

Decision Procedures, Moral Criteria, and the Problem of Relevant Descriptions in Kant's Ethics

Mark Timmons

Kant's supreme principle of morality, the Categorical Imperative, is often interpreted as providing a decision procedure or test that agents can use to figure out what, in a particular case, they ought or ought not to do. Again, as a supreme principle, it supposedly also provides a criterion of right action, in the sense that it specifies some fundamental morally relevant feature or property of actions in virtue of which actions have some particular deontic status including, of course, an action's being a duty. One way to think about how a moral principle like the Categorical Imperative provides us with a decision procedure is to think of the fundamental morally relevant feature specified by the principle as being some feature or property of actions whose presence the agent can detect and thereby form a judgment about whether or not the action in question is a duty. The Categorical Imperative, then, is often taken to provide us with both a decision procedure, for purposes of guiding deliberation, and a criterion of right action specifying an ultimate or fundamental right-making feature of actions that explains why, in the end, certain actions have the deontic status they do.

Consider, then, the Universal Law formulation of the Categorical Imperative, which, for our purposes can be expressed this way: *An action is morally right if and only if the maxim associated with the action is universalizable (i.e., can be consistently willed as universal law)*. If we construe this as giving us both a criterion of right action and a decision procedure, here is what we have: (1) the ultimate right-making feature of actions concerns facts about associated maxims and, in particular, whether or not such maxims have the property of being universalizable, and (2) in order to figure out and/or justify a claim about the deontic status of some action, one attempts to determine whether the ultimate right-making feature is present, and to do this, one is to employ the tests Kant describes for a maxim's being universalizable. Further, if one takes seriously Kant's claim that, in some sense, the various formulations of his supreme principle are equivalent, then one will be inclined to view them all as providing both a criterion¹ and a decision procedure.

¹ I plan to use the expression, 'moral criterion' (or just 'criterion') to refer to those features or properties (whatever they are) *in virtue of which* an action is right or wrong. This is

But there are reasons for being suspicious of this way of understanding how the various formulations of Kant's supreme principle of morality are related in his theory of right conduct. For instance, if we suppose that the Categorical Imperative presents us with a single criterion of right action, and if we then suppose that the various formulations play the same role in Kant's moral theory, we have to conclude (so it seems) that the Universal Law formula and the Humanity as an End in Itself formula both specify the same criterion of right action and that consequently the property of one's maxim being universalizable is the same property as an action's respecting humanity. But these aren't the same property, even though there may be some quite intimate relation between them. Some interpreters,² recognizing this, have wanted to go against the grain of standard Kant scholarship and claim that because the various formulations specify different criteria of right action, Kant is an ethical pluralist, in the sense that his theory specifies not one, but two basic right-making features of action each of which functions as criterion. What is being assumed here is that the various formulations express a criterion of right action, an assumption that I think is probably wrong.

Another, related reason for being suspicious of the idea that the various formulations of the Categorical Imperative play identical roles concerns the fact that there seems to be a fundamental difference in focus between these two formulations. That is, the Universal Law formulation, with an emphasis on a feature of the agent's maxims and consistent willing, seems to approach duty from a first-person perspective. However, the Humanity formulation, *Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as means, but always at the same time as an end*, especially as it gets used in the *Tugendlehre* as a basis for elaborating a scheme of duties, seems to approach duty more from a third-person perspective, with primary focus, not on agents and their maxims, but rather on the intrinsic nature of various actions. Again, the difference in focus suggests that perhaps these formulations play importantly different roles in Kant's theory of right action.

To lay my cards on the table, I want to propose what I will call a 'differential roles' interpretation of the Categorical Imperative, according to which (roughly) the various formulations of this principle play importantly different roles in the overall economy of Kant's theory of right conduct. Space does not permit me to discuss all of the various formulations of the Categorical Imperative; instead, I will focus exclusively on the Universal Law and Humanity formulations. I begin with two rather intuitive ideas that I want to develop and defend as I proceed: (1) The first formulation of the Categorical Imperative is best understood as being primar-

in keeping with how it is used by some contemporary philosophers in drawing a contrast between a moral principle as a decision procedure and a moral principle as a criterion of right action. For more on this, see section I.

² See, for example, *H.J. McCloskey, Meta-Ethics and Normative Ethics*, The Hague: Nijhoff, 1969, ch. 8.

ily a decision procedure and hence giving us an agent-centered perspective on duty, while (2) the second formulation is best understood primarily as a criterion of right action, specifying those features of actions and circumstances (including, of course, features of agents) in virtue of which actions are, objectively speaking, right or wrong. Here are some implications that I plan to draw from this way of viewing these two formulations. First, and most obviously, on this way of viewing these principles they play importantly different roles in the overall economy of Kant's moral theory. Second, because the Humanity formulation provides, as it were, the objective basis of right action, in telling us what (objectively) makes an action right, it provides the basis for determining which features of actions are morally relevant. In a sense, the two formulations work together: the second one provides some raw material, as it were, for use in the first. A maxim is what Kant calls a subjective principle of action, representing an agent's intention and thus how the agent views the circumstances and action. The notorious problem of relevant descriptions stems from the fact that there are numerous ways of conceptualizing one's circumstances and action, and hence many alternative maxims that might be associated with any one action. The problem is one of specifying which features of one's circumstances and action should be reflected in one's maxim for purposes of moral deliberation. The second formulation, I suggest, gives us some sort of handle on this problem. Third, Kant's solution to the problem of relevant descriptions provides the basis for a response to various forms of pessimistic appraisal of the utility of Kant's universality tests as moral decision procedures. Finally, I hope to shed some light on the question of how these two formulations are related in Kant's ethics.

Here is the plan of the paper. The first section is a brief overview of the distinction between decision procedures and criteria, and how they are related to the enterprise of moral theory. Next, in section two, I begin with the idea that the Universal Law formulation is a decision procedure which leads to a discussion, in section three, of a certain challenge, owing to Onora O'Neill, of our understanding of the power of Kant's universality tests. To meet her challenge requires showing how Kant's tests can test for the objective rightness of actions and not just their subjective rightness. One way to meet the challenge is to provide a theory of objective moral relevance, which I take up in section four devoted to the role of the Humanity formulation of the Categorical Imperative. Finally, in section five, I turn to various interpretive questions that my so-called differential roles interpretation prompts with the mainly critical aim of seeing how well the interpretation holds up as a reading of Kant.

I. Decision Procedures, Criteria, and Moral Theory

There are two main aims of moral theory, one practical and one theoretical, that are often enough not clearly distinguished. The practical aim of traditional moral theory is to provide agents with a decision procedure for use in the context of prac-

tical deliberation. Henry Sidgwick spoke of the practical aim in terms of *methods*, which he defined this way:

[A] 'Method of Ethics' is explained to mean any rational procedure by which we determine what individual human beings 'ought' – or what it is 'right' for them – to do, or to seek to realise by voluntary action.³

But Sidgwick also recognized a theoretical aim of moral theorizing, since in addition to discovering a method, "the student of Ethics seeks to attain systematic and precise general knowledge of what ought to be, and in this sense his aims may properly be termed 'scientific'."⁴ The sort of 'general knowledge' Sidgwick had in mind has to do with the 'deep' nature of right and wrong that we obtain when we understand those features or properties in virtue of which actions have whatever deontic status they have.

It has been common to suppose that the two aims neatly coincide in the sense that satisfying either one of the aims will result in automatic satisfaction of the other. But not so. To see why, we need to say a bit more about these aims and how they are related.

Take the practical aim first. Sidgwick defines a method of ethics as a *rational procedure* for coming to beliefs about the deontic status of actions. One way to think about such a procedure is that in following it a person will be led to form correct (or at least justified) moral views about the rightness or wrongness of particular concrete actions or perhaps action types. Here the focus is on finding a procedure that yields a certain output, however the procedure might do it, so long as it is both useful and reliable. There are two things of primary importance here. First, there must be some reliable connection between the procedure and the truth about the deontic status of actions and, second, the procedure must be something that is useful (so it can't place demands on agents that outstrip their ordinary capacities of human deliberation).⁵

As for the theoretical aim, the idea is that ethical inquiry seeks to uncover the underlying nature, so to speak, of right and wrong actions. As it is typically put these days, the properties of being right and of being wrong supervene on other properties of actions – the subvening properties being the more basic ones that ultimately explain why a certain action is right or wrong. The subvening properties, then, represent the ultimate right or wrong making properties of actions. Now one guiding assumption of much moral theorizing is that underlying all right and wrong action is a single feature, or perhaps small set of them, whose presence makes right actions right and whose absence makes wrong actions wrong. Knowl-

³ Henry Sidgwick, *The Methods of Ethics*, 7th ed., New York: Macmillan, 1907, p. 1.

⁴ *Ibid.*, p. 1.

