Towards an Adequate Theory of Universalizability

Marian-Ellen Ring

Department of Philosophy

McGill University, Montreal

July, 1993

A Thesis submitted to

the Faculty of Graduate Studies and Research

in partial fulfilment of the requirements of

the degree of Masters of Philosophy

Abstract

This thesis looks at two theories of universalizability: Immanuel Kant's deontological one and R.M.Hare's utilitarian one. It also looks at criticisms of both theories by David Wiggins. It concludes that his arguments against Hare are decisive because the moral theory that follows from Hare's version of the claim that moral judgements must be universalizable is incompatible with several basic requirements on moral theories. Wiggins' criticism of Kant, on the other hand, centres on a technical point that is overcome by an interpretation of Kant's tests for the universalizability of maxims that is given by Onora Nell. Finally the thesis argues that Kant's rational theory of ethics is superior to Wiggins' subjectivist claims because it both reflects our common sense conception of ethics and provides a rational basis for evaluating moral judgements.

Cette thèse a pour sujet deux théories à l'égard de l'universalisabilité des principes moraux: celle de Immanuel Kant et celle de R. M. Hare. Elle considère aussi des critiques de ces théories énoncées par David Wiggins. Elle conclue que les critiques de Wiggins sont décisives contre Hare parce que la théorie morale qui découle de ses arguments ne sont pas compatibles avec plusieurs critères auxquels une théorie morale doit se conformer. Par contre, les critiques de Wiggins envers Kant se concentrent sur un aspect de sa théorie qui admet d'une interprétation plus favorable, tel que celle donnée par Onora Nell. Finalement, la thèse maintient que la théorie de Kant est supérieure à celle de Wiggins, car elle reflète notre conception ordinaire de la morale, et elle lui donne une justification rationelle.

Table of Contents

Introduction

Do unto others as you would have others do unto you. Most people will recognise this as the Golden Rule they learned as children. It may seem like a trite platitude, a simple homily of what could be called "folk ethics", with no real value for either explicating ethical thought or providing guidance in making ethical decisions. To expect it to do either would be like modern meteorologists checking the evening and morning skies for redness in order to inform sailors of approaching weather conditions. Familiarity, not to mention simplicity and antiquity, breeds contempt. But should we dismiss it so lightly? It is often held to be a basic tenet of right conduct, and is not felt by most to need further explication or justification. Philosophers, on the other hand, require something more because they do not just want to know what people accept as a moral guide, but whether they are justified in doing so. Further, they do not want to merely know what guidance people take the rule to provide, but what guidance one can justifiably derive from it.

The Golden Rule has two roles. It indicates, although implicitly, a way of seeing oneself and others that I shall argue is fundamental to the moral point of view, and it gives a decision procedure for moral action in given situations. I shall concentrate mainly on this latter claim and show that the Golden Rule, or a philosophically more sophisticated version of it, can be a guide to moral action. The

1

question will then arise of how this can be so, and an answer will be given in terms of the fundamental moral idea that is the basis of the Rule.

In the idea expressed by the Golden Rule of including ourselves under a rule that we would have everyone follow we can recognize the philosophical concept of universalizability. Essentially, the thesis of universalizability is that if I act under a rule in a given situation, it must be a rule that I can will that everyone follow in a like situation. So the test for moral action is whether I can will that everyone act on the rule that I propose to follow. There are two ways in which I can apply this test, the first of which may be called the logical and the second the empirical. In the first case I take no notice of any facts about particular people; the only conditions that are admitted for consideration are those mentioned in the rule itself. I then ask whether it is logically possible to will that everyone act on that rule, or whether this would engender a contradiction.

In the empirical test, the answer to whether I can will that *everyone* act on the rule I propose is taken to be "no" if there is even *one* case in which I cannot will that the rule be acted upon. In order to see whether such a case exists, I take into account specific facts about particular people in order to find one set of (hypothetical) conditions under which I could not will the rule to be acted upon.

