

Creating the Kingdom of Ends

Christine M. Korsgaard

Harvard University



CAMBRIDGE
UNIVERSITY PRESS

4 Kant's Formula of Humanity

I INTRODUCTION

The Second Section of the *Groundwork of the Metaphysics of Morals* contains three arguments that have the form: if there were a categorical imperative, this is what it would have to be like.¹ Each of these arguments leads to a new set of terms in which the categorical imperative can be formulated. In summarizing these arguments, Kant tells us that universality gives us the form of the moral law; rational nature or humanity as an end in itself gives us the material of the law; and autonomous legislation in a kingdom of ends represents a complete determination of maxims and a totality of ends. The Formula of the Universal Law is to be used in actual decision making, we are told; the other two, which bring the moral law "closer to intuition" and "nearer to feeling" can be used to "gain a hearing for the moral law" (G 436).

Attention to these remarks about the relations among the three formulas has perhaps obscured the fact that the three formulas represent a progression in the argument that leads from "popular moral philosophy" into "the metaphysics of morals." I think that it is sometimes supposed that Kant's claim that the categorical imperative is a principle of reason rests squarely on the Formula of Universal Law – i.e., on that formula's "formality." The claims of the other two formulas to be rational principles are then taken to be based upon their presumed equivalence to the Formula of Universal Law. Those who make such a supposition err not only by ignoring the fact that the Categorical Imperative is not "deduced" in the *Groundwork* until the Third Section, but also by ignoring the fact that each formulation is intended to represent some characteristic feature of

rational principles. In particular, "humanity" is argued to be the appropriate material for a rational principle, just as universality is its appropriate form. Furthermore, the addition of each new feature represents a step further into the metaphysics of morals, with the idea of autonomy providing the stepping-stone that will make the transition to a critique of practical reason possible. In this paper, I am concerned with the argument for the Formula of Humanity. Specifically, I want to consider what characteristic feature of "humanity" as Kant thinks of it makes humanity the appropriate material for a principle of practical reason.

At the end of the discussion of the Formula of Universal Law and the examples of its application, Kant claims to have shown that duty must be expressed in categorical imperatives and to have "clearly exhibited the content of the categorical imperative," if there is one (G 425). Having established that a categorical imperative would say that we should act only on such maxims as we can will to be universal laws, Kant raises a new question: "Is it a necessary law for all rational beings that they should always judge their actions by such maxims as they themselves could will to serve as universal laws?" (G 426). To answer this question we must discover an *a priori* connection between the law and the will of a rational being, and in order to discover this connection we will be driven into metaphysics (G 426). That is, we have to investigate the possibility of "reason thus determining conduct" (G 427). This investigation is a motivational one.

Kant proceeds to tell us that what "serves the will as the objective ground of its self-determination" is an end. Ends may be either objective or subjective, depending on whether they are determined by reason or not. A formal principle is one that disregards all subjective ends; not one that disregards ends altogether (G 427). It is Kant's view throughout his moral philosophy that every action "contains" an end; there is no action done without some end in view. The difference between morally worthy action and morally indifferent action is that in the first case the end is adopted because it is dictated by reason and in the second case the end is adopted in response to an inclination for it. For instance, in the *Groundwork I* example of the comparison between morally worthy beneficence and morally indifferent beneficence, the difference is found to rest in the different grounds on which each of the two men have adopted the welfare of others as his end. It is

a mistake to suppose that Kant is contrasting a man who helps others as a mere means to his own pleasure with a man who does so from duty. Kant says explicitly that the man of sympathetic temperament is "without any motive of vanity or selfishness" (G 398). Each of these characters genuinely has the welfare of others as his end – that is, each values it for its own sake.² The difference is that the morally worthy man has adopted this end because it is a duty to have such an end. Of course, in the case of action that promotes the obligatory ends it is obvious that the morally motivated person has an end in view. But is there an end in view in the sort of moral action that is required as strict duty, and does not involve one of the two obligatory ends? Kant's answer is that there is – the end in view is humanity. The difference between the person who acts merely in accordance with duty and the person who acts *from* duty is described, in the *Metaphysical Principles of Virtue*, in terms of this end:

Although the conformity of actions to right (i.e., being an upright man) is nothing meritorious, yet the conformity to right of the maxim of such actions regarded as duties, i.e., *respect* for right, is meritorious. For by this latter conformity a man makes the right of humanity or of men his end. (MPV 390)

It is important here to keep in mind that there are two different roles an end can play in the determination of conduct; it can serve as a purpose pursued, or it can play a negative role and serve as something one must not act against. To take an ordinary example: we do not often get into situations where self-preservation serves as a positive incentive to any action, but it might quite frequently keep us from taking undue risks in the pursuit of our other ends: without much thought and in an everyday way, one might, under the influence of this end, avoid a dangerous area or going out at night. Kant thinks that the end of humanity functions in this negative way: "the end here is not conceived as one to be effected but as an independent end, and thus merely negatively. It is that which must never be acted against . . ." (G 437). In the *Metaphysical Principles of Virtue* Kant explains the constraining role of humanity as an end this way:

The doctrine of right had to do merely with the formal condition of external freedom. . . . Ethics, on the other hand, supplies in addition a matter (an object of free choice), namely, an *end* of pure reason which is at the same time represented as an objectively necessary end, i.e., as a duty for

man. For since sensible inclinations may misdirect us to ends (the matter of choice) which may be contrary to duty, legislative reason cannot guard against their influence other than, in turn, by means of an opposing moral end, which therefore must be given a priori independently of inclination. (MPV 380–81)

The role that Kant here assigns to this end stands in a specific relation to *human* reason, for it is human reason that has the obstacles provided by sensuous inclination to overcome. The sensuous inclinations present themselves falsely as sufficient reasons for action, because of a tendency in human nature which is described in *Religion Within the Limits of Reason Alone* as a "propensity to evil" (R 34–36), and in the *Critique of Practical Reason* as "self-conceit" (C2 73–74). These obstacles to goodness are controlled by making humanity an unconditional end which must never be acted against; and it is in this that human virtue consists (C2 84–89). There is no contradiction between this view and the many passages where Kant insists that morality needs no end as an incentive. Having humanity as an end is not an incentive for adopting the moral law; rather, the moral law commands that humanity be treated as an end. Although the role of this end in checking the inclinations is specific to human reason, it has a metaphysical point. Human *freedom* is realized in the adoption of humanity as an end in itself, for the one thing that no one can be compelled to do by another is to adopt a particular end (MPV 381), and this end, freely adopted, checks the power of the inclinations. In the *Groundwork*, the argument for the Formula of Humanity is preceded by warnings that the motive we are seeking not only must not be an empirical or subjective feeling or propensity, but also must not be derived from "a particular tendency of the human reason which might not hold necessarily for the will of every rational being" (G 425). Therefore, it must turn out that freely acting *from* duty and adopting humanity as one's unconditioned end are one and the same thing.

