

Comment on Thomas Hill, Jr.: **Kant's Theory of Practical Reason**
by Christine M. Korsgaard

Professor Hill has made commentary difficult, by presenting a lucid account of Kant's theory, with which I agree almost completely. His account falls into two parts. In the first part of his paper, Hill gives us an interpretation of Kant's theory of practical reason in general, drawing our attention to some important differences between Kant's view and some familiar contemporary views. In the second part of his paper, Hill shows how Kant's theory of practical reason naturally gives rise to a theory of moral reasons, in part because of its distinctive features.

Hill contrasts Kant's theory to three contemporary views which Kant would have found unacceptable. Substantive principles direct us to pursue intrinsically valuable goods or perform intrinsically valuable actions. Maximizing principles direct us to maximize some end such as pleasure or the satisfaction of desire. These ends may of course be thought of as intrinsically valuable, but more commonly they are selected because they are universal or thought to render our other ends commensurable. Substantive and maximizing views are naturally associated with the idea that all practical reasoning is instrumental, but as Hill points out, even Kant's theory of instrumental reason is different from them. The third contemporary view is closer to Kant's in that its representatives develop non-instrumental reasons from an appropriate standpoint, but in these contemporary views adopting the standpoint and the reasons it provides remains optional. Kant, by contrast, develops practical reasons from a standpoint that is supposed to be inescapable. The standpoint in question is that of the autonomous rational deliberator, seeking good reasons for what she chooses or does. As Hill carefully brings out, this standpoint is involved in non-moral as well as moral reasoning. This is why Kant's theory of practical reason in general leads naturally to a theory of moral reasons.

Let me first say something by way of endorsement of Hill's position with regard to non-moral reason. One of the most important features of Kant's theory of non-moral reason, and perhaps most alien to those trained in an empiricist tradition, is that even non-moral ends are determined to some extent by reason, and are to be regarded as objects of free choice. Kant says this in various places, but it is clearest in the essay "Conjectural Beginning of Human History" in which he discusses the way the awakening of reason and the associated power of comparison proliferates human ends. This, Kant says, poses a problem because it makes us capable of free choice but does not tell us how to choose. Morality is what ultimately solves the problem. Maximizing principles are unacceptable even as principles of non-moral reasoning, for their ends are not determined by free rational choice. Kant is also clear in his denial that the ends we pursue are intrinsically valuable, for, he says, "Ends that a rational being adopts arbitrarily as effects of his action (material ends) are in every case only relative; for it is solely their relation to special characteristics in the subject's power of appetite which gives them their value." (G.427/HJP.95) Our ends get their value from the fact that they satisfy our needs and inclinations, and in general from their suitability to the physiological and psychological conditions of human life. Even happiness is only valuable because we are, as Kant says, "in need of it." (K2.110/LWB.114) This means that the things we desire are only conditionally valuable. And this is one of the ways in which Kant's theory of practical reason leads naturally to morality. For the consideration that desired ends are conditionally valuable sets us on a quest for an unconditional reason for holding an end, and this can only be supplied by morality.

But it is conceivable, as far as that goes, that the only reason not to satisfy a desire comes from some other desire, or that the only condition on the value of your happiness is that it be compatible with that of others. In that case the only sense in which an end could be unconditionally valuable would be that it was desirable on the whole. As Hill

points out, Kant's theory of instrumental rationality always leaves the agent an option: adopt the means, or give up the end. We are negatively free and can always give up the end if there is reason. But a further argument is needed to show that rational agents are committed to any reasons for giving up ends other than that the ends are not conducive what is desirable on the whole. As Hill shows in the second half of his paper, Kant's view is that the perspective of the autonomous rational agent does provide grounds for such further argument. From this perspective, what is called for are reasons and principles that autonomous rational agents as such are committed to. As Professor Hill argues, one such principle is that of the value of autonomous rational agency itself, as the source of value of our ends and actions generally. This justifies the Formula of Humanity.

Hill professes skepticism about justifying the Formula of Universal Law in this manner, but I am going to make a stab at saying how it may be done. First, however, I must mention one passage in Hill's paper with which I want to quibble. Warning us against an "existentialist" interpretation according to which individuals simply choose what is to count as a reason for them, Hill remarks that

[the will with autonomy] ... is a will that acknowledges what is rational and commands accordingly ... rather than a "voluntarist's" will that "invents" the standards, making the non-rational rational by choice alone. Autonomy means identifying with rational constraints, not seeing them as imposed by others or as something we could give up without ceasing to be ourselves; but it does not mean we "voluntarily" choose to commit ourselves to them the way we might choose to accept the rules of a club, a church, or a private code. Hill, p. 13 (shorter version)

While I think that there is something right about this, I also think it can be misleading. It makes it sound as if we could specify the rational constraints which the

autonomous will must accept independently of a consideration of the essentially legislative function of the autonomous will. Whereas it seems to me that the Kantian project quite generally might be characterized as an attempt to justify the content of purportedly rational principles on the basis of the tasks and functions of reason. Since choice is the function of the will, arguments about its principles should be cast as arguments about what it has reason to choose.