⁵ There is much more to be said about the sorts of constraints imposed on a truly useful decision procedure, having to do with all sorts of cognitive and practical limitations on human beings, which we can't go into here.

edge of just what those features are would provide us with the sort of systematic and *general* knowledge that Sidgwick mentions in connection with the theoretical aim of ethics. Sidgwick calls such features *ultimate reasons* for the deontic status of actions, and assuming there are such reasons, moral theory seeks to discover those which could then be expressed as basic moral principles – very general moral claims that connect a moral property such as rightness with whatever morally relevant nonmoral property or feature it is that makes an action right. Such principles purport to express what Sidgwick calls a *standard* or *criterion* of right action.⁶

Now it is quite common to suppose that having discovered a basic moral principle of the sort just described, one would automatically have a rational decision procedure. The procedure is to take the moral principle which specifies that some nonmoral property P is the ultimate right-making feature of actions, investigate whatever concrete action is the object of moral assessment to see whether or not it has the relevant property P, and then draw a conclusion about the deontic status of the action. This is usually called the 'subsumptive model' of moral reasoning. So, given the dual aim of moral theory, it is sometimes taken for granted that such a theory must provide a single moral principle (or small set of them) that both (1) expresses the ultimate right/wrong-making property of actions and (2) is useful as a decision procedure. The difficulty of doing just this has been the basis of attacks on particular moral theories and indeed on the whole project of moral theory so conceived.

But it is just not true, as many moral theorists have pointed out, that a moral theory needs to accomplish its dual aim in the manner just indicated. Two points are worth making here. First, the connection between moral principles expressing 'ultimate reasons' – moral criteria – and decision procedures can be more or less direct. Second, objections to moral theory, or to some particular moral theory, may misfire if one fails to recognize the first point. Let us take these points in order.

Sidgwick, somewhat obscurely, distinguished between methods that are 'logically connected' with criteria and methods that are connected to criteria only by means of some 'plausible assumption'.⁷ The distinction here can perhaps be expressed by saying that some methods bear a direct connection to some moral criterion in the sense that the method has one make determinations of rightness and wrongness based on what, intuitively, is a direct determination of the presence or absence of the ultimate right-making feature that functions as a moral criterion. Thus, consider the sort of method or decision procedure directly connected with an act utilitarian principle. The principle (as usually interpreted) specifies that facts about the utilities of actions are the fundamental right and wrong-making features

⁶ For Sidgwick's discussion of the relation between methods and principles, see *Sidgwick* (op.cit. fn. 3), Book I, ch. VI.

⁷ *Ibid.*, pp. 78 and 83 ff. For a discussion of these matters in Sidgwick, see *J.B. Schneewind*, *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Oxford University Press, 1977, pp. 194–98.

of actions and, if we interpret utility hedonistically, then the principle specifies that facts about the production of states of pleasure and pain constitute the fundamental right and wrong-making features of actions. The method directly associated with this sort of act utilitarian principle is one that would have us calculate the utilities of various actions in order to come to a judgment about the overall rightness of the actions being considered. But suppose that someone were to believe, based on good evidence, that given human limitations in doing direct utility calculation, one is more likely to perform utility maximizing actions if one 'minds one's own business' and simply attempts to maximize one's own well-being. Here, the method being recommended bears an indirect relation to the utilitarian principle, related, as Sidgwick would say, by a 'plausible assumption.' The criterion of rightness is utilitarian, the method is egoistic.⁸

Now, there are numerous objections to act utilitarianism that focus on its impracticality. The objections in question (1) construe the theory as giving us both an ultimate criterion of right action and a decision procedure, and take the decision procedure to be one that is directly related to the principle in the sense just explained and then (2) urge rejection of the theory either because as a decision procedure it makes practically impossible demands on deliberators, or that as a method it involves some sort of incoherence.⁹ In defense, some utilitarians have wanted to insist that they propose their theory as primarily an account of right-making features and so as primarily a criterion of rightness, not a decision procedure.¹⁰ A utilitarian decision procedure (one directly related to the principle of utility) would have us act from the motive of maximizing utility, but Sidgwick reminds us:

Finally, the doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is the only right or always best *motive* of action. For, as we have before observed, it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on Utilitarian principles.¹¹

⁸ My example here is simple minded. In *The Methods of Ethics* (op.cit. fn. 3), Sidgwick devotes Book II to an elaborate discussion of the various direct and indirect methods associated with the principle of hedonistic egoism.

⁹ For a discussion of these (and other) such objections, see *R. Eugene Bales*, "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8 1971, pp. 257 - 65.

¹⁰ In the recent literature, *David Brink*, *Moral Realism and The Foundations of Ethics*, Cambridge, MA: Cambridge University Press, 1989, ch. 8, makes this move in defense of a version of act utilitarianism. But as *Owen Flanagan*, *Varieties of Moral Personality*, Cambridge, MA: Harvard University Press, 1991, p. 35, points out, granted "The motivational structure required of agents by some moral theory cannot in every case simply be read off its criterion of rightness. Nevertheless, every moral conception owes us at least a partial specification of the personality and motivational structure it expects of morally mature individuals ..."

It is important, then, when examining a moral theory of right action to keep the main practical and theoretical aims of moral theory distinct and not assume that the moral principles featured in a theory give us both a decision procedure and a criterion. So with this lesson in mind, let us turn to Kant's moral theory.

II. The Universal Law Formulation as a Decision Procedure

In the second chapter of the *Grundlegung*, Kant gives us a number of ways of formulating the supreme principle of morality, most notably, there is the Universal Law formulation, the Humanity-as-an-End-in-Itself formulation, and the Kingdom of Ends formulation. In some sense, they are all supposed to be equivalent, though the precise nature of their equivalence (assuming there is one) is not clear from what Kant says. Space does not permit me to discuss the roles of each of these formulations and examine how they might be related, rather, as I mentioned at the outset, I will focus on the Universal Law and Humanity formulations, beginning with the former.¹²

I think that it is fairly safe to say that the Universal Law formulation of the Categorical Imperative represents a decision procedure for coming to moral conclusions about the deontic status of actions.¹³ In the second section of the *Grundlegung*, having just gone through the four examples in connection with the Universal Law formulation, Kant summarizes this 'canon' for all moral judgment which clearly involves a procedure to follow in coming to conclusions about the deontic status of actions:

We must *be able to will* that a maxim of our action should become a universal law – this is the general canon for all moral judgment of action. Some actions are so constituted that their maxim cannot even be *conceived* as a universal law of nature without contradiction, let alone be *willed as* what *ought* to become one. In the case of others we do not find this inner impossibility, but it is still impossible to *will* that their maxim should be raised to the universality of law of nature, because such a will would contradict itself.¹⁴

¹¹ *The Methods of Ethics* (op.cit. fn. 3), p. 413.

¹² I shall ignore questions about how many distinct formulations there are of Kant's supreme moral principle. I shall also ignore the differences between the original Universal Law formula and its 'typic', the Universal Law of Nature formula – the latter being important for understanding the details of the sort of decision procedure associated with the Universal Law formula.

¹³ In *Die Metaphysik der Sitten* VI: 225, 51, Kant is explicit about this. He remarks that the Universal Law formulation requires that agents subject their maxims to "the *test* of conceiving yourself as also giving universal law through it [one's maxim]." My emphasis. All of my references to Kant's works cite the volume and page numbers of the Prussian Akademie edition of his works, followed by the page numbers of the English translations. English edition used: *The Metaphysics of Morals*, Mary Gregor, trans., Cambridge: Cambridge University Press, 1991.

In the secondary literature, the two universality tests of one's maxim that Kant is here describing are labeled the 'contradiction in conception' and 'contradiction in the will' tests respectively,¹⁵ and commentators have spent considerable effort trying to spell out just how these tests are supposed to go – what precise procedure one is being told to follow in coming to conclusions about the deontic status of actions.

Again, in the first section of the *Grundlegung* at 402, 70 Kant has just introduced the Categorical Imperative for the first time which he expresses in terms of the requirement that one's maxim be universalizable. He then writes:

In studying the moral knowledge of ordinary human reason we have now arrived at its first principle. This principle it admittedly does not conceive thus abstractly in its universal form; but it does always have it actually before its eyes and does use it as a norm of judgment. It would be easy to show here how human reason, with this compass in hand, is well able to distinguish, in all cases that present themselves, what is good or evil, right or wrong – provided that, without the least attempt to teach it anything new, we merely make reason attend, as Socrates did, to its own principle; and how in consequence there is no need of science or philosophy for knowing what man has to do in order to be honest and good, and indeed to be wise and virtuous.¹⁶

Here, Kant is fairly optimistic about the power of moral judgment possessed by ordinary folks, and a few lines after the quoted passage he remarks that "we cannot observe without admiration the great advantage which the power of practical judgment has over that of theoretical in the minds of ordinary men."¹⁷ One main aim of a philosophical treatise on morality, as Kant sees it, is to safeguard common sense moral thinking against skeptical attacks, and to do this, one needs to expose the underlying principle or 'norm of judgment' implicit in common sense moral thinking (at least when such thinking is done properly) and then justify that principle.¹⁸ Presumably, this 'norm of judgment' will bear some important and illuminating relation to whatever features are the ultimate right-making features of actions, but (heeding the lesson learned in the previous section) we need not assume that in order to fulfill its function as a basic norm for rational decision making, this norm must itself express such features. As we shall see below, on the sort of reading I am defending, facts about the universalizability of one's maxims are not the features in virtue of which actions are objectively right or wrong, though such facts

¹⁴ *Grundlegung zur Metaphysik der Sitten* IV: 424, 91. English translation used: *The Groundwork of the Metaphysics of Morals* (third edition), H.J. Paton, trans., New York: Harper & Row, 1965.