The argument for the Formula of Humanity as an End in Itself has two parts: Kant first argues that there must be an unconditional end; second, that the end must be humanity. The first part of this argument is simple; one can make it in either direction. If there is a necessary end, then there is a categorical imperative, for this end would be "a ground of definite laws" (G 428). If there is a categorical

imperative, then there must be some necessary end or ends, for if there is a categorical imperative there are necessary actions, and every action contains an end (MPV 385). In the *Metaphysical Principles of Virtue* this consideration also serves as the basis for what Kant calls a deduction of the duties of virtue from pure practical reason:

For practical reason to be indifferent to ends, i.e., to take no interest in them, would be a contradiction; for then it would not determine the maxims of actions (and the actions always contain an end) and, consequently, would not be practical reason. (MPV 395)

This shows that if there is a categorical imperative, it must have as its material a necessary end or ends. This end, Kant argues, must be "humanity."

II HUMANITY

Before looking at Kant's argument that the necessary end must be humanity, I want to review the available evidence about what Kant means by that term. The argument itself will show us what Kant has in mind, but preliminary evidence will pave the way for the argument. In the *Groundwork*, Kant interchanges the terms "humanity" and "rational nature." And he tells us that

Rational nature is distinguished from others in that it proposes an end to itself. (G 437)

The fullest statement of his notion of humanity is found in the *Metaphysical Principles of Virtue*:

The capacity to propose an end to oneself is the characteristic of humanity (as distinguished from animality). The rational will is therefore bound up with the end of the humanity in our own person, as is also, consequently, the duty to deserve well of humanity by means of culture in general, and to acquire or promote the capacity of carrying out all sorts of ends, as far as this capacity is to be found in man. (MPV 392)

In clarifying the idea of cultivation he has referred to "humanity, by which he alone is capable of setting himself ends" (MPV 387).

As these passages indicate, Kant takes the characteristic feature of humanity, or rational nature, to be the capacity for setting an end. Ends are "set" by practical reason; human beings are distinguished

from animals by the fact that practical reason rather than instinct is the determinant of our actions. An end is an object of free choice (MPV 384). A rational being, as possessor of a will, acts on maxims of his or her own choosing; but every maxim contains an end, and in choosing the maxim one also chooses an end. In the case of morally worthy actions, the end is chosen because of the necessity of the principle embodied in the maxim; but it is not only the morally obligatory ends that are freely chosen under the agency of practical reason. All maxims are freely adopted and so all ends are so chosen.

While it will be obvious that Kant thinks that the obligatory ends are objects of reason, the idea that all human ends are in some sense set by reason requires a little more explanation. It might seem to some that it is more natural to say of ends other than the obligatory ends that they are "set" by inclination or "passion," and that reason's only role with respect to these is that of determining the means by which they are to be realized. To see that this is not Kant's view is important for an understanding of the Formula of Humanity: it is the capacity for the rational determination of ends in general, not just the capacity for adopting morally obligatory ends, that the Formula of Humanity orders us to cherish unconditionally. I would therefore like to cite some additional evidence for this point.

First, there are the remarks about reason and happiness in the *Groundwork*. In the teleological argument concerning the purpose of practical reason in Section One, Kant argues that if happiness were nature's end for us, instinct would have been a better guide; nature would have allowed us theoretical reason with which to contemplate out happy state, but

would have taken care that reason did not break forth into practical use nor have the presumption, with its weak insight, to think out for itself the plan of happiness and the means of attaining it. Nature would have taken over not only the choice of ends but also that of the means, and with wise foresight she would have entrusted both to instinct alone. (G 395)

This remark could be read either as suggesting that nature has the choice of ends but would also have taken over the means as well, or as suggesting that in us nature has relinquished the control of both ends and means (and would not have done so if happiness had been her purpose). The remark that practical reason tries to think out the plan of happiness as well as the means to it, however, suggests the

latter reading. The latter remarks to the effect that happiness is an indefinite "ideal of the imagination" support this reading (G 418). If happiness were some plain and obvious thing – for example pleasure as Bentham thought of it – the problem of determining the means to it could be no more serious than the problem of determining the means to anything whatever. The difficulty Kant points to is that in constructing the imperative of prudence reason must specify the end before it can determine the means; but there is no possible rule for specifying "the plan of happiness."³

The second and best piece of evidence for the role of reason in the selection of ends in general comes from the essay *Conjectural Beginning of Human History*. In this essay, Kant uses Genesis as the basis for a speculative reconstruction of the steps taken by humanity in its transformation from a creature governed by instinct to a rational being. The first object of free choice is the apple and Kant explains how it comes about. Humans are guided by instinct through the sense of smell and taste to their natural food. But by means of comparison they notice that other foods are visually similar to the things they eat. This operation of comparison is assigned to reason, and it leads to new desires; not only desires that go beyond instinct, but desires that are positively contrary to it (CBHH 111). The result of this event is described by Kant:

The original occasion for deserting natural instinct may have been trifling. But this was man's first attempt to become conscious of his reason as a power which can extend itself beyond the limits to which all animals are confined . . . this was a sufficient occasion for reason to do violence to the voice of nature (3:1) and, its protest notwithstanding, to make the first attempt at a free choice . . . He discovered in himself a power of choosing for himself a way of life, of not being bound without alternative to a single way, like the animals . . . He stood, as it were, at the brink of an abyss. Until that moment instinct had directed him toward specific objects of desire. But from these now opened up an infinity of such objects, and he did not yet know how to choose between them. (CBHH 111–12)