The first of these methods can be found in Immanuel Kant's theory of the Categorical Imperative. In the first chapter I will give a brief account of his position as expressed in his *Foundations of the Metaphysics of Morals*. In discussing criticisms

2

of his theory I will mainly be concerned with the claim that it is not possible to derive substantive results from applications of the Categorical Imperative. I will also look at the objection that applying it would involve disassociating ourselves from the human emotions and relationships that give life much of its value.

In the second chapter I will discuss arguments that R. M. Hare gives for the empirical version of the universalizability test. Instead of determining what actions are morally permissible by a strictly logical application of the test, he looks at the consequences of actions with respect to the satisfaction of the preferences of all those who are affected by them.

In the third chapter I look at David Wiggins' criticisms of universalizability tests. Wiggins argues that Kant's test cannot stand as is, and that it requires an empirical aspect in order for it to have any content. He sees Hare's theory as a reasonable attempt to give it this content, but he maintains that Hare's theory fails also, and therefore that the notion of universalizability cannot be the basis of a workable ethical theory.

If we agree with Wiggins about Hare but do not agree with his conclusions, then his argument must have gone wrong somewhere else. In the fourth chapter I look at an interpretation of Kant's theory that argues that his test for moral action can include enough empirical content to give it substance while still remaining formal. I also argue that Kant's theory meets the conditions that Wiggins rightly shows that any working moral system must meet.

I conclude that the logical version of the universalizability test is the one that is successful, and that it is more successful in choosing specific moral acts because it better formalizes our most basic moral idea. However, one may question how helpful this test is. If the rules it generates are not generally disputed, then it does not help us to solve difficult moral dilemmas. I want to suggest that the difficult cases are difficult just because there is no common sense decision procedure for them, and since the test is a formalization of the common sense way of deciding moral questions, it cannot help. The Categorical Imperative is useful for delimiting what J. L. Mackie calls "narrow" morality. More than this is beyond the reach of ethical theory.

What this means is that for the vast area of human interaction that can be called ethical, what Mackie calls "broad" morality, there are no right or wrong answers, although there may be better or worse ways of arriving at answers. Thus in Hare's example of the Navy captain who must choose between saving the survivors of a sunken ship or torpedoing the submarine that sunk it before it destroys other ships in the fleet, thus killing those in the water, the anguish that torments him whatever he decides is not at the thought that he may have done something wrong, but at the thought that he has done something horrible. If we feel that moral theory has let us down just where we need it most - and we need it here most just because this is where it lets us down - we can at least take comfort in the knowledge that whatever we do, we are not transgressing any moral law, or flouting any moral obligation.

4

Chapter One

In this chapter I shall first give a brief account of Kant's moral theory as given in his *Foundations of the Metaphysics of Morals.*[1] I will then look at some criticisms of Kant's theory.

In order to discover the supreme principle of morality, Kant asks what it is that we mean when we talk of moral worth and moral obligation with respect to action. Because, according to Kant, nothing can be held to be unequivocally good but a good will,[2] in order for an action to have moral worth it must be done for the right reasons, i.e. be motivated by moral considerations and not by inclination nor by the consequences of the action. The only consequences that may be considered are those intended to result from the action. If the only things that can provide a motive for action are reasons and desires, then if an action cannot be motivated by inclination it must be motivated by reason, regardless of any contingent facts of a situation. Not only must the action be in accordance with a moral law, but adherence to the law can be the only motive if the action is to have moral worth.[3]

An action that is morally obligatory is one that we have a duty to perform. Kant therefore examines the concept of duty. The duties that are imposed by moral law are exceptionless; Kant holds that because a moral law imposes an obligation it must be necessary and therefore binding on all rational beings. The ground of

5

this obligation cannot be found in any contingent facts about human beings, but only by examining reason itself.[4]

Kant distinguishes four types of action according to their relation with duty.[5] There are actions that go against duty; actions that are in accordance with duty but are caused by an indirect inclination, as a means to a further end; actions that are in accordance with duty but toward which one has an immediate inclination; and actions that are in accordance with duty and toward which one has no inclination. The first case is not a candidate for being a dutiful, and therefore moral, act; in the second it is easy to discern whether the act was done from duty or as a means to an end. In the third case it is difficult to tell which of the two motives, inclination or duty, is the actual reason for an act. It is only when a dutiful act goes against inclination that it is certain that it has moral worth.

