

Kant's Formulations of the Categorical Imperative

- (1) Kant's first formulation of the CI states that you are to “act only in accordance with that maxim through which you can at the same time will that it become a universal law.”

(G 4:421) O'Neill (1975, 1989) and Rawls (1989, 1999), among others, take this formulation in effect to summarize a decision procedure for moral reasoning, and I will follow them: First, formulate a maxim that enshrines your reason for acting as you propose. Second, recast that maxim as a universal law of nature governing all rational agents, and so as holding that all must, by natural law, act as you yourself propose to act in these circumstances. Third, consider whether your maxim is even conceivable in a world governed by this law of nature. If it is, then, fourth, ask yourself whether you would, or could, rationally *will* to act on your maxim in such a world. If you could, then your action is morally permissible.

- (2) The Humanity formulation of the CI. This formulation states that we should never act in such a way that we treat Humanity, whether in ourselves or in others, as a means only but always as an end in itself. This is often seen as introducing the idea of “respect” for persons, for whatever it is that is essential to our Humanity. Kant was clearly right that this and the other formulations bring the CI ‘closer to intuition’ than the Universal Law formula. Intuitively, there seems something wrong with treating human beings as mere instruments with no value beyond this. But this very intuitiveness can also invite misunderstandings.

First, the Humanity formula does not rule out using people as means to our ends. Clearly this would be an absurd demand, since we do this all the time. Indeed, it is hard to imagine any life that is recognizably human without the use of others in pursuit of our goals. The food we eat, the clothes we wear, the chairs we sit on and the computers we type at are gotten only by way of talents and abilities that have been developed through the exercise of the wills of many people. What the Humanity formula rules out is engaging in this pervasive use of Humanity in such a way that we treat it as a *mere* means to our ends. Thus, the difference between a horse and a taxi driver is not that we may use one but not the other as a means of transportation. Unlike a horse, the taxi driver's Humanity must at the same time be treated as an end in itself.

Second, it is not human beings *per se* but the ‘Humanity’ in human beings that we must treat as an end in itself. Our ‘Humanity’ is that collection of features that make us distinctively human, and these include capacities to engage in self-directed rational behavior and to adopt and pursue our own ends, and any other capacities necessarily connected with these. Thus, supposing that the taxi driver has freely exercised his rational capacities in pursuing his line of work, we make permissible use of these capacities as a

means when we behave in a way that he could, when exercising his rational capacities, consent to — for instance, by paying an agreed on price.

Third, the idea of an end has three senses for Kant, two positive senses and a negative sense. An end in the first positive sense is a thing we will to produce or bring about in the world. For instance, if losing weight is my end, then losing weight is something I aim to produce. An end in this sense guides my actions in that once I will to produce something, I then deliberate about means of producing it. Humanity is not an ‘end’ in this sense, though even in this case, the end “lays down a law” for me. Once I have adopted an end in this sense, it dictates that I do something: I will act in ways that will bring about that end.

An end in the negative sense lays down a law for me as well, and so guides action, but in a different way. Korsgaard (1996) offers self-preservation as an example of an end in a negative sense: We do not try to produce our self-preservation. Rather, the end of self-preservation prevents us from engaging in certain kinds of activities, for instance, picking fights with mobsters, and so on. That is, as an end, it is something I do not act *against* in pursuing my positive ends, rather than something I produce.

Humanity is in the first instance an end in this negative sense: It is something that *limits* what I may do in pursuit of my other ends, similar to the way that my end of self-preservation limits what I may do in pursuit of other ends. Insofar as it *limits* my actions, it is a source of *perfect* duties. Now self-preservation is a *subjective* end, while Humanity is an *objective* end. Self-preservation is subjective in that it is not an end that every rational being must have. We do place more importance on it than most of our other positive ends. Because self-preservation is more important to me than excitement, I am not a base-jumper, and so self-preservation puts a limit on my behavior. But I could make self-preservation less important if I wish, and perhaps put excitement in its place so that it, and not self-preservation, limits pursuit of my other ends. Humanity is an *objective* end, because it is an end that every rational being must have insofar as she is rational. Hence, it limits what I am *morally permitted* to do when I pursue my positive *and* subjective negative ends.