Kant goes on to trace further steps by which the powers of reason are developed. Reason not only directs the human being to objects around it for which there is no instinctual desire, but leads to the development of specifically human desires, such as love and the taste for beauty, and later concern for the future (CBHH 112–15). Morality comes only at the end of this development. But the develop-

ment represents a logical or rational completion as well as a genesis. The possession of practical reason, through such operations as comparison and foresight, directs our desires to an ever-increasing range of objects, but so far it does not teach us how to choose among them. Reason makes it possible to set new ends, but its guidance at this stage is only partial. This is a crucial point, for it is because of this fact that these ends are still "subjective" and not yet "objective" ends. Reason plays a role in determining our interest in them, but they are not dictated by reason. Human reason, by directing us to "an infinity" of new possible objects of desire without determining more definitely which are worthy of choice, sets up a problem. The *Groundwork* argument suggests that the idea of making a plan for happiness will not solve this problem. Rather, it is only through the development of morality that reason can give us *complete* guidance in choosing ends. In any case, there can be no question that in this essay Kant thinks of all human ends as being partially "set" by the operations of reason. They may be objects of desire or inclination, but it is reason that is responsible for the unique human characteristic of having non-instinctual desires.

The third piece of evidence for Kant's views about the specific nature of "humanity" comes from *Religion Within the Limits of Reason Alone*. In a discussion of the question whether human beings are good or evil by nature, Kant describes an "Original Predisposition to Good in Human Nature," divided into three parts: predispositions to animality, humanity, and personality (R 26–27). Of these three predispositions the first is associated with the instinctual desires and the last with respect for the moral law as sufficient incentive of the will. In between them comes the predisposition to humanity which can be brought under the general title of self-love which is physical and yet *compares* (for which reason is required); that is to say, we judge ourselves happy or unhappy only by making comparison with others. (R 27)

Kant adds that

The first requires no reason, the second is based on practical reason, but a reason thereby subservient to other incentives, while the third alone is rooted in reason which is practical of itself, that is, reason which dictates laws unconditionally. (R 28)

It might be possible to read "practical reason . . . subservient to other incentives" as referring to a "hypothetical" use of practical

reason – that is, a discovery of means – except that the role actually assigned to practical reason here is not the discovery of means but “comparison.” This is the role of reason in *Conjectural History* as well, although in the *Religion* the comparison is not among possible objects of desire but a comparison of one’s own lot with another’s. Again, however, the result must be the acquisition of new, specifically human ends, for Kant claims that “nature, indeed, wanted to use the idea of such rivalry . . . only as a spur to culture” [R 27].

Throughout the historical writings, culture represents the development towards the perfect freedom or rule of reason that will only be achieved by morality.

When Kant says that the characteristic of humanity is the power to set an end, then, he is not merely referring to personality, which would encompass the power to adopt an end for moral or sufficient reasons. Rather, he is referring to a more general capacity for choosing, desiring, or valuing ends; ends different from the ones that instinct lays down for us, and to which our interest is directed by the operations of reason. At the same time, of course, it is important to emphasize that this capacity is only completed and perfected when our ends are fully determined by reason, and this occurs only when we respond to moral incentives. Humanity, completed and perfected, becomes personality, so that in treating the first as an end in itself we will inevitably be led to realize the second. Thus, in the *Critique of Practical Reason*, humanity in one’s own person and personality are spoken of as if they were the same thing [C2 87]. But the distinctive feature of humanity, *as such*, is simply the capacity to take a rational interest in something: to decide, under the influence of reason, that something is desirable, that it is worthy of pursuit or realization, that it is to be deemed important or valuable, not because it contributes to survival or instinctual satisfaction, but as an end – for its own sake. It is this capacity that the Formula of Humanity commands us never to treat as a mere means, but always as an end in itself.

III THE BASIS OF THE ARGUMENT

But suppose that there were something the existence of which in itself had absolute worth, something which, as an end in itself, could be a ground

of definite laws. In it and only in it could lie the ground of a possible categorical imperative, i.e., of a practical law [G 428].

With these words, Kant, in the *Groundwork*, establishes the connection between the existence of a categorical imperative and the existence of an unconditionally valuable end. Immediately after, he asserts, and then argues, that this end must be “man and, in general, every rational being.” In the next section, I want to reconstruct that argument in order to show why humanity must be this unconditional end and the material of a rational principle. In this section I want to say something about the theory of rational action upon which that argument is based.

In the discussion of good and evil in the *Critique of Practical Reason* (57–71), Kant discusses what he refers to as an old formula of the schools: *Nihil appetimus, nisi sub ratione boni; nihil aversamus, nisi sub ratione mali*. If this is taken to mean that “we desire nothing except with a view to our weal or woe” it is “at least very doubtful.” But if it is read as saying “we desire nothing, under the direction of reason, except in so far as we hold it to be good or bad” it is “indubitably certain” (C2 59–60). Similarly, in the *Groundwork* Kant says that “the will is a faculty of choosing only that which reason, independently of inclination, recognizes as practically necessary, i.e., as good” (G 412–29). Insofar as we are rational agents we will choose what is good – or take what we choose to be chosen as good.

As the identification of “good” with “practically necessary” in the *Groundwork* quotation suggests, Kant takes “good” to be a rational concept. This means two related things. First, reason must determine what is good. On this basis Kant argues in the *Critique of Practical Reason* that if the end were set by inclination and reason determined only the means, then only the means could be called “good” (C2 62). Thus, if an end is good, it must be set by reason; and if an action is done under the full direction of reason, then the end must be good. Second, and correlatively, if an end is deemed good it provides reasons for action that apply to every rational being:

What we call good must be, in the judgment of every reasonable man, an object of the faculty of desire, and the evil must be, in everyone’s eyes, an object of aversion. Thus, in addition to sense, this judgment requires reason. [C2 60–61]

It is this that gives rise to the *Groundwork* requirement, associated with the Formula of Humanity, that others "must be able to contain in themselves the end of the very same action" (G 430). If one's end cannot be shared, and so cannot be an object of the faculty of desire for everyone, it cannot be good, and the action cannot be rational.