After arguing that only actions done from duty have moral value, Kant goes on to discuss a second proposition of morality, namely that it is not the results of dutiful actions that give them their moral value, but their maxim.[6] An agent's maxim is "the subjective principle of volition,"[7] that is, the rule the agent takes himself to be following in acting. When an act is not determined by inclination or by anything material, it must be determined by "the formal principle of volition."[8] This leads to the third principle: "duty is the necessity of an action executed from respect for law."[9] It is law that objectively determines action, and respect for the law that determines it subjectively.[10] Thus an action has moral worth if it is done

6

out of respect for the law that commands it and for no other motive. This respect is the recognition of the law as one that my will is subject to. For Kant the only object of such respect is the law that we impose upon ourselves because we see that it is necessary.[11]

Next, Kant must find a law that is necessary in that the will must impose it upon itself if it is to be a good will. Because the will in this capacity cannot be influenced by any effects, including the effects of any particular law, it must conform to law as such. He concludes, therefore, that "I should never act in such a way that I could not also will that my maxim should be a universal law."[12] For example, if I ask myself whether I should lie if it would save me from some trouble, I can answer myself in two ways. I might decide that I should not, but only because it might lead to some greater inconvenience later on. In this case, I have looked to see what the consequences would be for me of telling a lie. On the other hand, I might decide that I should act according to my duty, that is, according to law. In order to do so I must ask myself whether I can will that the maxim of my action - that one should lie to save oneself from trouble - should become universal law.[13] According to Kant I could not. If everyone lied when in trouble, no one would believe anyone, and therefore lying would not help me in my present situation if my maxim was a universal law. My maxim is that I should lie, but if it were universal law, its result would be that I cannot lie, and therefore my maxim

7

cannot at the same time be a law. Since to act according to duty is to act according to law, I cannot dutifully, and therefore morally, act on this maxim.[14]

Kant is concerned to show that empirical facts can play no part in determining the fundamental principle of morality. (Which is not the same as saying that they play no part in telling whether a specific maxim conforms to the principle.) If it is to be universally binding it must be based on reason alone. That duties exist and that we must obey them is an undeniable fact of reason even if no one in the history of the world has ever in fact acted out of pure duty.

A rational being has the capacity, which is the will, to act according to principles.[15] For a will that acts completely according to reason, actions that are recognized as objectively necessary by reason are also subjectively necessary. A will that recognizes as necessary the same thing that reason does is free, and is what Kant calls a holy will. If the will is influenced by other considerations, what reason sees as necessary will appear to it as contingent. Thus, if such a will acts according to reason, it is acting under constraint rather than freely as a holy will does. Kant believes that this is the normal condition of man, and is the reason that we speak of obligation in this case.

An objective principle that is to constrain the will is a command or an imperative, and is formulated as an "ought" statement.[16] A moral imperative is one that commands an action as necessary regardless of the agent's individual ends. It is thus a categorical, rather than hypothetical, imperative. A categorical imperative

8

does not address the result of any action, but rather its motive. It commands that an action be performed because it conforms to a moral principle.

The next question that Kant asks is how imperatives are possible, that is, how is it possible that the will is thus constrained. It is easy to see how this works with respect to hypothetical imperatives, as one who wills the end wills the means.[17] But what about categorical imperatives? Such an imperative cannot be empirically shown to exist, because even actions that seem to be in accordance with it may actually have some other motive.