The Humanity in myself and others is also a *positive* end, though not in the first positive sense above, as something to be produced by my actions. Rather, it is something to realize, cultivate or further by my actions. Becoming a philosopher, pianist or novelist might be my end in this sense. When my end is becoming a pianist, my actions do not, or at least not simply, produce something, being a pianist, but constitute or realize the activity of being a pianist. Insofar as the Humanity in ourselves must be treated as an end

in itself in this second *positive* sense, it must be cultivated, developed or fully actualized. Hence, the Humanity in oneself is the source of a duty to develop one's talents or to 'perfect' one's Humanity. When one's makes one's own Humanity one's end, one pursues its development, much as when one makes becoming a pianist one's end, one pursues the development of piano playing. And insofar as Humanity is a positive end in others, I must take account of their ends in my own plans. In so doing, I further the Humanity in others, by helping further the projects and ends the adoption and pursuit of which constitutes that Humanity. It is this sense of Humanity as an end-in-itself on which Kant's arguments for imperfect duties rely.

Finally, Kant's formula requires "respect" for the Humanity in persons. Proper regard for something with absolute value or worth requires respect for it. But this can invite misunderstandings. One way in which we respect persons, termed "appraisal respect" by Stephen Darwall (1977), is clearly not consistent with the Humanity formula: I may respect you as a rebounder but not a scorer, or as a researcher but not as a teacher. When I respect you in this way, I am positively appraising you in light of some achievement or virtue you possess relative to some standard of success. If this were the sort of respect Kant is counseling, then clearly it may vary from person to person and is surely not what treating something as an end-in-itself requires. For instance, it does not seem to prevent me from regarding rationality as an achievement and respecting one person as a rational agent in this sense, but not another. And Kant is not telling us to ignore differences, to pretend that we are blind to them on mindless egalitarian grounds. However, a distinct way in which we respect persons, referred to as "recognition respect" by Darwall, better captures Kant's position: I may respect you because you are a student, a Dean, a doctor or a mother. In such cases, cases of respecting you because of who or what you are, I am giving the proper regard to a certain fact about you, your being a Dean for instance. This sort of respect, unlike appraisal respect, is not a matter of degree based on your having measured up to some standard of assessment. Respect for the Humanity in persons is more like Darwall's recognition respect. We are to respect human beings simply because they are persons and this requires a certain sort of regard. We are not called on to respect them insofar as they have met some standard of evaluation appropriate to persons.

- (3) The third formulation of the CI is "the Idea of the will of every rational being as *a will that legislates universal law*." (4:432) Although Kant does not state this as an imperative as he does in the other formulations, it is easy enough to put it in that form: Act so that through your maxims you could be a legislator of universal laws. This sounds very similar to the first formulation. However, in this case we focus on our status as universal law *givers* rather than universal law *followers*. This is of course the source of the very dignity of Humanity Kant speaks of in the second formulation. A rational will that is merely bound by universal laws

could act accordingly from natural and non-moral motives, such as self-interest. But in order to be a legislator of universal laws, such contingent motives, motives that rational agents such ourselves may or may not have, must be set aside. Hence, we are required according to this formulation to conform our behavior to principles that express this autonomy of the rational will — its status as a source of the very universal laws that obligate it. As with the Humanity formula, this new formulation of the CI does not change the outcome, since each is supposed to formulate the very same moral law, and in some sense “unite” the other formulations within it. Kant takes each formulation that succeeds the first in its own way to bring the moral law “closer to feeling”. The autonomy formula presumably does this by putting on display the source of our dignity and worth, our status as free rational agents who are the source of the authority behind the very moral laws that bind us.

The Kingdom of Ends Formula as a variant of formula 3: This formulation has gained favor among Kantians in recent years (see Rawls, 1972; Hill, 1992). Many see in it as introducing more of a social dimension to Kantian morality. Kant states that the above concept of every rational will as a will that must regard itself as enacting laws binding all rational wills is closely connected to another concept, that of a “systematic union of different rational beings under common laws”, or a “Kingdom of Ends”. (4:433) The formulation of the CI states that we must “act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends” (4:439). It combines the others in that (i) it requires that we conform our actions to the maxims of a legislator of laws (ii) that this lawgiver lays down universal laws, binding all rational wills including our own, and (iii) that those laws are of ‘a merely possible kingdom’ each of whose members equally possesses this status as legislator of universal laws, and hence must be treated always as an end in itself. The intuitive idea behind this formulation is that our fundamental moral obligation is to act only on principles which could earn acceptance by a community of fully rational agents each of whom have an equal share in legislating these principles for their community.

Robert Johnson, “Kant’s Moral Philosophy,” *The Encyclopedia of Philosophy*.