capacity. If this is reasonable, then having rational agency capacity is not incomparably more valuable than having fun. Or consider going across town, increasing one’s risk of premature loss of rational agency capacity, in order to get tastier bread than one can obtain in one’s neighborhood. Or consider becoming drunk at a wedding, to be festive, in safe surroundings. One sacrifices some amount of time in the possession of full rational capacity in exchange for conviviality. If these activities can be permissible, then the humanity formula as interpreted by Hill is delivering implausible verdicts. Consider also the idea that dignity value cannot be quantified. Suppose one has to choose between preventing one person from being injured in a way that damages rational agency capacity and preventing a thousand persons from suffering the same injury. If dignity values cannot be aggregated, it seems one cannot say 1000 persons

losing rational agency is worse than one. To some, the notion of an incomparable and nonquantifiable value, a value to be respected not promoted, sounds fishy.

Further comment: Some commentators have suggested that if we interpret Kant's Categorical Imperative principle as a test for deciding what maxims it is morally OK to act on, we run into problems, some of which have been canvassed in these notes. If we interpret Kant's Categorical Imperative principle as giving us a procedure to determine what our moral duties are, e.g. that lying is wrong or that we morally ought not to borrow money on the basis of a false promise to repay, we run into problems. These problems are serious. Perhaps we should think of the point of the CI procedure in a different way. One proposal is that the supreme principle of morality is that we should act only in ways that express proper respect for the unconditional and incomparable value of rational agency capacity, but what this requires in practice cannot be set forth in any determinate principles. We just have to figure out in each particular context what respect for humanity (rational nature) requires in that context. The norm of respecting rational nature then is functioning somewhat like Ross's list of prima facie duties—there is no formula according to Ross that specifies what, given these duties, we ought to do in any particular circumstances. {{Don't worry about the references to the author Ross—he is not a course author.}} In some ways this way of reading the CI is close to construal #1 of the humanity formula plus the additional claim that no other formula of the CI supplies determinate content as to what we morally ought to do. On this latter view, the CI is a way of conceiving of morality and not a substantive principle that specifies or determines what we ought to do. On the "express respect for humanity" reading, in contrast, the CI is substantive but interpreting its requirements always requires additional evaluation case by case.

The overall argument of the three sections of the *Groundwork* book is as follows.

1. In the first section Kant argues that we all assume that we are subject to the categorical imperative, that is, that ordinary moral consciousness presupposes the validity of the categorical imperative. To get to this conclusion Kant argues that the only thing in the world that is good without qualification is a good will. The next step is to argue for the claim that a good will is a will that acts from duty and not merely in accordance with duty. An act done from duty is motivated by respect for duty. When (if) I manifest a good will in action, my respect for duty determines that I do the dutiful action, regardless of whether or not my desires incline me to do the act that satisfies the duty. For example, if I am honest because I like the people I am dealing with or because I think that I will make more profit in the long run if I am honest or because I am afraid that if I am dishonest I will get caught and punished or even because I happen to like being an honest person, my act lacks moral worth. Only if I am honest because I recognize that honesty is my duty does my act have moral worth according to Kant. Roughly, Kant's idea is that it is not under my control what kind of desires I happen to have, but I am morally responsible only for what lies within my control, so I am not responsible for the

quality of my desires, whether they are nice or nasty. Therefore, I can take no credit for my desires if they are nice and I cannot be blamed if they are nasty. But on the assumption that one can always choose to do an act because one recognizes it is required by reason, then each of us always has it within our power to conform to duty from recognition of duty. If my will is good or not good, this must be so on account of some feature of my will that I can be responsible for, that lies within my control, and this cannot be a matter of having nice or nasty desires, but must rather be decided by whether or not I act from duty. Kant finally argues to the conclusion that an act that is done from duty must be an act that is done for the sake of the law and this must be an act that aims at conformity to the categorical imperative. The idea of conforming to law as such gives us the idea I ought never to act except in such a way that I can also will that my maxim should become a universal law. This is a formulation of the Categorical Imperative.

2. The second section distinguishes hypothetical from categorical imperatives, introduces three formulations of the categorical imperative (the formulas of universal law, humanity as an end in itself, and autonomy), shows how to apply the categorical imperative as a test for the permissibility of proposed maxims, uses the idea of autonomy to connect the idea of the categorical imperative to the ideal of a kingdom of ends, and links the categorical imperative idea to human dignity and worth.

The second section also argues that if categorical imperatives are possible at all, if there are any categorical requirements of reason binding us, then any such imperatives will take the form of the rule that Kant calls the Categorical Imperative. This argument occurs in the 30th paragraph of section 2, just before Kant discusses his four examples of applying the categorical imperative.

Just a word about Kant's view on autonomy and rationality as developed in the second section: Details aside, Kant's idea is that morality is not an external command that limits our freedom like the command of a king or a political ruler. The commands of morality are dictates of rationality, and as free beings it is our nature to act rationally. We best express our nature as free rational beings by acting in conformity with the moral law. The commands of morality are in this respect just like the commands of mathematical logic. Logic is not best viewed as a limitation on our ability to think as we please. In following a mathematical proof correctly and accepting its conclusion, we are "constrained" only by requirements of reason that as rational beings we will to accept. Each person who correctly follows a proof legislates the conclusion for herself, and according to Kant morality is self-legislation in the same sense. When we follow the commands of morality we are following commands that as rational beings we legislate for ourselves, so in being moral, we are autonomous, self-governing. Just as all rational beings working through mathematical proofs correctly will reach the same conclusions, all rational beings legislating rules for themselves will be legislating commands that turn out to be the same for everybody. In this way we get from autonomy to the idea of a kingdom of ends, a community of self-legislating rational agents whose wills are in perfect harmony insofar as

they are willing the same rational moral laws. See the tail end of section two on the link between autonomy, the categorical imperative, and the ideal of a kingdom of ends.

3. The third section takes up the issue of free will. The problem Kant wrestles with in the third section is how to make room for the ideas of human choice, reason, and deliberation in a world that we believe to be a world where every event is caused. Section II has argued in effect for the claim that if we have free will, if reason can motivate us, then we are bound by the Categorical Imperative. It would be nice if Kant could prove that we do have free will, reason can motivate us, hence we are bound by the Categorical Imperative. But Kant thinks that you cannot establish the metaphysical claim that we have free will. Moreover, when we observe the world and try to describe it, we are committed to the claim that every event has a cause, which seems to preclude free will. Nonetheless Kant tries to show (a) it is not inconceivable that we have free will and (b) whenever we deliberate and choose actions we must presuppose that we have free will, that reason can motivate us.

The *Groundwork* argument in a nutshell:

Section 1: Our ordinary moral beliefs commit us to believing we are bound by the Categorical Imperative.

Section 2: If reason can move us to action, we are bound by the Categorical Imperative.

Section 3. (It would be nice if Kant could prove in the third section that reason can move us to action. But Kant thinks no such thing can be proved. Instead Kant tries to show the following)

1. It is possible that reason can move us to action
and
2. Whenever we deliberate and choose what to do, we must assume that reason can move us to action.