From these considerations it follows that if there are perfectly rational actions, there must be good ends, and that when we act under the direction of reason, we pursue an end that is objectively good. But human beings, who act on their conception of laws, take themselves to act under the direction of reason. In the argument for the Formula of Humanity, as I understand it, Kant uses the premise that when we act we take ourselves to be acting reasonably and so we suppose that our end is, in his sense, objectively good. Perhaps it will at first seem odd that he uses that premise in an argument leading to a formula of the categorical imperative, since only if there is a categorical imperative will anything be in his sense objectively good. Here it is crucial to remember that the arguments leading to the formulations of the categorical imperative all tell us what the imperative will be like *if it exists*. Only *if* there is a categorical imperative will there be perfectly rational action; but if there is perfectly rational action there will be ends that are good.

Since good is a rational concept, a good end will be one for which there is reason – an end whose existence can be *justified*. But this by itself is not enough to establish a categorical imperative, for reasons can be relative: means for example can be called good, but only relatively to a given end (C2 62). If the goodness of an end is only relative, it will not have that claim upon all rational beings that Kant associates with the rationality of the concept "good," and cannot provide the basis for a categorical imperative:

The ends which a rational being arbitrarily proposes to himself as the consequences of his action are material ends and are without exception only relative, for only their relation to a particularly constituted faculty of desire in the subject gives them their worth. And this worth cannot, therefore, afford any universal principles for all rational beings or valid and necessary principles for every volition. [G 427–28]

What is required for a categorical imperative, therefore, is an end for which there is sufficient reason – an end whose existence can be

completely justified, and which therefore has a claim on every rational will. This is why Kant seeks "something the existence of which in itself has absolute worth" or an "end in itself" (G 428). Justification – the giving of practical reasons for ends and actions – is in one sense subject to the same fate as explanation – the giving of theoretical reasons for events. Reason seeks the "unconditioned," as the basis for an account (justification or explanation) that provides a sufficient reason.

As these comparisons suggest, the argument for the Formula of Humanity depends upon the application of the unconditioned/conditioned distinction to the concept of goodness. This follows from the fact that good is a rational concept. In any case where anything is conditioned in any way, reason seeks out its conditions, not resting until the "unconditioned condition" is discovered (if possible). An inquiry in the "analytic" or as Kant in one passage more helpfully calls it the "regressive" style (PFM 276n) is an argument in which something is taken as given or actual and the conditions of its possibility are explored. The arguments of the *Groundwork*, at least in the first two sections, like the arguments of the *Prolegomena*, are "regressive." If there is a categorical imperative, then there is fully rational action. If there is fully rational action, how is it possible? In the case of the Formula of Humanity the material of the law is sought through an investigation of the question: what is capable of fully justifying an end? What is unconditionally good?

In one sense, this question has already been answered in the first section of the book. There, Kant asserts that the only thing that can be conceived to be unconditionally good is a good will. The location of this claim shows us that Kant attributes it to "common rational knowledge of morals." It is used as a starting point for his analysis. In the remarks that follow, Kant elucidates the claim by explaining that the good will is the only thing that has its "full worth in itself" (G 394); and is the only thing whose value is in no way relative to its circumstances or results. Its value is independent of "what it effects or accomplishes" (G 394); it is in the strictest sense *intrinsically* good.

The good will is also said to be the condition of all our other purposes (G 396). This follows from its being the only unconditionally good thing. The value of anything else whatever is dependent upon certain conditions being met. Kant mentions talents of the

mind, qualities of temperament, gifts of fortune such as power, wealth, and health, and happiness among the things whose value is conditional. If the value of something is conditional, however, an inquiry into the conditions of its value should lead us eventually to what is unconditioned. This is partly affirmed in these early passages, for Kant tells us that the talents and temperamental qualities must be directed, the advantages used, and the happiness possessed by one with a good will in order that they be good. The good will is, in all cases, the unconditioned condition of the goodness of other things.

As the inclusion of happiness among the conditional goods shows, although Kant is claiming that the good will is the only thing whose value is intrinsic, he is not claiming that the good will is the only thing that is valuable as an end.⁴ Means are obviously conditional goods, for their goodness depends upon the goodness of the ends to which they are instrumental. But happiness, although clearly an end, and an end under which Kant thinks all of our other ends are subsumed, is also a conditional good whose value depends upon the good will. So Kant tells us that the good will is not the sole or complete good but "the condition of all others, even of the desire for happiness" (G 396). It is for this reason that "an impartial observer" disapproves "the sight of a being adorned with no feature of a pure and good will, yet enjoying uninterrupted prosperity" (G 393). But the impartial observer is equally dismayed by the idea that the virtuous person be without happiness:

That virtue (as the worthiness to be happy) is the supreme condition of whatever appears to us to be desirable and thus of all our pursuit of happiness and, consequently, that it is the supreme good have been proved in the Analytic. But these truths do not imply that virtue is the entire and perfect good as the object of the faculty of desire of rational finite beings. For this, happiness is also required, and indeed not merely in the partial eyes of a person who makes himself his end but even in the judgment of an impartial reason, which impartially regards persons in the world as ends-in-themselves. (C2 110)

A thing, then, can be said to be objectively good, either if it is unconditionally good or if it is conditionally good and the condition under which it is good is met. The happiness of the virtuous, for this reason, forms the other part of the "highest good": virtue, and happiness in proportion to virtue, together comprise all that is objectively good. A conditionally good thing, like happiness, is objectively good

when its condition is met in the sense that it is fully justified and the reasons for it are sufficient. Every rational being has a reason to bring it about, and it is this that makes it a duty both to pursue the happiness of others and, in general, to make the highest good one's end.

Since all objective value must come from unconditioned value, the good will is the source of all the good in the world. The highest good, as virtue and happiness in proportion to virtue; or the Kingdom of Ends, as "a whole of rational beings as ends in themselves as well as of the particular ends which each may set for himself" (G 433) are representations of a system of ends which can be said to be "synthesized" by the categorical imperative. This system is the totality of all that is objectively good under the unconditioned good; it is the systematic whole or unity formed by practical reason.