First we notice that only a categorical imperative can act as a law upon the will, because only it is necessary, and not contingent upon other ends. Kant proposes that the very concept of a categorical imperative will yield the "formula containing the proposition which alone can be a categorical imperative."[18] A categorical imperative states a law for the will, and also declares that this law is necessary and therefore one to which the maxims of the will must conform. Since it has no specific content, it gives as a law "the universality of law as such to which the maxim should conform."[19] Thus the only Categorical Imperative is: "Act only according to that maxim by which you can at the same time will that it should become a universal law." By analogy with the idea that nature is what is determined by universal laws, Kant reformulates the principle to read: "Act as though the maxim of your action were by your will to become a universal law of nature."[20]

If we can derive all of the imperatives of duty from this imperative, then we will see what we mean by duty, even if it turns out that no one ever acts according to it. Kant looks at some examples of what are normally accepted as duties to see whether they can be derived from the Categorical Imperative.[21] He looks at examples of duties to oneself and others, both perfect and imperfect. The first example is that of suicide. The maxim Kant claims for the person contemplating suicide is that out of self-love he should shorten his life when prolonging it would cause him more misery. Kant maintains that this could never hold as a law of nature as self-love has the natural role of leading one to improve one's life, and thus it would be contradictory for a system of nature to have that same feeling impel one to end it. (This is an unfortunate example, as Kant goes beyond what the test can legitimately take into consideration by appealing to the proper end of a special motivating force. But this does not count against the legitimacy of the test itself, which can be shown to work within the limits that Kant sets for it.)

Secondly, I can never will as a universal law of nature the maxim to make, when in distress, a promise that I know I cannot keep in order to relieve that distress. If there was such a law no one would believe what was promised, and it would be impossible to make the false promise that I intend.

Thirdly, though out of laziness I may not want to develop my talents, I cannot will that not developing one's talents be a law of nature. Even though such a law could exist without being incompatible with my maxim, I cannot will it to be

10

a law because as a rational being I necessarily will that my faculties be developed, as they are given to me for "all sorts of possible purposes."[22]

Fourthly, one cannot, although not wishing to help those in need and even renouncing aid if needed, will as a universal law that no one help those in need. Although such a law could exist without contradiction, one could not consistently will it to exist as it would contradict one's rational will to be helped if the need arose.

The 'contradiction in conception' test, i.e. that which maxims fail if they cannot without contradiction be thought of as universal laws, fails those maxims that go against perfect duties. The 'contradiction in willing' test, i.e. that which maxims fail if they cannot be willed as universal laws, fails those maxims that go against imperfect duties.

If there are any such duties, they can only be revealed by the Categorical Imperative, but we still do not know whether such a thing exists. It cannot be derived from any aspect of human nature, because duty is binding on all rational beings, so Kant asks whether it is "a necessary law for all rational beings that they should always judge their actions by such maxims as they themselves could will to serve as universal laws?"[23] He answers that if so, "it must be connected with the concept of the will of a rational being as such."[24]

The will is determined by its ends and if it is to be determined solely by reason, that end must be an objective motive for all rational beings. Such an end

11

cannot be subjective, that is an end for someone, but must be an end in itself. Kant proposes that rational beings, and among them human beings, are the only such ends, and are therefore the only ground of a categorical imperative. "Man necessarily thinks of his own existence in this way,"[25] and thus it is a subjective principle. But he also recognizes that every rational being considers his existence in the same way "by means of the same rational ground,"[26] and so it is an objective principle. Thus the imperative is: "Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only."[27] Kant claims that this is merely a reformulation of the Categorical Imperative and that it gives the same results when applied to the same examples. In the case of perfect duties one's actions cannot conflict with the idea of a person as an end in himself; in the case of imperfect duties, action must harmonise with it.[28]

Moral imperatives cannot have interests as incentives because they are derived from the concept of duty, and if there is action from duty it renounces all interest. This idea is built into the Categorical Imperative by the idea of the will recognising universal laws, which do not depend on any individual interest and therefore provide an objective ground for action. But action requires a subjective ground also, and this is given by reference to an end, in this case an objective end. Further, "the subject of all ends is every rational being as an end in itself"[29] Therefore, although the will is constrained by duty to obey a law, it is so only by a law that the will gives to itself. This idea of the will giving to itself a law because

it gives it as a universal law, is autonomy of the will. A case of the will obeying a law given to it conditionally for the fulfilment of some interest is called heteronomy of the will.[30]