¹⁵ I believe *Onora O'Neill* (*Nell*, when the book was published) in *Acting on Principle*, New York/London: Columbia University Press, 1975, first introduced these terms which are now commonly used.

¹⁶ *Grundlegung zur Metaphysik der Sitten* IV: 403 - 04, 71 - 2.

¹⁷ *Ibid.*, 404, 72.

¹⁸ At the end of the Preface to the *Grundlegung* at 392, 60, Kant writes: "The sole aim of the present *Groundwork* is to seek out and establish *the supreme principle of morality*."

are correlated with what, on Kant's view, are the ultimate right-making features of actions.

Now all sorts of questions about the Universal Law formulation as a decision procedure can be raised. Obviously, there are questions about how the two tests associated with this formulation are supposed to generate contradictions, and whether, assuming they can generate contradictions of some sort, they are reliable tests for determining the deontic status of actions. In addition, there are questions about how exactly to understand the role of the Universal Law formulation as a decision procedure.¹⁹ For the most part, I don't plan to add to the already large literature addressing these questions. I am here interested in coming to a clearer understanding of the roles of the various formulations of the Categorical Imperative in Kant's moral philosophy, including how the various formulations are related.

III. The Categorical Imperative Deflated: O'Neill's Challenge

One question important for our concerns has to do with the status of any moral conclusions that can be derived from the use of Kant's universality tests. Normally, moral theory is understood to be the search for principles that will help us discover and better understand the *objective rightness* of actions. Talk of the objective rightness of an action, as opposed to its *subjective rightness*, can perhaps be best understood in terms of perspective. As I am using the notion, the objective rightness (and in general, the deontic status of an action, objectively considered) is independent of anyone's conception or beliefs about the morality of the action; an action is objectively right if it really is right. And presumably, whether or not an action really is right depends upon objective features of the action and not on the beliefs or attitudes of individuals or groups. Here, the perspective is an appropriately idealized one.²⁰ The subjective rightness of an action, by contrast, has to do with the first-person perspective of agents. And here it is useful to draw a distinction. Let

¹⁹ See *Barbara Herman*, "Moral Deliberation and the Derivation of Duties," in *her* *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press, 1993. In this essay, she considers what she calls the 'derivation-of-duties' and 'moral-deliberation' models for understanding how the Universal Law formulation plays a role in moral deliberation. She finds reasons to reject these models and proposes a model that might fairly be called the 'derivation-of-presumptive-duties' model.

²⁰ Those who maintain a robust conception of objectivity will appeal to a 'God's eye perspective' in articulating the objective point of view. *Sidgwick* (*The Methods of Ethics*, op.cit. fn. 3), p. 382, refers to the 'point of view of the universe' in expressing the sort of God's eye perspective from which the good of every individual can be understood to be on a par. Other philosophers, who accept a less robust conception of objectivity, articulate the relevant perspective in terms of some idealized human point of view. This tradition includes Adam Smith, who appeals to the device of an 'impartial spectator' and includes among more recent writers, R.M. Hare, R. Firth, J. Rawls, and R. Brandt.

us say that an action is *weakly subjectively right* just in case the agent honestly believes that the action in question is objectively right, and let us say that an action is *strongly subjectively right* just in case the agent's belief that the action is objectively right is a belief that is rationally held. (For simplicity's sake, I am ignoring the obvious temporal indexing of these two notions of subjective rightness.)

It has been standard to interpret Kant's Universal Law formulation as giving us tests for determining the objective rightness of actions, though some have wanted to argue that all that Kant's tests give us is a way of determining the subjective rightness of actions. Since whether or not Kant's Universal Law formulation can yield conclusions of objective rightness or only conclusions of subjective rightness is important for getting clear about its role in Kant's moral theory, we need to explore this issue.

This particular issue regarding the status of moral conclusions reached through the application of the Universal Law formulation is tied to the problem of relevant descriptions. This problem is one facing any universality test in ethics.²¹ Generically speaking, a universality test for determining the deontic status of actions is one that proceeds by subsuming particular actions up for moral assessment under a general description of the action. But any action admits of a vast multitude of true descriptions, and applying a universality test to some action will yield different conclusions about the deontic status of the action depending on which description is used in characterizing the action. Thus, unless there is some principled way of specifying which features of an action are morally relevant and thus should be mentioned in a description of the action, a universality test can be made to yield inconsistent deontic conclusions about concrete actions. The problem of relevant descriptions (in so far as it pertains to the use of universality tests), then, is to provide a criterion for determining which features of an action should be mentioned in a description of the action for purposes of moral evaluation.

This problem, as it relates to Kant's universality tests, has to do with the following considerations: (1) employing Kant's universality tests requires that one formulate a maxim upon which the action is to be performed (assuming for convenience that we are evaluating an action that has not yet been performed); (2) in formulating a maxim, one describes one's action and circumstances; (3) there are numerous ways in which one can describe some concrete action and the circumstances in which it is to be performed; and so (4) a maxim may include features that are not morally relevant and may fail to include all of the morally relevant features that would be required in order to come to a correct assessment of the objective deontic status of the action in question (assuming, of course, that Kant's tests are reliable

²¹ I am following *O'Neill* (op.cit. fn. 15) in labeling such views as involving universality tests. Of course, such views can also be construed as providing criteria of right action, and not tests or decision procedures, though, as I have already mentioned, a principle representing some theory of right conduct is typically taken as being a criterion that can be used as a decision procedure.

when used on 'correct' maxims). Kant tells us that maxims are 'subjective principles' of action – principles that may involve distortion and error since they are formed "in accordance with conditions of the subject (often his ignorance or again his inclinations)."²² The problem, then, is one of specifying which descriptions of one's action and circumstances are relevant for purposes of moral evaluation and which of them, consequently, should be reflected in the formulation of one's maxim. What Kant needs, then, in order to solve this problem of relevant descriptions, is some principled account of those features of one's circumstances and actions that bear on the objective deontic status of actions and which thus ought to be included in one's maxim. Barring a solution to this problem, Kant's universality tests can't deliver, or so it is claimed.

Among Kant's interpreters, we find both pessimists and optimists regarding the prospects for such a solution. Pessimists despair of any solution to this problem in Kant's ethics, though they sometimes disagree about the implications for Kant's moral theory. One kind of deep pessimist would conclude that Kant's moral theory, without a solution to the problem is hopeless. Others think that there being no solution to the problem only shows that the Universal Law formulation is hopeless, but that other formulations of the moral law don't face the problem and so Kant's theory can get along without a solution to the problem.²³ Optimists, like myself, think there is solution to the problem. In fact, I am a happy-go-lucky optimist since I don't think the problem is very much of a problem in Kant's ethics and that a pretty obvious solution has been overlooked (strangely enough) by critics and interpreters. But more of this later.

For now, I want to consider an interesting in-between interpretation of Kant's Universal Law formula according to which Kant does have a way of determining which descriptions of one's situation and action are relevant for employing the universality tests, but the solution in question is not a solution to the problem of relevant descriptions as posed: it is not a solution to the problem of objective relevant descriptions. The idea is that in any case of moral decision making, the complete morally relevant description of an agent's action and circumstances for purposes of using the universality tests just is whatever happens to be reflected in one's maxim. This is Onora O'Neill's 1975 solution to the problem of relevant descriptions.²⁴ She writes:

²² Grundlegung zur Metaphysik der Sitten IV: 421, 88.

²³ Thomas Hill is someone who accepts this milder form of pessimism. See *Thomas E. Hill, Jr., Dignity and Practical Reason*, Ithaca, NY: Cornell University Press, 1992, especially essays 2 and 4.