As pure practical reason it likewise seeks the unconditioned for the practically conditioned (which rests on inclinations and natural need); and this unconditioned is not only sought as the determining ground of the will but, even when this is given (in the moral law), is also sought as the unconditioned totality of the object of pure practical reason, under the name of the *highest good*. (C2 108)

I have said that practical reason shares the "fate" of theoretical reason insofar as it, too, is driven to "seek the unconditioned." In an important sense, however, the fate of practical reason is different from that of theoretical reason; this is one of the most central tenets of Kant's philosophy. Theoretical reason, in its quest for the unconditioned, produces antinomies; in the end, the kind of unconditional explanation that would fully satisfy reason is unavailable. Practical reason in its quest for justification is subject to no such limitation.⁵ This is part of Kant's doctrine of the primacy of practical reason. The argument for the Formula of Humanity provides an initial access to that doctrine, by showing that Humanity can be regarded as an unconditionally good thing, and a source of justification for things that are only conditionally good.

IV THE ARGUMENT FOR THE FORMULA OF HUMANITY

Having established that if there is a categorical imperative there must be something that is unconditionally valuable, Kant proceeds to argue that it must be humanity. Here is what he says:

All objects of inclinations have only a conditional worth, for if the inclination and the needs founded on them did not exist, their objects would be without worth. The inclinations themselves as sources of needs, however, are so lacking in absolute worth that the universal wish of every rational being must be indeed to free himself completely from them. Therefore, the worth of any objects to be obtained by our actions is at all times conditional. Beings whose existence does not depend on our will but on nature, if they are not rational beings, have only a relative worth as means and are therefore called "things"; on the other hand, rational beings are designated "persons" because their nature indicates that they are ends in themselves, i.e. things which may not be used merely as means. Such a being is thus an object of respect and, so far, restricts all [arbitrary] choice. . . . For, without them, nothing of absolute worth could be found, and if all worth is conditional and thus contingent, no supreme practical principle for reason could be found anywhere. (G 428-29)

In one sense, it seems as if Kant is just reviewing the available options in his search for something unconditionally good: considering objects of inclinations, inclinations, natural beings or "things", and finally persons, that being the one that will serve. But it is also possible to read this passage as at least suggesting a regress towards the unconditioned: moving from the objects of our inclinations, to the inclinations themselves, finally (later) back to ourselves, our rational nature. The final step, that rational nature is itself the objective end, is reinforced by this consideration:

The ground of this principle is: rational nature exists as an end in itself. Man necessarily thinks of his own existence in this way; thus far it is a subjective principle of human actions. Also every other rational being thinks of his existence by means of the same rational ground which holds also for myself; thus it is at the same time an objective principle from which, as a supreme practical ground, it must be possible to derive all laws of the will. (G 429)

I have quoted these rather long passages because my aim in what follows is to give a reconstruction along these lines – that is, on the assumption that the argument is intended as a regress upon the conditions. The reconstruction depends upon the ideas set forth in the previous sections. A rational action must be done with reference to an end that is good, and a good end is one for which there is a sufficient reason. It must be the object of every rational will, and it must be fully justified. If it is only conditionally good, the uncondi-

tioned condition for its goodness must be sought. Although we know already that the good will is this condition, the argument helps to show us what the good will must be, by showing us what will serve as such a condition.

Suppose that you make a choice, and you believe what you have opted for is a good thing. How can you justify it or account for its goodness? In an ordinary case it will be something for which you have an inclination, something that you like or want. Yet it looks as if the things that you want, if they are good at all, are good because you want them – rather than your wanting them because they are good. For "all objects of inclinations have only a conditional worth, for if the inclinations and the needs founded on them did not exist, their objects would be without worth" (G 428). The objects of inclination are in themselves neutral: we are not attracted to them by their goodness; rather their goodness consists in their being the objects of human inclinations.

This, however, makes it sound as if it were our inclinations that made things good. This cannot be right, for "the inclinations themselves, as sources of needs, however, are so lacking in absolute worth that the universal wish of every rational being must be indeed to free himself completely from them" (G 428). Now even without fully endorsing what Kant says here, we can easily agree that there are some inclinations of which we want to be free: namely those whose existence is disruptive to our happiness. Take the case of a bad habit associated with an habitual craving – it would not be right to say that the object craved was good simply because of the existence of the craving when the craving itself is one that you would rather be rid of. So it will not be just any inclination, but one that we choose to act on, that renders its object good.

Even consistency with our own happiness does not make the objects of inclination good, however. (Now I am, admittedly, departing from the passage I am interpreting – for Kant leaves it at the undesirability of having inclinations at all.) This is partly because we are not certain what our happiness consists in, but more because of a claim that has already been made in the opening lines of the *Groundwork*: we do not believe that happiness is good in the possession of one who does not have a good will. This is, of course, our great temptation – to believe that our own happiness is unconditionally good. But it is not really a tenable attitude. For either one must have

the attitude that just one's own happiness is unconditionally good, which is rather a remarkable feat of egocentricism,⁶ or one has to have the attitude that each person's happiness is unconditionally good. But since "good" is a rational concept and "what we call good must be, in the judgment of every reasonable man, an object of the faculty of desire" (C2 60–61), we cannot rest with the position that everyone's happiness, whatever it might be, is absolutely good. For:

Though elsewhere natural laws make everything harmonious, if one here attributed the universality of law to this maxim, there would be the extreme opposite of harmony, the most arrant conflict, and the complete annihilation of the maxim itself and its purpose. For the wills of all do not have one and the same object, but each person has his own. . . . In this way a harmony may result resembling that depicted in a certain satirical poem as existing between a married couple bent on going to ruin, "Oh, marvelous harmony, what he wants is what she wants"; or like the pledge which is said to have been given by Francis I to the Emperor Charles V, "What my brother wants (Milan), that I want too." (C2 28)

Given that the good must be a consistent, harmonious object of rational desire and an object of the faculty of desire for every rational being, one can take neither everyone's happiness nor just one's own happiness to be good without qualification: the former does not form a consistent harmonious object; and the latter cannot plausibly be taken to be the object of every rational will if the former is not. Thus happiness cannot in either form be the "unconditioned condition" of the goodness of the object of your inclination, and the regress upon the conditions cannot rest here. We have not yet discovered what if anything makes the object of your choice good and so your choice rational.