Morality consists in acting according to the autonomy of the will. It is forbidden to perform any action given by a maxim that is incompatible with being a universal law. If one's maxims are not already necessarily universalizable, as those of a holy will are, then one's will is subject to the constraint of acting only according to those that are, and this constraint is called duty. "Reason relates every maxim of the will as giving laws to every other will and also to every action toward itself; it does so not for the sake of any other practical motive or future advantage but rather from the idea of the dignity of a rational being who obeys no law except that which he himself also gives."[31]

Morality is the condition under which a rational being can make universal laws, so morality, and humanity so far as it is moral, alone has dignity because "the dignity of humanity consists just in this capacity of giving universal laws, although with the condition that it is itself subject to that same legislation."[32]

This is why acting according to duty is merely obeying a law and, although it is obligatory, is not morally worthy, whereas there is value and dignity in acting out of respect for duty and this comes from the fact that the law is one that we give to ourselves.

13

According to Kant, that the principle of autonomy is the sole principle of morals can be shown by an analysis of the concepts of morality, because we find that its principle must be a categorical imperative and that what this imperative commands is the principle of autonomy. Now Kant wants to see if this categorical imperative of autonomy which is necessary to morals actually exists, i.e. whether morality is possible. It is, he claims, because the will is free, that is, the laws that it obeys are given to it by itself. That the will is free for all rational beings is shown by the fact that "every being which cannot act otherwise than under the idea of freedom is thereby really free in a practical respect."[13] I take this to mean that even the staunchest determinist deliberates and chooses between alternatives, and so acts exactly as though he were free to choose. If we did not feel free to choose we could never act but only wait and see. All action that is motivated by reason takes freedom for granted, and is therefore free for all practical purposes, even if freedom cannot be proven from the theoretical point of view.

That Kant's criterion of the Categorical Imperative, with respect to both types of duties, has intuitive appeal can be seen by any parent who has ever admonished a child with the questions "How would it be if everyone did that?" and "How would you like it if someone did that to you?" The latter question is more obviously an appeal to the Golden Rule, and corresponds with the contradiction in willing test. However, Kant specifically repudiates any attempt to identify the Categorical Imperative with the Golden Rule.[14] He specifically warns against this

14

identification with respect to the necessary duties to others, who are to be seen as ends and not merely as means. Because each person is to be treated as an end, one cannot commit an act that affects another person unless that person also has the (intended) effect of that act as an end. The Golden Rule, Kant claims, allows one to commit any act as long as its effect on another person is an end that one has, as an effect on oneself. Thus one can escape one's duty to help others by agreeing not to be helped when in need. This objection does not apply to the Categorical Imperative because, for Kant, being able to rationally will something is not the same as being able to agree to something. Thus if we agree not to be helped, it is not in accord with what we would will if perfectly rational. Therefore if the Golden Rule is to serve as a moral principle based on reason, it must be interpreted so that it does not allow our actions to be determined by arbitrary inclinations or renunciations.

Kant's theory has engendered a lot of criticism, but some of it has been due to misunderstanding, or even ignoring, certain aspects of his argument. Two common objections can, it seems to me, be met by reminding the objectors of what they are leaving out. I do not want to focus on them, so I will give here merely a rough approximation of an argument against them. The first objection is that of triviality, i.e. that the Categorical Imperative only shows us what actions are permissible, not which are moral. For example, I can will without contradiction that everyone act on the maxim "to brush one's teeth every day," yet one would not

want to say that in so doing one was acting morally. As Lewis Beck points out, there is one other criterion that an action must meet in order to have moral worth: it must be done out of respect for duty, i.e. *because* it can be a universal law. I do not act on my maxim "to brush my teeth every day" *because* it can be a universal law, but because I want my teeth to be healthy, thus it is a hypothetical imperative.[35]