²⁴ However, there are places in *Acting on Principle* (op.cit. fn. 15), ch. 5, where O'Neill suggests that agents ought to do what they can to make sure that they propose and act on maxims that avoid self-deception, ignorance and bias which would tend to distort one's conception of one's action and circumstances. She suggests, then, that maxims may be more or less appropriate for purposes of applying Kant's universality tests. But it is not clear that this recommendation is consistent with her pessimism about the prospects for giving an account

Kant's universality test includes an explicit solution to the problem of relevant descriptions. The principle or instantiation of a principle whose universality it is relevant to test is the one on which the agent acts or proposes to act on a given occasion. This principle or instantiation is the agent's maxim. It is Kant's contention that when an act is morally acceptable . . . the agent's maxim must be a principle of a particular sort. The Categorical Imperative both states a condition on principles and also instructs us as to which principle or instantiation of a principle we should assess in any given context. It incorporates a solution to the problem of relevant descriptions.²⁵

Of course, since the actual maxims upon which agents act can involve distortion and error that comes from self-deception, ignorance and biases of various sorts, O'Neill is forced to conclude that Kant's tests can only be counted upon to give us conclusions about the subjective rightness of actions. She distinguishes 'contexts of action' which represent the first person perspective of the deliberating agent, from 'contexts of assessment' that view actions from an idealized perspective. While admitting that since maxims may not, from an idealized perspective, accurately represent the agent's situation and action, and thus that Kant's tests may not be relied upon to yield correct judgments of the objective deontic status of actions, still, in contexts of action,

Kant's theory does provide adequate guidance. Agents cannot do better than to act on carefully vetted maxims. If their act conforms to a maxim which can be shown to be obligatory or permissible [by the universality tests] . . . then they can do no more to ensure that their act is obligatory or permissible. . . . Kant's theory of right is not in fact open to the sorts of counter example which arise from agents' ignorance, bias, or self-deception. But it is a theory which tells us how to decide rightly rather than whether actions are right.²⁶

Her talk of deciding rightly is equivalent to talk of an action's being subjectively right, and so on her interpretation, Kant's Universal Law formulation gives us a decision procedure for determining whether or not an action is subjectively right. And since the tests represent what Sidgwick would call a rational method of ethics,

of objective moral relevance since it looks as if the recommendation requires just such an account. For instance, not all sorts of ignorance about my action (e.g., about the neurophysiological processes in me that are involved in its production) are normally relevant in coming to a correct deontic assessment of some action of mine; I need only worry about morally relevant information. But then it seems that one needs some account of objective moral relevance in order to make judgments about matters of ignorance as they bear on the deontic status of actions.

²⁵ Onora O'Neill (op.cit. fn. 15), p. 13. More recently, O'Neill has offered an apparently different solution to the problem, according to which (roughly), maxims are to be understood as *fundamental* intentions that include comparatively general descriptions of one's action and circumstances that then guide agents to adopt more specific maxims in specific contexts. See O'Neill, "Consistency in Action," in *her* *Constructions of Reason*, New York: Cambridge University Press, 1989. One problem with this proposal is that maxims may involve descriptions that are more or less general, and it isn't clear just how to fix on a maxim of the appropriate generality.

²⁶ Onora O'Neill (op.cit. fn. 15), p. 129.

they can presumably generate correct conclusions about the strong subjective rightness of actions – conclusions about what an agent is rational in believing to be right, even if, in fact, the act is not right.

O'Neill admits that theories of right action typically aim to give us a story about the objective deontic status of actions, and so she admits that restricting Kant's universality tests to only giving us a decision procedure for determining the subjective deontic status of actions is quite a retreat from what the theory seems to promise. But after surveying a few proposals for solving the problem of objective descriptions and finding none of them acceptable, she concludes that a theory of subjectively right action is the best Kant can do.

O'Neill's interpretation is quite useful for the sort of challenge it sets for anyone who would want to get more from Kant's universality tests than her interpretation allows. The challenge is to come up with, on Kant's behalf, what we can call a theory of objective relevance that will specify which sorts of descriptions of one's action and circumstance are morally relevant and so must be included in one's maxim if the universality tests are to yield reliable conclusions about the objective deontic status of actions. And here is where I see the Humanity formulation coming into the picture.

IV. The Humanity Formulation as a Criterion of Right Action

I see the Humanity formulation of the Categorical Imperative as providing us with a 'deep' account of the deontic status of actions – it represents Kant's theory about those ultimate features in virtue of which an action is right or wrong. The basic idea is simply this. On Kant's view about the deontic status of actions, facts about how a concrete action, were it performed, would affect 'humanity' or what he calls 'personality' (whether in others or one's self) constitute ultimate morally relevant features of actions – those features in virtue of which actions are right or wrong. The various duties to oneself and to others that Kant lays out in Part II of *Die Metaphysik der Sitten* (the *Tugendlehre*) are by and large duties that he derives by considering the bearing of certain types of action on facts about the maintenance and promotion of humanity. I maintain that this scheme of duties constitutes a (partial) account or theory of relevant descriptions. But before going into the details here, we should first pause to consider the problem of relevance in a bit more detail, and then we can turn to my reading of Kant's Humanity formulation of the Categorical Imperative.

1. Remarks about Theories of Moral Relevance

As noted in the previous section, the problem of relevant descriptions poses a challenge to a certain type of moral theory, viz., those involving the complication of assessing the deontic status of concrete actions under some description. Some

writers have pointed out that this sort of problem does not apply to all moral theories. Thus, for instance, in his discussion of utilitarian generalization and various forms of rule utilitarianism, David Lyons remarks that “this sort of complication does not arise in connection with simple [act] utilitarianism, for the simple utility of an act is not description-relative.”²⁷ Since the problem of relevant descriptions now under discussion apparently only presents a problem for certain moral theories, I’m going to call it the ‘special problem of relevance.’

What I will call the ‘general problem of moral relevance’ does not, unlike the special problem, pertain to some special class of moral theories like Kant’s. It can be posed as the question of what features of an action (and one’s circumstances) are the fundamental morally relevant features in virtue of which an action is right or wrong.²⁸

My reason for distinguishing the two problems, in addition to simply getting clear about the problem of moral relevance generally, is that so far as I can tell, a solution to the special problem of relevance (regardless of the particular moral theory for which we are seeking a solution) really depends on the sort of solution to the general problem that the theory in question represents. What I mean is that a moral theory is in the business of giving a reasoned answer to the general problem of relevance, and those moral theories that must deal with the special problem of moral relevance will be able to do so only by appeal to the answer they give to the general question. At least, this is what I will be claiming in connection with Kant’s moral theory. If these remarks strike the reader as cryptic, let me try to clear things up by considering in more detail the general problem of relevance.

Let us first consider two broad approaches we find among philosophers who think that it is possible to give some sort of general account or theory of moral relevance.

One approach would be to specify, independently of any normative moral theory, features that are morally relevant, or at least rule some out as irrelevant. Some philosophers have claimed that there are broadly ‘logical’ (including semantic) constraints on moral terms and moral judgment that serve as a basis, independently

²⁷ David Lyons, *Forms and Limits of Utilitarianism*, Oxford: Oxford University Press, 1965, p. 35.

²⁸ Posing the problem in terms of this question is not quite right, since it presupposes that there are some (at least one) fundamental features of the sort in question that can be captured by some set of moral generalizations (principles). However, so-called ‘particularists’ in ethics would reject this presupposition, and so would a kind of moral skeptic who rejects moral predications as mistaken. But the general problem of relevance can be posed in a non-question begging manner in terms of two questions: (1) What features of an action (if any) are relevant in determining the deontic status of actions? and (2) Is it possible to give a general account or theory of moral relevance that would serve to codify such features? The sort of radical moral skeptic just mentioned would deny that there are any features of the sort mentioned in (1), while particularists would answer (2) in the negative. For simplicity, I leave these complications out of the text.

of commitment to any normative theory, for specifying what is and what is not relevant, either wholly or at least in part. For instance, appealing to the principle that ‘ought’ implies ‘can’, we can rule out as irrelevant for purposes of moral appraisal facts about ‘options’ which, in my circumstances, were not open to me. The fact that the drowning child could have been saved by some able nearby person is not relevant in evaluating my not doing so, if, though nearby, I am confined to a wheelchair. Again, some have argued, on formal grounds, that moral evaluations must be based on descriptions in non-indexical, universal terms. Proper names, for instance, are sometimes thought to be irrelevant for purposes of moral evaluation since (roughly) an action of mine being right or wrong does not depend the fact that it was performed by Mark Timmons even though descriptions using ‘universal terms’ that in fact serve to pick me out might be relevant. Of course, even if one accepts these remarks about relevance, such restrictions do not go very far toward giving us a full account of relevance.

Some have wanted to get a more robust account of relevance by appealing to the meaning of ‘moral’ or perhaps the meanings of moral terms, as a basis for claiming that certain features of actions cannot be relevant and that others must be. Philippa Foot,²⁹ for instance, gives the example of clasping and unclasping one’s hands as a kind of fact about an action that cannot be relevant, and G. J. Warnock³⁰ claims that facts about production of benefits and harms are always relevant in moral evaluation. But I agree with Hare³¹ that these claims about relevance, even if they are correct, do not just depend on the meaning of ‘moral’ or on the meanings of moral terms, but rely on substantive moral claims. What this means is that although there may be some theory-independent restrictions on moral relevance (of the sort indicated in the last paragraph), any robust account of moral relevance of a sort needed to yield determinate moral judgments about the deontic status of actions will be theory-dependent.