Now comes the crucial step. Kant's answer, as I understand him, is that what makes the object of your rational choice good is that it is the object of a rational choice. That is, since we still *do* make choices and have the attitude that what we choose is good in spite of our incapacity to find the unconditioned condition of the object's goodness in this (empirical) regress upon the conditions, it must be that we are supposing that rational choice itself *makes* its object good. His idea is that rational choice has what I will call a value-conferring status. When Kant says: "rational nature exists as an end in itself. Man necessarily thinks of his own existence in this way; thus far it is a subjective principle of human actions" (G 429), I read

him as claiming that in our private rational choices and in general in our actions we view ourselves as having a value-conferring status in virtue of our rational nature. We act as if our own choice were the sufficient condition of the goodness of its object: this attitude is built into (a subjective principle of) rational action. When Kant goes on to say: "Also every other rational being thinks of his existence by means of the same rational ground which holds also for myself; thus, it is at the same time an objective principle from which, as a supreme practical ground, it must be possible to derive all laws of the will" (G 429), I read him as making the following argument. If you view yourself as having a value-conferring status in virtue of your power of rational choice, you must view anyone who has the power of rational choice as having, in virtue of that power, a value-conferring status. This will mean that what you make good by means of your rational choice must be harmonious with what another can make good by means of her rational choice – for the good is a consistent, harmonious object shared by all rational beings. Thus it must always be possible for others "to contain in themselves the end of the very same action" (G 430).⁷

Thus, regressing upon the conditions, we find that the unconditioned condition of the goodness of anything is rational nature, or the power of rational choice. To play this role, however, rational nature must itself be something of unconditional value – an end in itself. This means, however, that you must treat rational nature wherever you find it (in your own person or in that of another) as an end. This in turn means that no choice is rational which violates the status of rational nature as an end: rational nature becomes a limiting condition (G 437–38) of the rationality of choice and action. It is an unconditional end, so you can never act against it without contradiction. If you overturn the *source* of the goodness of your end, neither your end nor the action which aims at it can possibly be good, and your action will not be fully rational.

To say that humanity is of unconditional value might seem, at first sight, somewhat different from the claim with which the *Groundwork* opens: that the good will is of unconditional value. What enables Kant to make both claims without any problem is this: humanity is the power of rational choice, but only when the choice is fully rational is humanity fully realized. Humanity, as I argued in Section III, is completed and perfected only in the realiza-

tion of "personality," which is the good will. But the possession of humanity and the capacity for the good will, whether or not that capacity is realized, is enough to establish a claim on being treated as an unconditional end.

V TREATING HUMANITY AS AN END IN ITSELF

Readers have often been puzzled by the prescription "treat humanity as an end." Kant's claims that the Humanity formula is closer to intuition (G 436), that the Formula of Universal Law gives the content the categorical imperative (G 420–21; 425), and that the latter ought to be used in actual decision making (G 437), might make it seem as if Kant does not intend this formulation to give definite directions for application, independently of its equivalence with the Formula of Universal Law. In opposition to this is the fact that Kant re-explains his *Groundwork* examples in terms of this formulation; in one case – the suicide example – providing a rather better account in terms of this formulation than he does in terms of universal law. Even more important, however, is the fact that all of the duties described in the *Metaphysical Principles of Virtue* are derived from the idea that humanity must always be treated as unconditionally valuable.

In fact, the argument that reveals the unconditional value of humanity also teaches us how to apply the Formula of Humanity. In order to know what is meant by "treating humanity as an end," we need only consider this argument, and see how humanity got to be an end in itself. What was in question was the source of the goodness of an end – the goodness say, of some ordinary object of inclination. This source was traced to the power of rationally choosing ends, exercised in this case on this end. So when Kant says rational nature or humanity is an end in itself, it is the power of rational choice that he is referring to, and in particular, the power to set an end (to make something an end by conferring the status of goodness on it) and pursue it by rational means.

The question is then: what is involved in treating your own and every other human being's capacity for the rational choice of ends – that is to say, for conferring value – as an end in itself? There are several things that are important to keep in mind. First, Kant thinks that this end functions in our deliberations negatively – as some-

thing that is not to be acted against. The capacity for rational choice is not a purpose that we can realize or something for us to bring into existence. Second, it is an unconditional end, and that has two important implications. The first is that as an unconditional end it must *never* be acted against. It is not one end among others, to be weighed along with the rest. The second implication in a sense gives the reason for the first: as an unconditional end it is the condition of the goodness of all our other ends. If humanity is not regarded and treated as unconditionally good then nothing else can be objectively good. As Kant puts it in the *Groundwork*: "the subject of a possible will which is absolutely good . . . cannot be made secondary to any other object without contradiction" (G 437). No relative end can be pursued as if it were better or more important than humanity itself without a kind of contradiction.

While it would not be feasible to go through all of the many cases Kant gives of duties derived from the Formula of Humanity, I want to say something about the two major kinds of derivations from this formula that exist, and how each is supposed to work. In order to do this, I will concentrate on the *Groundwork* examples.

The first treatment of the examples, used to illustrate the workings of the Formula of Universal Law, divides them into two groups. The duty not to commit suicide because of the prospect of wretchedness, and the duty not to make a false promise because of a financial emergency, illustrate cases in which the maxim cannot be thought as a universal law without contradiction. Under the Formula of Humanity, these are classified as cases in which humanity, in your own person or another, is treated as a mere means. The duties of developing your talents and powers, and of helping others, are classified under Universal Law as cases where the maxim, though thinkable as a universal law, cannot be willed as such. Under the Formula of Humanity, these are classified as cases in which the action is not in conflict with humanity, but fails to "harmonize" with it (G 430). In both cases, what is involved is a failure to properly acknowledge in your conduct the value-conferring status either of another or of yourself. We can make this plausible, and also see why Kant takes the two formulas to be identical, by considering the examples and the way Kant explains them.