Another criticism is that Kant neglects feelings, and in particular does not give credit for wanting to do good, or feeling good about acting morally. Some have gone so far as to claim that Kant holds that someone who hates his duty but does it, is a better person than someone who likes to do his duty and does it. I do not see any problem with the view that the former shows greater strength of character, but Kant does not say that such a person is more moral, only that it is easier to identify the moral component of an action in these cases. Kant does allow that the more one likes to do one's duty the better, because then one is more likely to do it, but that the pleasure one takes in doing it can never be the motivation for doing it. In fact he claims that the more a person acts from duty, the more of a sense of personal well-being he will have. But it is pointless to act in order to experience this feeling, because by definition one would not then be acting from duty, and one would not in that case have that feeling.[36]

Another argument concerning the lack of feeling in Kant's doctrine comes from Bernard Williams. He claims that if we act only out of duty and never

because of special relationships we have with others, we become alienated from our feelings and those relationships and people. As an example he states the difference between saving a loved-one's life out of duty and out of love. In the former case he maintains that one has asked one question too many in that one places consideration of duty above one's most important feelings.[37] However, many actions are merely permissible, saving loved ones included, and so are not done out of duty at all, although they must not be contrary to it. This then provides the reason for subjecting actions, even those that are motivated by our strongest values to the test: as Henry Allison points out, it is possible that acting on these feelings in some cases would result in actions that are not permissible, e.g. if I sacrifice someone else in order to save someone I love.[38] Our deepest attachments cannot provide justification for every possible action and it is the role of morality to set the limits.

The main criticism of Kant's Categorical Imperative that I want to discuss is that it lacks empirical content, and that if one brings in the empirical considerations required for the idea of universalizability to do any work, it would go against the Kantian emphasis on the rational basis of moral agency. Thus the only way to preserve the method is to lose the ideal.

The specific argument I want to discuss is put forth by David Wiggins. His claim is that universalizability cannot do the work that Kant set for it to do, and that "the only real chance of effecting a meet between the moral point of view and universalizability is to ...[make] heteronomy welcome."[39]

Wiggins starts by giving his account of Kant's moral theory. He reads Kant as maintaining that moral judgements must be made by the agent *qua* rational agent, and therefore independently of any empirically given concerns that the agent might have. In deciding whether a maxim can be universalized or not, one need only examine its meaning, and see whether it can be consistently applied. One need only know what it is to make a promise to know that if it was universally permissible to make a promise with no intention of keeping it, then the whole system of promising would crumble. Wiggins objects that this test has some results that we would not want to accept. It would also be the case that if every lender released his debtor from his obligation, the whole system of lending would disintegrate.[40] However, it is likely that Kant would think that the conclusion in this case was quite correct. One should not release a debtor from a debt that he has voluntarily incurred, as this would be to treat him as less than purely rational because it would not be in accord with an end that he has chosen for himself.* If the debtor were to ask for release one could do so, or if it was impossible for him to pay, one might refrain from demanding payment; but in both these cases the maxim can be universalized without a problem.

Wiggins then looks at Kant's second test of maxims, for consistency in willing. Some maxims can be universalized without contradiction, but a world in

---

* This argument is similar to the one Kant gives in the *Rechstlehre* (Second Part, section 49) for retributive punishment in that it involves seeing a person's action as free and rational and not to be negated.

which everybody acted on those maxims might not be compatible with some of the basic requirements of agency. For instance, the maxim "not to help others in need" can be universalized without contradiction, but would be incompatible with what Kant supposes is the will of every rational agent to be helped when in need. Wiggins objects that this brings in an empirical element, in that the agent takes into account his own contingent desires in deciding to endorse a maxim. And this seems to go against Kant's injunction that the agent, with respect to morals, only act on maxims that he can endorse *qua* rational agent.[41] Wiggins argues that Kant's claim that certain ends are necessarily held by all rational agents illegitimately involves teleological considerations. Wiggins claims that if this move is disallowed, then admitting empirical content in the form of the agent's desires is the only path universalizability can take. The agent discovers whether he can endorse his maxim being acted on in all relevantly similar situations by asking whether he would endorse it if he were in the situation of the person who is now suffering from the maxim being acted on. If his desires conflict with his maxim, then it must be rejected. The test for maxims becomes their "suitability for being acted upon in any relevantly similar situation, including situations where my own position is quite different, and for me to accept this on a desiderative or modified desiderative basis".[42]

R.M. Hare takes precisely such a path as a result of his belief in universalizability as a requirement of moral judgements and in the next chapter I will look at his theory in detail.