This conclusion should come as no surprise when it is recalled that one main aim of a moral theory of right conduct (the theoretical aim) is to specify those ultimate features of actions in virtue of which actions are right or wrong. Moral principles expressing such features just are principles of moral relevance. Hare puts the point this way:

It is a great mistake to think that there can be a morally or evaluatively neutral process of picking out relevant features of a situation, which can then be followed by the job of appraising or evaluating the situation morally. We can indeed describe a situation without committing ourselves to any moral judgments about it . . . nevertheless, when we decide what features of the description are morally relevant, we are already in the moral business. There are exceptions to this rule; some features of situations can be ruled out as irrelevant on purely formal grounds, as we have seen. But in the main to call a feature

²⁹ Philippa Foot, “Moral Beliefs,” *Proceedings of the Aristotelian Society* 59, 1958/9.

³⁰ G.J. Warnock, *Contemporary Moral Philosophy*, London: Macmillan, 1967, p. 67.

³¹ R.M. Hare, “Relevance,” from *Values and Morals*, A.I. Goldman/J. Kim, eds: Dordrecht: Reidel, 1978. Reprinted in *Essays in Moral Theory*, Oxford: Clarendon Press, 1989.

morally relevant is already to imply that it is a reason for or against making some moral judgment; and to say this is already to invoke a moral principle.³²

Returning now to Kant's moral theory, testing the morality of actions by the procedures involved in the Universal Law formulation of the Categorical Imperative faces what I have called the special problem of relevance. What I am suggesting is that if we are looking for a solution to the special problem, we ought to be looking at the sorts of moral principles featured in his theory which would furnish a solution to the general problem of moral relevance. And, if we are interested in the objective rightness of actions, and if the objective deontic status of actions depends on features of those actions, then it seems (to me anyway) that we ought to look to the Humanity formulation for an account of moral relevance.

2. *Humanity as an End in Itself*

Doing so requires that we say something about the concept of humanity as an end in itself, though space does not permit anything approaching a fully adequate analysis of this very rich concept and its role in Kant's moral philosophy.³³ I will confine myself to a few brief remarks.

First, the idea of something being an end in itself is, for Kant, the idea of what he calls in the *Grundlegung* an 'objective end.'³⁴ Roughly, to say that something has this status is to say that it is something (some state of affairs) that morally ought to be, either by being maintained and promoted if the state of affairs already exists, or brought about if it does not already exist. Now according to one sort of analysis of 'ought' statements in Kant, they are equivalent to statements about the activities of fully rational agents.³⁵ Statements expressing what one morally ought to do in some

³² Ibid., p. 193. David Lyons agrees with Hare. In *Forms and Limits of Utilitarianism* (op.cit. fn. 27), he considers the proposal that "questions of relevance for a given substantive principle are settled by appeal to general, independent moral criteria or principles." He responds by noting that "This approach is methodologically unsound and leads to a dead end. After all, how does one determine the criteria of moral relevance in general? To what substantive criterion might one appeal? It would seem that relevance should and can be determined by reference to the nature and content of the substantive principles in question" (pp. 101 - 02).

³³ For more detail on this concept in Kant's ethics, see *Thomas E. Hill, Jr.*, "Humanity as an End in Itself," *Ethics* 91, 1980, pp 84 - 90. Reprinted in his *Dignity and Practical Reason* (op.cit. fn. 16).

³⁴ To classify some state of affairs or object as an objective end is to indicate that it is something that is unconditionally valid for all rational agents to bring about or maintain. In relation to imperfectly rational agents, such as human beings, whose volitions are not necessarily in accord with reason, such ends are properly called 'obligatory ends', which is what Kant calls them in the *Metaphysik der Sitten*.

³⁵ For an attempt to work out in some detail an analysis along these lines, see *Mark Timmons*, "Necessitation and Justification in Kant's Ethics," *Canadian Journal of Philosophy* 22, 1992, pp. 223 - 261.

set of circumstances are thus equivalent to statements about what all fully rational agents would do in those circumstances, and so statements about what ought to be, which express judgments about what is intrinsically valuable, are equivalent to statements about what all fully rational agents would necessarily value as such.

Although I can't argue for the claim here, it is plausible to suppose that one object of necessary value for all fully rational agents is their own rational capacities or what Kant refers to as their 'rational natures.'³⁶ Part of what is involved in having a rational nature is the capacity for acting on the basis of reasons which in turn involves, at a minimum, the capacity to deliberate and choose goals that one thinks are worthy of pursuit. Kant's way of putting this idea is to say that part of what is distinctive of rational agents is their capacity to set ends. But being capable of setting ends and acting upon them is compatible with the claim that all such ends (and thus all reasons for action) have their source in the various desires and aversions one has. However, Kant thinks that humans are agents in a much stronger sense than this: humans, qua rational agents, are capable of autonomy of the will which, he says, "is that property the will has of being a law to itself (independently of every property belonging to objects of volition)."³⁷ So one object of necessary value for all fully rational agents is their autonomy. Moreover, it is plausible to suppose that fully rational agents would necessarily value those various tendencies and capacities that are intimately tied to the maintenance and promotion of autonomy. The sorts of tendencies and capacities will depend on the sort of rational agent under consideration. Kant claims that basic moral requirements hold for all rational (not just human) agents. So in determining what all is involved in valuing the autonomy of some type of rational agent, we need to consider the specific nature of that type of agent. When it comes to human beings we find that certain capacities pertain to our brute animal nature, others pertaining to our natures as social creatures. These human capacities are represented in chapter 1 of the *Religion* as innate features of human beings; they are part of what Kant calls 'the original predisposition to good in human nature.' He writes:

We may conveniently divide this predisposition, with respect to function, into three divisions, to be considered as elements in the fixed character and destiny of man:

- (1) The predisposition to *animality* in man, taken as a *living* being;
- (2) The predisposition to *humanity* in man, taken as a living and at the same time *rational* being;
- (3) The predisposition to *personality* in man, taken as rational and at the same time an *accountable* being.³⁸

³⁶ For a recent attempt to base a moral principle on the necessary objects of rational volition, see *Alan Gewirth*, *Reason and Morality*, Chicago: The University of Chicago Press, 1977.

³⁷ *Grundlegung zur Metaphysik der Sitten*, 440, 108.

³⁸ *Die Religion innerhalb der Grenzen der bloßen Vernunft* VI: 26, 21. English edition used: *Religion Within the Limits of Reason Alone*, Theodore M. Greene/Hoyt H. Hudson, trans., New York: Harper & Row, 1960.

In the paragraphs immediately following this one, Kant elaborates each of these divisions. The predisposition to animality concerns those innate tendencies to self-preservation, preservation of the species, and community with others – tendencies that are opposed by ‘beastly vices’, including gluttony, drunkenness, and lasciviousness. This aspect of our predisposition to good is at the root of the various duties to oneself considered in light of our animal natures that Kant elaborates in the *Tugendlehre*, Part I, chapter 1. The predisposition to humanity is explained as pertaining to our comparative judgments of well-being and the desire for equality with all others. It is opposed by ‘vices of culture’, including envy, ingratitude, and malice – vices featured in Kant’s *Tugendlehre* discussion of duties of respect to others (Part II, chapter 1). Finally, the predisposition to personality (which Kant says is ‘the idea of humanity considered quite intellectually’)³⁹ is described as the capacity for respect for the moral law being a sufficient motive of the will – which is a necessary presupposition of being bound by moral requirements as well as being capable of moral goodness through coming to have a good character (*Gesinnung*). Again, this division of the original predisposition is the basis for duties of virtue to oneself as a moral being that we find in Part I, chapter 2 of the *Tugendlehre*.

If we use the term ‘humanity’ to cover all aspects of these various innate tendencies and capacities in human nature,⁴⁰ the idea is that all fully rational human agents would necessarily value the various ingredients mentioned above that are innate in human nature, presumably because they are, for human beings at least, intimately connected to autonomy, understood negatively as a capacity to act for reasons independently of inclinations and desires, and positively as being capable of acting for unconditionally valid reasons. Because humanity is necessarily valued by all fully rational agents, it has the status of an objective end – one set by reason, not by inclination, as Kant would say. Because it has this status, it can thus serve as a norm of right and wrong conduct.

3. Kant’s Theory of Moral Relevance

If the things I have said about theories of moral relevance and about Kant’s notion of humanity as an end in itself are roughly correct, then we can look to the Humanity formulation of the Categorical Imperative for a Kantian theory of moral relevance. Here is one way (perhaps not the only way) of understanding the theory.

As we have seen, a fundamental moral principle, in its theoretical role, purports to indicate those features of actions that are (or are among) the ultimate right and

³⁹ *Ibid.*, 28, 23.