In the suicide case, Kant says that "if, in order to escape from burdensome circumstances, he destroys himself, he uses a person

merely as a means to maintain a tolerable condition up to the end of life" (G 429). As mentioned before, Kant takes it to be a consequence of his argument that humanity cannot be made secondary to any relative end without contradiction. This is what happens in the case of suicide: the end, in the example, is "a tolerable condition" and the means is the destruction of a rational being – hence a rational being is being used as a mere means to a relative or conditional end. The reason why this is said to be a contradiction rather than merely a case of misordered values is that the relative end must get its value from the thing that is being destroyed for its sake. However obvious it may seem that a "tolerable condition" is a good thing, it is good only because of the value conferred upon it by the choice of a rational being. Destroy the rational being, and you cut off the source of the goodness of this end – it is no longer really an end at all, and it is no longer rational to pursue it.

The false promising case is slightly more complicated, and easier to explain if we use the result of applying the Formula of Universal Law. Because we know from that test that the maxim of false promising in order to escape a financial emergency could not be universalized, we know that false promising could not be the universal method of escaping financial emergencies. From this it follows, because reasons must be universal, that the desire to escape a financial emergency cannot justify (be a sufficient reason for) making a false promise. If you make a false promise, then, you accord to your value-conferring capacity a greater power, so to speak, than you do that of others. You act as if your desire to avoid financial trouble has a justifying power that someone else's exactly similar desire would not have. But this cannot be right: if your desire gets its justifying power from your humanity, then any other person's similar desire would have the same justifying power. If the end of your action is not good because of your humanity, on the other hand, it cannot be good at all, and your action is not rational. There is another way to describe this kind of case that is perhaps even better for bringing out the violation of humanity that is involved. Whenever you violate the first contradiction test under the Formula of Universal Law, and act on a maxim that cannot be universalized, you must be using some method to achieve your end that not everyone could use to achieve that end. The efficacy of your action depends upon the fact that others do not act as you do, and that in a sense means that

others are making your method work. This is characteristic of the kind of violation of duty that is most amenable to treatment under the first contradiction test. For example, when you tell a lie for a certain purpose, the lie works to achieve the purpose only because most people tell the truth. That is why you are believed, and so why the lie achieves its purpose. In such a case it is not just the person to whom you lie that you treat as a means but all of those who tell the truth. This is because you allow their actions to fuel your method, and that is explicitly treating their rational nature as a mere means: indeed it is making a tool of other people's good wills. Whenever you use a method that works only because others do not use it – which is the first contradiction test reveals – you make an instrument of the rational nature of others, and treat them as mere means.

The third and fourth examples, of the duty of self-perfection and the duty to promote the happiness of others, admit of very clear accounts in terms of the idea of acknowledging the value-conferring power of rational beings as ends in themselves. In the case of the duty of self-perfection, it is a question of developing and realizing the capacities which enable you to exercise your power of rational choice – the talents and powers that make it possible for you to set and pursue ends. It is your powers as an agent that are to be promoted. This, indeed, is as close as Kant comes to assigning a positive function to humanity as an end. What makes this possible is the fact that rational nature is a sort of capacity. It is, as Kant says, not an end to be effected (G 437), for rational nature is not something that we can create; nevertheless, we can realize our rational capacities more or less fully, and this is what generates the various positive duties of self-perfecting. In the case of the duty to promote the happiness of others, Kant says:

For the ends of any person, who is an end in himself, must as far as possible also be my end[s]^B if that conception of an end in itself is to have its full effect on me. (G 430)

This is because the full realization and acknowledgment of the fact that another is an end in itself involves viewing the end upon which this person confers value as *good* – and when one acknowledges that something is good, one acknowledges it to be "in the judgment of every reasonable man, an object of the faculty of desire." To treat another as an end in itself is to treat his or her ends as

objectively good, as you do your own. To treat anyone as an end in itself is to regard that person as one who confers value on the objects of his or her choice.

VI CONFERRING VALUE

In this last section I want to bring in as a final piece of support for the reading I have given a set of passages from the *Critique of Judgment*. These passages seem to me to support in a very forceful way the idea that it is our power to confer objective value that Kant thinks of as having unconditional worth.

In the "Methodology of Teleological Judgment," Kant is concerned with the question of what might appropriately play the role of final purpose of creation. He has established the idea of a natural purpose, which provides the basis for a teleological interpretation of nature; but in order to view nature as a teleological *system*, we must discover its final purpose: we must discover, that is, a reason for the existence of nature itself. This will not be a purpose internal to nature, but one outside or independent of it: "an objective supreme purpose, such as the highest reason would require for creation" (C3 436). In carrying out his inquiry, Kant undertakes a familiar sort of regress argument. In this case, the condition of a given thing is its purpose, and the regress must end with something which is in itself a final purpose, "that purpose which needs no other as the condition of its possibility"; something about which "it can no longer be asked why" it exists (C3 434-35). Starting with the idea of a natural purpose, the argument proceeds to what Kant calls an ultimate purpose: this will be that which we judge to be the purpose *within* nature towards which all nature is organized. This ultimate purpose, being as it were nature's contribution to the final purpose, will give us an idea of what the final purpose is.

Beginning with consideration of vegetable nature, we can reason back to the ultimate purpose as follows:

a more intimate knowledge of its indescribably wise organization does not permit us to hold to this thought (that it is a mere mechanism), but prompts the question: What are these things created for? If it is answered: For the animal kingdom, which is thereby nourished and has thus been able to spread over the earth in genera so various, then the further question comes:

What are these plant-devouring animals for? The answer would be something like this: For beasts of prey, which can only be nourished by that which has life. Finally we have the question: What are these last, as well as the first-mentioned natural kingdoms, good for? For man in reference to the manifold use which his understanding teaches him to make of all these creatures. He is the ultimate purpose of creation here on earth, because he is the only being who can form a concept of purposes and who can, by his reason, make out of an aggregate of purposively formed things a system of purposes. (C3 426-27)

Kant then goes on to inquire for the more specific feature of human life that is the purpose of nature – that is to say, for something "found in man himself which is to be furthered as a purpose by means of his connection with nature" (C3 429). This Kant supposes must be either human happiness or human culture (C3 429-30). At this point Kant takes up an argument that is repeated throughout the teleological historical writings and also appears in the teleological argument at the beginning of the *Groundwork*. Happiness does not seem to be something that nature can achieve or aims at achieving: the evidence favors culture. But furthermore, the argument requires that the ultimate purpose of human beings within nature teaches us what the final purpose of nature itself is. And happiness cannot be the final purpose of nature, for the same reasons that we found, in the argument for the Formula of Humanity, that happiness could not be the unconditioned good.