<u>Notes</u>

1. All quotations from Kant are taken from Immanuel Kant, *Foundations of the Metaphysics of Morals*. trans. Lewis White Beck (New York: Bobbs-Merrill, 1959). References are to the edition of the Prussian Academy of Sciences.

2. Ibid., p. 393.

3. Ibid., p. 399.

4. Ibid., p. 389.

5. Ibid., pp. 397-399.

6. Ibid., p. 399.

7. Ibid., p. 400, ft. 1.

8. Ibid., p. 400.

9. Ibid., p. 400.

10. Ibid., p. 400.

11. Ibid., p. 401, ft. 2.

12. Ibid., p. 402.

13. Ibid., p. 403.

14. Ibid., p. 403.

15. Ibid., p. 412.

16. Ibid., p.413.

17. Ibid., p. 417.

18. Ibid., p. 420.

19. Ibid., p. 421.

20. Ibid., p. 421.

21. Ibid., p. 421-423.

22. Ibid., p. 423.

23. Ibid., p. 426.

24. Ibid., p. 426.

25. Ibid., p. 429.

26. Ibid., 429.

27. Ibid., p. 429.

28. Ibid., p. 430.

29. Ibid., p. 431.

30. Ibid., p. 433.

31. Ibid., p. 434.

32. Ibid., p. 440.

33. Ibid., p. 448.

34. Ibid., p. 430.

35. Ibid., p. xvii.

36. Paul Dietrichson, "What does Kant mean by 'Acting from Duty"?," *Kant: A Collection of Critical Essays*, ed. Robert Paul Wolff (Garden City: Doubleday and Co.) p.326.

37. Bernard Williams, "Persons, Character, and Morality," in *Moral Luck* (Cambridge: Cambridge University Press, 1981) p. 18.

38. Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990) p. 197.

39. David Wiggins, "Universalizability, Impartiality, Truth," *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford: Basil Blackwell, 1987) p. 69.

40. Ibid., p. 68.

41. Ibid., p. 69.

42. Ibid., p. 69.

Chapter Two

Hare's interpretation of the universalizability test is different from Kant's and so is his starting point. The foundation of Hare's theory lies in the claim that the meanings of moral words such as "must" and "ought" lie entirely in their logical properties. Thus we come to know what they mean by knowing what statements that contain them logically commit us to, just as saying "*a* and *b*" commits us to not saying either "not *a*" or "not *b*". Hare maintains that we can come to know how to think about questions containing these words by studying their logic. Because of their logical properties, we reason differently about statements that contain those words than those, e.g. statements of fact, that do not. The logical properties of "ought" and the deontic "must" are such that judgements containing them are universalizable and prescriptive.

The moral theory that Hare derives from this claim is a form of utilitarianism that combines a formal aspect that he claims is close to Kant's injunction to universalize, with a substantive aspect that looks to the preferences of others in deciding whether a given maxim is universalizable. He further believes that a perfect command of logic and the facts would so constrain moral judgements that we would in practise all arrive at the same ones.

Hare claims that it is a logical feature of the modal use of the word "must" - the use "in which it corresponds to the necessity operator of ordinary modal logic" -

24

that it is a contradiction to say "he must be in the garden, but I can conceive of another situation, identical in all its universal properties to this one, except that the corresponding person is not in the garden."[1] We know this because we know how the word "must" is used. Hare claims, further, that it is a misuse of the word "ought" to say "You ought, but I can conceive of another situation, identical in all its properties to this one, except that the corresponding person ought not."[2] Making an "ought" judgement commits us to making the same judgement in any relevantly similar situation, thus sentences containing "ought" are universalizable. The same logical property is true of the deontic "must."