⁴⁰ This usage seems to be in keeping with how Kant uses the term throughout most of his ethical writings, although in the passage from the Religion quoted above, Kant distinguishes humanity and personality.

wrong-making features of action. In Kant’s moral theory, facts about how an action affects humanity, whether in oneself or others, are what (ultimately) make an action right or wrong. But saying this much is not terribly useful until one illuminates the notion of humanity in question. As I have said, for Kant, it is a complex notion, but in *Die Metaphysik der Sitten* Kant gives us some indication of its content, particularly in the *Tugendlehre*, where he specifies a system of general duties by appealing (in almost all cases)⁴¹ to the idea of humanity as an end in itself. Think of the various derived negative duties (e.g., against suicide and immoderate consumption of food and drink) as indicating types of actions that necessarily interfere in some way with the maintenance and promotion of humanity, and think of the various positive duties (e.g., duties to develop our talents and help others in need) as types of action the principled omission of which necessarily interferes with the maintenance and promotion of humanity. The system of duties, then, can be viewed as a specification (at least in part) of the content of the concept or idea of respecting humanity.

What does this have to do with the issue of relevant descriptions? To consider an action as of a certain type is, in effect, to consider it under some description or other. For many such descriptions, because of their importance in moral and social life, we have single terms that serve to pick out actions under a description.⁴² To classify an action as a case of suicide is short for classifying the action under a more cumbersome description. The same goes for terms like ‘gluttony’, ‘lying’, ‘avarice’, ‘servility’, ‘ingratitude’, ‘malice’, ‘envy’ featured in Kant’s system of duties. And, of course, what this means is that the system of duties outlined in the *Tugendlehre* represents at least part of a principled account of objective moral relevance. And so in cases where some action of mine can be correctly classified as being an instance of one or more of the various actions mentioned in the system of duties, then that fact about it is morally relevant.⁴³

⁴¹ One notable exception is his derivation of the general duty of beneficence. See *Die Metaphysik der Sitten* VI: 393, 196–97.

⁴² This point has been often noted. For instance, *Hilary Putnam*, *Reason, Truth and History*. Cambridge: Cambridge University Press, 1981, notes that “We invent moral words for morally relevant features of situations”, p. 144. Again, *Eric D’Arcy*, *Human Acts*, Oxford: Oxford University Press, 1963, discusses moral ‘case terms’ like ‘murder’, ‘lying’, ‘adultery’, ‘embezzlement’, and so on. He notes that “The essential features of morally significant behaviour and situations constantly recur in recognizable patterns which usually enable the most originally contrived deeds to be subsumed under the traditional case terms, which have been fashioned because of the frequent occurrence of such deeds, and the special importance they have for human welfare”, p. 24.

⁴³ The sorts of features of actions, agents, and their circumstances referred to by terms like ‘suicide’, ‘lying’ and the rest, represent a system of basic morally relevant features – features that are always morally relevant whenever they are instantiated and relevant in the same way. Thus, any time an action constitutes a lie, that fact about it is morally relevant and always counts as a wrong-making feature of the act (though in special cases, this fact about it may not be decisive, or so I would argue in connection with Kant’s theory of right conduct despite his sometimes rigoristic claims to the contrary). Other features of someone’s particular cir-

If what I have been saying is correct, then Kant does, in effect, provide us with an account of moral relevance. Moreover, it is an account that can be used to respond to O'Neill's challenge of coming up with an account of 'objective' moral relevance or else settling for a deflated conception of Kant's Categorical Imperative. In order to use the universalization tests to come to conclusions about the objective deontic status of actions, one's maxim ought to mention all and only those features of one's action and circumstances that are morally relevant for the assessment of the objective deontic status of the action.⁴⁴ There is, of course, much more to say about this view of moral relevance, but some of what there is to say will emerge in the following section taking up various interpretive issues and questions to which I now turn.

V. The Differential Roles Interpretation of the Categorical Imperative

On the differential roles interpretation of the Categorical Imperative I am defending, the various formulations (at least the two we have been considering) play importantly different roles in Kant's theory of right conduct. Of course, given that I have not said anything about the other formulations of the Categorical Imperative, and that what I have said about the Universal Law and Humanity formulations has been brief and somewhat sketchy, I am in no position to confidently advocate this interpretation. What I can do in the space remaining is explore various interpretive issues and questions that mainly concern the relations between the Universal Law and Humanity formulations of Kant's supreme principle of morality if one goes for the differential roles interpretation of these formulations that I have been sketching. After saying more precisely what the differential roles interpretation comes to,⁴⁵ I shall proceed by posing and then answering various questions that seem obviously suggested by what I have been saying.

cumstances on some occasion might be relevant depending on contingent facts about the bearing of such features on humanity as an end. For instance, in Kant's casuistical remark related to the issue of avarice in *Die Metaphysik der Sitten* at 434, 230, he asks "Ought I to economize on food or only in my expenditures on external things? In old age, or already in youth?" One question being raised here concerns the relevance of one's stage of life in applying a maxim of thrift. Considerations pertaining to how old one is and one's life expectancy are morally relevant in this sort of case, even though such considerations may be completely irrelevant in other cases.

⁴⁴ This way of putting things may be too simplistic. Coming to a correct deontic assessment of some concrete action may require a series of applications of Kant's tests to a series of maxims, especially in cases that involve a good deal of complexity. So in addition to providing a criterion of objective relevance, one must also tackle the issue of how one is to use such information in applying Kant's universality tests. (I thank Tom Hill for urging this point on me at the Smith conference.)

⁴⁵ One can, of course, defend a differential roles interpretation of the Categorical Imperative and disagree with the details of my account.

I think it is fairly uncontroversial that the various formulations of the Categorical Imperative are often assumed to play multiple roles in Kant's theory of right conduct: as mentioned at the outset, a common view of the Universal Law formulation is that it is both a decision procedure and a moral criterion.⁴⁶ In opposition to this assumption, I have been defending the two following claims, the first methodological, the second substantive. (1) It proves illuminating to distinguish the various roles a moral principle might play vis-à-vis the main aims of moral theory – I have discussed the main theoretical and practical aims – and not just assume that all formulations of Kant's supreme moral principle play both roles. (2) The Universal Law formulation functions primarily as a decision procedure and the Humanity formulation functions primarily as a moral criterion. To these claims, I want to add a third: (3) The Universal Law formulation does not express a criterion of right action. The basis of this claim is simply that the feature of a maxim's being universalizable is not the same feature as respecting humanity. Since the Humanity formulation does present us with a moral criterion, then assuming that Kant presents us with a single ultimate moral criterion of right action, it follows that the feature of being universalizable is not a moral criterion in the sense in which we have been using this term. This remark, however, prompts our first question.

1. How are the Universal Law and Humanity formulations related?

In the second chapter of the *Grundlegung*, Kant claims that these two principles are "at bottom the same."⁴⁷ I won't be able to thoroughly discuss this claim, but I owe the reader some indication of how these formulations are related on the interpretation I am proposing.

Since the concepts featured in the Universal Law formulation (action on a maxim, willing, and *universal law*) are distinct from those featured in the Humanity formulation (*action, means and ends, and humanity*), it doesn't appear as if there is

⁴⁶ In addition to this *equivalence thesis*, Kant also claims at *Grundlegung* 436, 103-04 that the Universal Law formulation is the form of a good maxim, the Humanity formulation is its matter, while the Kingdom of Ends formulation is the 'complete determination' of a good maxim. This suggests a *non-equivalence thesis*, in opposition to the equivalence thesis, in that from Kant's remarks about form, matter, and complete determination, none of these formulations would appear to be complete expressions of the moral law and so would not be equivalent. I won't be able to discuss the differences between these two theses and the question of how they might be reconciled. (For an illuminating discussion of this matter, see Philip Stratton-Lake, "Formulating Categorical Imperatives," *Kant-Studien* 83, 1993, pp. 317-40.) I should also note that my discussion of the various formulations of the Categorical Imperative has been restricted to consideration of the roles they play in Kant's theory of right action. A thorough examination of their various roles in Kant's ethics would require that we consider other features of his views, including most obviously his theory of moral worth.

⁴⁷ *Grundlegung zur Metaphysik der Sitten* IV: 437, 105.

a conceptual equivalence between the two formulations.⁴⁸ However, even if they are not conceptually equivalent, there may be some tight conceptual connection between them that would reveal them to be (in some sense) at bottom the same. One way to get at this issue, given how I interpret the two formulations, is by asking why the Universal Law formulation (when applied to maxims that reflect all morally relevant features of an agent's circumstances) is a fully reliable (i.e., infallible) decision procedure. After all, on my interpretation, this formulation does not express a criterion of right action; that is what the Humanity formulation does. But then if the Universal Law formulation reliably reveals (when properly applied) the objective deontic status of actions, there must, it seems, be some intimate connection between Kant's concept of universalizability and his concept of humanity as an end in itself. The question is: how are we to understand the connection between the two concepts?