Happiness, on the contrary, as has been shown in the preceding paragraphs by the testimony of experience, is not even a *purpose of nature* in respect of man in preference to other creatures, much less a *final purpose of creation*. Men may of course make it their ultimate subjective purpose. But if I ask, in reference to the final purpose of creation, Why must men exist? then we are speaking of an objective supreme purpose, such as the highest reason would require for creation. If we answer: These beings exist to afford objects for the benevolence of that supreme cause, then we contradict the condition to which the reason of man subjects even his own inmost wish for happiness (viz. the harmony with his own internal moral legislation). (C3 436n)

The answer then will be that the ultimate purpose in nature is "culture" in the specific sense of the development of humanity:

The production of the aptitude of a rational being for arbitrary purposes in general (consequently in his freedom) is *culture*. (C3 431)

Culture, according to Kant, is an appropriate ultimate purpose in nature, for the development of culture is something that "nature can do in regard to the final purpose that lies outside it" (C3 431).

The final purpose of nature, Kant argues, is morality itself. It is in morality that the aptitude of setting purposes before ourselves finds its completion, for an end or purpose must be objectively good, and only in morality do we find the unconditioned condition of its goodness:

Only in man, and only in him as subject of morality, do we meet with unconditioned legislation in respect of purposes, which therefore alone renders him capable of being a final purpose, to which the whole of nature is teleologically subordinated. (C3 435-36)

It is our capacity to set ends – to freely choose what shall be an end by means of reason, that not only makes every rational being an end in itself, but which forms the only possible final purpose of nature, teleologically conceived. It is only this capacity that has its value completely in itself; so that this not only forms the basis of a possible categorical imperative, but also the only possible basis for a complete teleological view of creation.

Without men the whole creation would be a mere waste, in vain, and without final purpose. But it is not in reference to man's cognitive faculty (theoretical reason) that the being of everything else in the world gets its worth; he is not there merely that there may be someone to *contemplate* the world . . . we must presuppose for it a final purpose, in reference to which its contemplation itself has worth. Again it is not in reference to the feeling of pleasure or to the sum of pleasures that we think a final purpose of creation as given; . . . or, in a word, by happiness. For the fact that man, if he exists, takes this for his final design gives us no concept as to why in general he should exist and as to what worth he has in himself. . . . But it is that worth which he alone can give to himself and which consists in what he does, how and according to what principles he acts, and that not as a link in nature's chain but in the *freedom* of his faculty of desire. That is, a good will is that whereby alone his being can have an absolute worth and in reference to which the being of the world can have a *final purpose*. (C3 442-43)

Or as Kant puts it in an earlier footnote:

There remains, then, nothing but the value which we ourselves give our life, . . . in such independence of nature that the existence of nature itself can only be a purpose under this condition. (C3 434n)

On Kant's view it is human beings, with our capacity for valuing things, that bring to the world such value as it has. Even the justification of nature is up to us.

NOTES

- 1 The three arguments mentioned here are in the *Groundwork* at 420-21, 427-29; 431-32.
- 2 In fact, the example is explicitly given as one in which the agent has a "direct inclination" to the action and its purpose as opposed to the kind of case (like that of the "honest" grocer) in which the agent has no direct inclination but is "impelled to do [the action] by another inclination." That is, the action with its purpose is an end and not a means (G 397). Admittedly, Kant's discussion of happiness in the *Critique of Practical Reason* (C2 23) suggests a hedonistic view of the inclinations. But though the sympathetic character may value the beneficent action because it is pleasant, this *may* still be taken as an explanation of why he values it rather than a reduction of it to a mere means. (This is the way Mill, as I understand him, proposes to interpret hedonism in Chapter IV of *Utilitarianism*.) However, Kant's language in the *Critique of Practical Reason* is that of the more conventional sort of hedonism which makes pleasure and the avoidance of pain ends to which everything else desired must be regarded as means.
- 3 If Kant's view of the inclinations in the *Critique of Practical Reason* is understood to be a conventional hedonistic view [see note 2], then the account of happiness given there must count either against my reading of these remarks in the *Groundwork* or against the *Groundwork's* "Ideal of the Imagination" view itself. Beck thinks that the difficulty is in Kant's psychological views. See *A Commentary on Kant's Critique of Practical Reason*, p. 101. I would like to thank a reader for *Kant-Studien* for this reference, as well as for other useful references and comments.
- 4 A more extended version of some of the arguments of this section appears in my paper *Two Distinctions in Goodness*, Chapter 9 in this volume. In that paper I compare G. E. Moore's conception of intrinsic value as something possessed by all ends that ought to be valued for their own sakes and Kant's conception of unconditional value as possessed only by the good will.
- 5 Practical reason does of course also have its antinomy, but it results not from a failure to locate the unconditioned as the "determining ground of the will" (the original source of all justification), but rather from the apparent failure of the unconditioned principle to produce the associated

unconditioned totality, the highest good, in the natural world (C2 108; 113-14).

- 6 Some of the standard arguments against the rationality of egoism might be useful in supporting this point: for example, that of G. E. Moore in *Principia Ethica*, Sections 59-61. See also the discussions by Thomas Nagel in *The Possibility of Altruism*, especially Chapter X.
- 7 The idea that choice in accordance with the moral law is the basis of the concept of "good" (and not the reverse) is also argued in the *Critique of Practical Reason*, in the chapter on "The Concept of an Object of Pure Practical Reason" (57-67).
- 8 Beck translates this passage "as far as possible also be my end." On my reading, the plural fits the sense of the passage better.