I am going to describe such an argument, based in part on the same texts in the *Groundwork* which Hill has been discussing, and in part on some things Kant says in *Religion Within the Limits of Reason Alone*. First of all, we will need to have in place a piece of technical terminology that Kant uses to describe the relation of the autonomous agent to the options among which she chooses when she acts. The term is incentive (*Triebfeder* - Paton uses "impulsion"). An incentive is something that makes an action interesting to you, that makes it a live option. It is a candidate for a motive. It does not become a motive, however, until the agent has adopted it into her maxim. (R.19) Desires and inclinations are incentives; so is respect for the moral law. One way to capture the idea that an inclination by itself is not a reason for action is to say that an inclination of itself is merely an incentive, and does not become a reason for action until the agent has adopted it freely into her maxim. The question will then be what justification she has for doing so.

Imagine an attempt to discover the fundamental motivational structure, or complete rationale, of an action. I ask to know why you are doing something, and you give me a reason, tell me your proximate end. I then ask why you want that, and most likely you cite some larger end or project. I press on, demanding your reason at every step, until we reach the moment when you are out of answers. You have shown that your action is calculated to assist you in achieving what is desirable on the whole, what, having deliberated scrupulously, you have determined that you want most.

The reasons that you have given can be cast in the form of maxims derived from imperatives. From a string of hypothetical imperatives, instrumental and prudential, you have derived the maxim that we might formulate:

I will do this action, in order to get what I desire.

According to Kant, this maxim only determines your will if you have adopted another maxim that makes it your end to get what you desire. This maxim is:

I will pursue as ends the things I desire.

Now suppose that I want to know why you have adopted this maxim.

Two answers we can dismiss immediately. First, suppose there is a psychological law of nature that runs something like "a human being always adopts as ends the things he desires." This is a denial of the negative freedom that Kant supposes we attribute to ourselves in any kind of deliberation.

Second, suppose that you have adopted this maxim randomly. There is nothing further to say. You could have adopted some other maxim, since your will is free. As we know, Kant rejects this, as being inconsistent with the very idea of a will, which does what it does for a reason. Since a reason is derived from a principle, we cannot say the will chooses and of its principles randomly.

This leaves us with a deep problem of a familiar kind. If you can give a reason you derive it from some more fundamental maxim, and I can ask why you've adopted that one. If you cannot it looks as if your principle was selected randomly. Obviously, to put an end to a regress of this sort what you need is something about which it is impossible, unnecessary, or incoherent to ask why. This must be something that, as Professor Hill puts it, the autonomous rational agent *as such* can be shown to be necessarily committed to. One way to move to the Formula of Humanity from this point is to observe that the maxim of adopting the objects of your desires as ends cannot be regarded as rationally unchallengeable as long as you concede that desires have only a conditional worth. You

can then ask what gives them worth, and the unconditional value of rational beings will be the answer. However, I want to take a different route, one that leads through a voluntarist argument to the Formula of Universal Law.

Kant tells us - rather grandly as Professor Hill says, that the Will is practical reason. Choices are always for reasons, and reasons are derived from principles, or laws. If the Will is autonomous, the Will must decide for itself what it will count as a law, hence what it will count as a reason. Although it is fanciful, we can think of the Will as if it were an agent, and imagine this determination of what is to count as a reason as if it happened at a certain time. The Free Will is called upon to choose its most fundamental principle. Yet, precisely because at this point the will has not yet determined what is going to count as a reason, it seems as if there is no reason for it to choose one principle rather than another. Kant calls this feature of the Will its "spontaneity."

As far as the argument stands now, it looks as if the Will could adopt any maxim we can construct. It could adopt the maxim of pursuing the things you want least, or pursuing the things you find detestable. Actually, however, these are not serious options, for reasons that come out most clearly in *Religion Within the Limits of Reason Alone*. Finding something detestable is not, for us human beings, an incentive for pursuing it. Although incentives do not yet provide reasons for the Will in the position of spontaneity, they do determine what the options are - which things, so to speak, are candidates for reasons. In the *Religion*, Kant claims that it is impossible for a human being not to be moved at all by incentives; our freedom, rather, is exercised in choosing the order of precedence among the different kinds of incentives. So the real choice will be between a maxim of self-love, which subordinates incentives of morality to incentives of inclination, and the moral maxim, which subordinates incentives of inclination to moral ones. The maxim of self-love says something like: I will do what I desire, and what is morally required if it doesn't interfere, and the moral maxim says, I will do what is

morally required, and what I desire if it doesn't interfere. The moral maxim is of course the maxim derived from the categorical imperative.

It looks at first as if the problem here is to show that there is some reason for the Will in a position of perfect spontaneity to choose the moral law rather than the law of self-love. Yet this seems impossible, since the Will by hypothesis has not yet determined what it counts as a reason. But on reflection we can see that the problem can be circumvented. We need only consider the position of the Will, and the content of the Formula of Universal Law.