Kant's universality tests can lead an agent to a correct assessment of the objective deontic status of actions by revealing facts about actions that bear on humanity as an end in itself. More precisely, the non-universalizability of a maxim indicates that the associated action is morally forbidden (or at least presumptively forbidden) and it helps reveal what it is about the action that makes it morally wrong. And what, in most general terms, the test reveals (in the case of non-universalizable maxims) is that were the action performed the agent would be acting in a way that fails to take rational agency as a constraint on action. It does this for two reasons. First, the features of one's action and circumstances relevant for deontic assessment just are those that bear on humanity as an end. Second, the universality tests in effect require agents to detach from their personal, desire-based interests and projects and view their action impartially in terms of its bearing on the maintenance and promotion of rational agency. So, in cases where one cannot will that one's maxim serve as a universal law for all rational agents, one is proposing to act in a manner that simply fails in one way or another to take humanity (or rational agency) as a constraint on action. This, at least, is the gist of the connection between the concepts of universalizability and humanity as an end, but it may help to clarify matters if we review a few of Kant's examples of duties.

Consider for instance the lying promise example. Presumably the lying promise maxim fails the contradiction in conception test because the possibility of making a lying promise depends on the stability of the practice of promising which would

⁴⁸ *Onora O'Neill*, "Universal Laws and Ends-in-Themselves", in *her* *Constructions of Reason* (op.cit. fn. 25) argues that contrary to their seeming non-equivalence, the two formulations are conceptually equivalent. *John Atwell* in "Are Kant's First Two Moral Principles Equivalent?" *Journal of the History of Philosophy* 7, 1969, pp. 273-284, also argues that they are equivalent (in the sense that they never yield conflicting deontic conclusions about the same action) based on his claim that both formulations "rest on the same basis, namely, [one's] own willingness to be treated in like manner", p. 283. I won't be able to discuss these views on this matter here. For defense of an interpretation that emphasizes the non-equivalence of the various formulations, see *Stratton-Lake* (op.cit. fn. 46).

be completely undermined were one's maxim to become universal law. When the agent reflects on this fact about her maxim, she must see the act of making a lying promise as a case of taking advantage of other agents by being a 'free rider.' This sort of deception, because of its manipulative nature, constitutes a violation of humanity as an end in itself since it interferes with an agent's autonomy, and so is morally wrong (at least presumptively). What makes the action wrong concerns those facts about it – its manner of manipulation – that constitutes a violation of humanity as an end in itself. One way in which, on Kant's view, the wrongness is revealed is by an application of one of his universality tests.

Now consider the case of helping others in need. Presumably the maxim of intentionally refraining from giving one's help to (at least some of) those in need cannot be consistently willed as a universal law given that as a rational agent one necessarily wills that one be helped in circumstances of need. This fact about one's maxim in effect requires the agent to view her act of refraining here as manifesting a kind of failure to promote what she recognizes to be an end in itself, viz., humanity. This fact about such actions – i.e., the fact that they fail to promote humanity – is what makes them (presumptively) wrong; the main purpose of the universality test is to reveal this fact.

The last few paragraphs have only roughly and rather vaguely indicated how the Universal Law formulation is related to the Humanity formulation; I have not worked things out in the sort of detail needed to defend my view of their relation. I must leave this for another occasion.

2. *Isn't the Universal Law formulation really superfluous on the differential roles interpretation?*

After all, the Humanity formulation presumably gives us a criterion of right action, and it seems to give us a criterion that can be directly applied. In the lying promise case, since we are to describe the case as one of telling a certain kind of lie, and since this kind of action constitutes a *prima facie* way in which one can violate humanity as an end, it seems that we can conclude straight away (barring any overriding considerations pertaining to one's circumstances) that the action in question is wrong.⁴⁹ We don't really need the Universal Law formulation at all, or so it seems.

Since Kant tells us that the two formulations are in some sense the same, and he illustrates how they may be applied to the same cases to yield the same deontic conclusions, there is a sense in which one need not use the Universal Law formulation in coming to justified moral conclusions about actions. Kant himself seems to

⁴⁹ For an example of a contemporary moral theory that rests upon a fundamental moral principle akin to Kant's Humanity formulation, see *Alan Donagan*, *The Theory of Morality*, Chicago: University of Chicago Press, 1976.

think of there being what he calls a “subjectively practical” difference between them, claiming that “it is better if in moral judgment we proceed always in accordance with the strict method and take as our basis the universal law formula of the categorical imperative . . .”⁵⁰ So for practical purposes, Kant recommends use of the Universal Law formulation. Unfortunately, Kant does not elaborate this claim, leaving it open to speculation why he thought the so-called ‘strict method’ is to be preferring in making moral judgments. Here, then, is a suggestion for understanding Kant’s recommendation.

I have been claiming that the Universal Law formulation functions primarily as a decision procedure, which, when correctly applied, reveals to the agent facts about the nature of the action in virtue of which it possesses a certain deontic status. The universality tests associated with this formulation do this by getting the agent to focus on certain morally relevant features of her action from a certain impartial perspective – a perspective in which she detaches from her personal desire-based interests and focuses on considerations of rational agency. This mode of moral thinking is to be recommended because, as Kant points out, human beings have a “disposition to quibble with these strict laws of duty, to throw doubt on their validity or at least on their purity and strictness, and to make them, where possible, more adapted to our wishes and inclinations.”⁵¹ If one thinks of the occasions prompting moral deliberation as often times those in which one is proposing to make an exception for oneself by performing an action against which there is a moral presumption, then the mode of thinking represented by the universality tests, will help reveal any proposed illicit exemptions on the agent’s part.

In addition to cases in which one proposes to make an exception to some presumptive moral rule, there are those cases in which the rigor of the universality tests might prove a more reliable guide to the deontic status of contemplated courses of action in which there is, in Kant’s terminology, “conflicting grounds of obligation.”⁵² In such cases, there are features of the course of action (and hence a description of the action) which provide a reason or ground for taking it to be morally wrong, but there are also features of that same course of action which provide reasons or grounds for taking it to be morally right, even obligatory. A much discussed example is the case of having to lie to a would-be murderer in order to save a life. The descriptions of the act as a lie and the description of it as saving a life are both morally relevant (since both sorts of consideration bear on treating humanity as an end), and so both should be reflected in a maxim to be tested for its being fit for universal law. The relevant maxim will be something like: To lie in circumstances in which I am being asked by a would-be murderer about the whereabouts of an intended innocent victim in order to save the life of the innocent person. Presumably, this maxim can be universalized. The point of the universality

⁵⁰ Grundlegung zur Metaphysik der Sitten IV: 436, 104.

⁵¹ Ibid., 405, 73.

⁵² Metaphysik der Sitten VI: 224, 50.

test in this case is to help reveal whether or not such actions are consistent with treating humanity as an end in itself; it provides for us an appropriate impartial ‘point of view’ from which to make such judgments.

Finally, before moving to the next question, let me remind the reader that on the interpretation of the Categorical Imperative I am defending, I am not claiming that the various formulations of this principle must each be construed as either a decision procedure or as a criterion of right action and never both. Rather, what I have been claiming is that the Universal Law formulation seems best interpreted as setting forth a decision procedure while the Humanity formulation seems best interpreted as primarily a criterion of right action. Saying this does not commit me to denying that the Humanity formulation can be used as a decision procedure, though I have denied that the Universal Law formulation is, strictly speaking, a criterion of right action; the feature of universalizability functions rather as a fully reliable indicator⁵³ of what it is about actions that makes them morally right or wrong.

3. If maxims are to include morally relevant descriptions provided by the humanity formulation, aren't they 'morally loaded'? And if they are, isn't there something objectionably circular about the universality tests?

I completely agree that the sorts of descriptions provided by the Humanity formulation give moral content, so to speak, to maxims.⁵⁴ In general, given that a theory or account of moral relevance generally (and hence an account of morally relevant descriptions) depends on some substantive moral theory, this is to be expected. But I don’t think we should infer from this that there is something objectionably circular about the universality tests. The suspicion that there is circularity here probably stems from the assumption that the Universal Law formulation (and the tests associated with it) is supposed to give us a procedure that, as it were, operates on some morally neutral characterization of one’s action (and circumstances), and yields deontic conclusions about actions being tested. Or, perhaps the

⁵³ Of course, an application of the Universal Law formulation may not lead to a correct deontic assessment of an action under moral scrutiny since, for one thing, the maxim may not reflect all of the morally relevant information about the case. But in arguing that the Universal Law and Humanity formulations are in some sense at bottom the same, I take there to be some conceptually tight connection between the concepts of universalizable maxim and treating humanity as an end in itself such that necessarily a correct application of the universality tests yields an accurate deontic assessment of any action. So, when correctly applied, the tests function as infallible indicators of the deontic status of actions.