At the standpoint of spontaneity, the Will must, in order so to speak to commence operations, choose a principle or law for itself. Nothing provides any content for that law. All that it has to be is a law.

Suppose it chooses the categorical imperative, as represented in the Formula of Universal Law. This formula merely tells us to choose a law. Nothing provides any content for that law. All that it has to be is a law.

By making the Formula of Universal Law its principle, the Free Will retains the position of spontaneity. The categorical imperative is thus shown to be a law of spontaneity. In a sense, it simply describes the task of an autonomous will.

On the other hand, suppose the Will chooses the law of self-love. In that case, it departs from its position of perfect spontaneity and puts itself in the service of inclination, acquiring a master. What is more, there is absolutely no reason for it to do so. Since we are just talking about the Will itself right now, not the whole person, the incentives of inclination cannot provide a temptation to do this. Incentives of inclination cannot move the Will to abandon its position of perfect spontaneity, since they cannot move the Will at all until after it has already abandoned that position by resolving to be so moved by them.

This argument shows that there are not really two choices, morality and self-love, on an equal footing. The Will that makes the categorical imperative its law merely reaffirms its independence of everything except law in general. Making the categorical imperative its law doesn't require the spontaneous Will to take action - it is already its principle. Surrendering the position to the law of self-love requires an action (R.26) - and furthermore, one for which there is no reason. Thus, not only are the two options not on a footing, but the choice of the maxim of self-love is unintelligible. Morality is the natural condition of a Free Will. The Free Will that puts inclination above morality sacrifices its freedom for nothing.

You will now think that the argument is going to turn out to be too strong. If the Will is Free, moral evil is incomprehensible, for on this argument, moral evil is the will's unmotivated abandonment of its freedom. But this is Kant's view. Moral evil is, as he puts it in the *Religion*, inscrutable. It is a Fall, in the Biblical sense, and it is exactly as hard to understand as the Fall in the Bible. (R.36ff)

But as Professor Hill points out, this does not lead to the all or nothing view that we are only responsible for our good actions, because only free when we perform those. A free but evil Will is shown to be unintelligible from the standpoint of pure practical reason, but not impossible. But it is not surprising that a deep form of irrationality should be unintelligible. And we are whole persons, not just spontaneous Wills. Unlike the anthropomorphized Will in the story I have just told, we are imperfectly rational, and this is in part because we *are* subject to temptation from inclinations. In the *Metaphysical Principles of Virtue*, Kant explains that this is why, for human beings, freedom can only be achieved through virtue - the adoption of morally worthy ends. For, he says, since sensible inclinations may misdirect us to ends which may be contrary to duty, legislative reason can only guard against their influence by setting up opposing moral ends. (slight paraphrase, MM.381/JWE.38/MG.38) Moral evil is unintelligible from a practical point

of view, but from the theoretical standpoint we use in explaining conduct, nothing is easier to understand. To suppose otherwise is to give a practical Idea a kind of theoretical or explanatory weight that Kant denies to such Ideas.

The argument about spontaneity is not supposed to show that moral evil is so irrational that it is impossible, but rather to provide an account of the interest that we attach to moral ideas. This is what Kant says is needed to break the supposed circle in *Groundwork* III. The interest is in the idea of ourselves as free in the sense described. The will that adopts self-love as its maxim is determined by inclinations, and inclinations, as Kant sees them, are ultimately determined by natural forces. Such a will becomes a mere conduit for natural forces. A moral will by contrast aspires to make a difference in nature, to make the world conform to rational demands, to make it a kingdom of ends. I believe Kant thinks of this as being, roughly speaking, the motivating thought of morality: one's autonomy means that one can make a difference, make the world a more rational place.

Let me conclude by summarizing what I have said. I have agreed completely with Professor Hill's contentions about the distinctive features of Kant's theory of practical reason. Practical reasons at every level are developed from the standpoint of an autonomous rational deliberator in search of good reasons for choice and action. From this point of view, not only means but ends must be supported by reasons, and the search for reasons acceptable from a standpoint of autonomy leads in a natural way to morality. Because the standpoint of the autonomous rational deliberator is inescapable, so also is the moral point of view. The content of moral principles is established by showing that there are principles an autonomous rational agent as such must accept. Professor Hill has done this for the Formula of Humanity, and I have tried to sketch a way it may be done for the Formula of Universal Law. The only thing I have disagreed with at all is his contention that the autonomous will should not be seen as choosing which considerations

are to count as reasons: I have suggested that in a limited sense a voluntarist conception is appropriate. In determining the order of the adoption of incentives into maxims, the will does decide what will count as a reason. Since there is only one decision for it to make, this is hardly a serious disagreement. Professor Hill claims that his paper is aimed at interpretation rather than defense, but I think it shows us the strategy that defense must take.