⁵⁴ Barbara Herman in “Moral Deliberation and the Derivation of Duties” (op.cit. fn. 19), sketches an account of the role of the Categorical Imperative in which it is important for purposes of moral deliberation that maxims have moral content. “[I]n circumstances where moral deliberation is appropriate, the deliberative conclusion is not arrived at as the product of the employment of the CI procedure on a maxim that itself has no moral content,” p. 145.

idea is that this formulation is supposed to provide an account of relevance as well as serve as a decision procedure. And so the thought is that if the maxims already have moral content, then the universality tests really aren't the sorts of non-question begging tests they are supposed to be.

But raising this circularity worry should only prompt us to examine the sort of role the Universal Law formulation should be understood to play in Kant's moral theory. Here is not the place to launch into what would be a complicated story, a few remarks will have to do.

I don't think the Universal Law formulation, construed as a decision procedure, should be cast in the role of having to operate on maxims without moral content, nor should it be thought of as itself providing an account of relevant descriptions. Maxims of the sort featured in Kant's famous *Grundlegung* examples already have moral content given the manner in which the cases are described, and so there is, going into the procedure, a presumption that actions of a certain sort are morally wrong or morally obligatory.⁵⁵ The tests confirm this of course, which is an important result in Kant's attempt to convince his readers that the Categorical Imperative lies behind common sense moral judgments of various sorts. The universalization procedure is particularly useful, then, to individuals who, owing to uncertainty about the deontic status of some particular action, need to think through the details of the case in order to determine whether or not there are considerations that would justify an assessment of the action that differs from what is presumed about its deontic status. Of course, the sorts of considerations upon which one may legitimately rebut a moral presumption are constrained by the theory of moral relevance provided by the Humanity formulation.⁵⁶

4. On this view the solution to the problem of relevant descriptions is easy; so why has it been overlooked?

I should say that although I know of no one who has made the particular proposal that I am making in connection with the problem of relevant descriptions in Kant's ethics, when I read what some others have written about Kant's universality tests, I detect hints and suggestions about the problem that are much like my own proposal.⁵⁷ But if not entirely overlooked, there are reasons why this solution to

⁵⁵ Nor, would I argue, does this view of maxims in relation to the universality tests violate Kant's idea that the moral law is a formal as opposed to a material practical principle, though I won't pause here to go into this matter.

⁵⁶ Thus, for example, the problem of 'false positives' (i.e., the problem that if we qualify the description of some action with regard to which there is, say, a presumption that it would be wrong to perform, we can fix things so that most any maxim will pass Kant's tests since if everyone were to act on some qualified maxim no contradiction would result) does not arise here, since one is only allowed to qualify one's maxim (i.e., fill it out with more detail) in terms of considerations that are morally relevant.

the problem of relevant descriptions (if it really is a solution) has been largely overlooked which have to do with certain prejudices (if I may call them that) that have guided much interpretive work on Kant's ethics. In the first place, many interpreters have been fascinated with the idea of being able to generate substantive moral conclusions from a meager starting point; the prospect of combining "formality and fertility".⁵⁸ Consequently there has been, perhaps until very recently, much more attention paid to Kant's Universal Law formulation than to his other formulations. There has also been a fascination with finding rigorous decision procedures in ethics; something the Universal Law formulation seems to promise. And, of course, Kant's universality tests are particularly intriguing because one has a sense that certain maxims really cannot be universalized, but it is not exactly clear how one is supposed to generate contradictions in connection with such maxims, and it is fun to try one's hand at working out the details. But in addition to the attractions of the Universal Law formulation, the Humanity formulation has received bad press. Some have argued, for instance, that the Humanity formulation is, to quote one influential interpreter, "... absolutely useless, [and] not a moral criterion at all."⁵⁹ The idea behind this dismissive attitude seems to be that in order to determine what counts as a violation of (or failure to positively promote) humanity as an end in itself, one must consult the universality tests – actions mentioned in maxims that fail one or other of the tests violate humanity because the maxim cannot be willed to be universal law. So, it has looked to many interpreters that somehow all of the real work in Kant's theory of right conduct is done by the Universal Law formulation.

But giving exclusive attention to the Universal Law formulation leaves completely out of account Kant's criterion of right action and thus his account of moral relevance, or so I have claimed. And, of course, I have been arguing that the Humanity formulation is a moral criterion and hence, the basis of an account of moral relevance.

Perhaps another reason why this solution has been largely overlooked is that interpreters of Kant have not understood clearly enough what I have called the

⁵⁷ I am thinking in particular of certain passages in the writings of Barbara Herman on Kant's ethics. See *Barbara Herman, The Practice of Moral Judgment* (op.cit. fn. 19), especially essays 4 and 5.

⁵⁸ This way of putting it is from O'Neill (op.cit. fn. 15), p. 1.

⁵⁹ *M.G. Singer, Generalization in Ethics*, New York: Alfred A. Knopf, 1961. Singer's reason for saying this is that the principle supposedly equivocates between purely 'rational ends' and 'phenomenal ends' and so, after the remark about the principle not being a moral criterion, he goes on to say: "It is an *ad hoc* device, serving only to give rational respectability to a judgment about what ought to be done made antecedent to its application. It is impossible to tell before the issue has already been decided on some other standard which way of acting would be treating someone merely as a means to 'ends' he cannot himself share. The line of reasoning involved is not, as it seems to be, 'This would be treating someone merely as a means; therefore it is wrong.' It is 'This would be wrong; therefore it would be treating someone merely as a means.'" p. 235.

general problem of moral relevance and how a solution to that problem (and hence the problem of relevant descriptions) is theory-dependent. But once it is viewed in this manner, then it is clear that we ought to be looking at the substantive content of Kant's moral theory which, I claim, is provided by the Humanity formulation.

VI. Conclusion

I began with what I took to be two rather intuitive ideas, viz., the Universal Law formulation is, if nothing else, a decision procedure, while the Humanity formulation, employing the rich notion of humanity as an end, gives us a criterion of right action and thereby a solution to the problem of relevant descriptions in Kant's ethics. The paper has been spent trying to work out these two ideas. This idea that the various formulations play importantly different roles in the overall economy of Kant's ethics is what I am calling a differential roles interpretation of the Categorical Imperative. I don't claim that my interpretation is a fix-all for the various objections that have been raised against Kant's theory of right conduct, but I do think it gives us a perspective on certain important elements of his view (or at least a fair reconstruction of his view) that improves our understanding of his overall theory of right conduct. I'm afraid, however, that my treatment of many issues has been sketchy, and I am sure that I haven't dealt with all of the challenges that can be made to such an interpretation.⁶⁰

Zusammenfassung

Eine Moraltheorie hat typischerweise zwei Hauptziele. Sie hat das praktische Ziel, ein Entscheidungsverfahren zu entwickeln, dem die Handelnden folgen sollen, um zu einer richtigen oder gerechtfertigten deontischen Einschätzung ihrer Handlungen zu kommen. Eine Moraltheorie hat auch das theoretische Ziel, ein moralisches Kriterium zu entwickeln, das bestimmt, welche Momente einer Handlung diejenigen sind, die die Handlung zu einer "richtigen" Handlung machen. Moralphilosophen haben festgestellt, daß die grundlegenden moralischen Prinzipien einer Moraltheorie beiden Zielen dienen können. Notwendig ist das freilich nicht. Kants erste Formel des Kategorischen Imperativs, die das Allgemeine Gesetz in den Vordergrund rückt, wird am besten verstanden, wenn sie dem praktischen Ziel dienen soll, die Handelnden mit einem Entscheidungsverfahren auszustatten. Man sollte die Formel nicht als ein Kriterium der Moral verstehen. Dage-

⁶⁰ This paper was presented at the Symposium on "Kant's Metaphysics of Morals: A 200 Year Tradition of Distinguishing Law and Ethics" held at Smith College, Northampton, MA, August 12 - 16, 1996. I wish to thank Sharon Byrd, Joachim Hruschka, and Jan Joerden for organizing the conference and also the conference participants for their stimulating comments on my paper. I also wish to thank Michael Gorr, Nelson Potter, and John Tienson who all gave me useful written comments on an earlier draft of this paper.

gen dient die zweite Formel, die die Menschheit als Zweck an sich selbst in den Vordergrund stellt, in Kants Moralphilosophie als ein moralisches Kriterium. Als moralisches Kriterium liefert sie auch ein Kriterium für das, was moralisch relevant ist und was nicht, und das sog. Problem der relevanten Handlungsbeschreibungen in Kants Ethik ist gelöst. Ich vertrete also eine Interpretation des Kategorischen Imperativs, die ich die der "unterschiedlichen Rollen" nennen möchte. Die erste und die zweite Formel des Kategorischen Imperativs spielen im Gesamtgefüge von Kants Theorie richtigen Handelns grundlegend verschiedene Rollen.

DEPARTMENT OF
LAW
UNIVERSITY OF
MICHIGAN

Jahrbuch für Recht und Ethik
Annual Review of Law and Ethics

Herausgegeben von

B. Sharon Byrd · Joachim Hruschka · Jan C. Joerden

Band 5



Duncker & Humblot · Berlin