The second logical feature of moral judgements is their prescriptivity. A statement is prescriptive, according to Hare, "if and only if, for some act A, some situation S and some person P, if P were to assent (orally) to what we say, and not, in S, do A, he logically must be assenting insincerely."[3] (Hare does not place any restrictions on what are to count as situations and acts, but seems that these must exclude speech acts, and situations described in terms of speech acts, if not all statements are to be prescriptive.) Thus, it would not be inconsistent for a person to assent to the fact that a room in a hotel faces the sea, and yet not take that room. This is because the statement "the room faces the sea" is not prescriptive. But if the person assents to the statement "this hotel is better than that one" and yet checks into that one, he must have been assenting insincerely. This is because the word "better" is a prescriptive one. Bernard Williams protests that one can

25

perfectly consistently say that one prefers not stay at better hotels.[4] This shows that it is possible to use "better" in a non-prescriptive way, but Hare claims that this does not matter as long as there is a clear sense in which it is used prescriptively. This prescriptivity also holds for "ought," and especially the deontic "must."[5] Hare claims that it is more obviously contradictory to say "I must not do this" while doing it, than it is to say "I ought not to do this" while doing it. Hare prefers to use the word "ought" because it is more commonly used in everyday language, but stresses that in critical moral thinking, it must be regarded as fully prescriptive.[6]

Hare claims that we know that "ought" and "must" are prescriptive and universalizable because of our linguistic intuitions about sentences that contain them. Linguistic intuitions are the basis of logic, but one cannot jump from this by analogy to claiming that moral intuitions are the basis of morals. Linguistic intuitions about moral terms show us how we must reason about morals, but they can never generate substantive moral claims. We cannot learn whether a moral claim is correct by looking at how language is used, or at whether the community agrees. We certainly cannot decide by looking at moral intuitions, because where a claim is in question, there are bound to be intuitions in favour of both sides of the question. Describing a moral system as the one in force in a society at a given time says nothing about whether the system is correct, just that it is workable. Therefore, on the basis of moral intuitions we could never criticise any moral system, even one that we have discarded.

Hare considers the objection that a theory that argues from facts about language to substantive moral principles violates what is commonly known as Hume's Law, that one may never derive an "ought" from an "is". He rejects this objection because although he reasons from factual premises, using canons of reasoning derived from linguistic intuitions, to moral judgements, this is not done by deduction or any other kind of "linear inference." It is the connection between other people's prescriptions and the prescriptive principles that are chosen by critical thinking that allows Hare to call his argument "non-linear." In *Freedom and Reason*, Hare compares his method to that of science.[7] Deduction does not proceed from facts to principles, but instead a hypothesis is tested against the facts and is held only if it is not falsified. In the same way one proposes a candidate for a moral judgement and tests it against the facts, i.e. by seeing whether it accords with the preferences of others. We derive moral principles from other people's preferences, not by inference, but by considering them as our own. This follows from the requirement that we universalize our prescriptions, which is a logical requirement.

Although it is more natural to talk of people having preferences than issuing prescriptions, it must be remembered that each preference has, according to Hare, a corresponding prescription. As prescriptions are in the form of statements, it is easier to "display the logical relations between them" than the states that they express.[8] The preference that x happen is inconsistent with the preference that it

27

not happen, because the corresponding prescription statements "let *x* happen" and "let *x* not happen" are mutually inconsistent.

Because of the features of moral words, one is constrained in certain ways. The judgement to endorse a maxim must hold in all relevantly similar situations, including situations where we are being acted on in the way in which we are now proposing to act on someone else. This is very different from the Kantian conception of universalizability, but accords with the revised universalizability requirement Wiggins gave in response to the problem he saw with Kant's formulation.[9] As Wiggins put it, the test for maxims becomes their "suitability for being acted upon in any relevantly similar situation, including situations where my own position is quite different, and for me to accept this on a desiderative or modified desiderative basis."[10]

We must now see exactly what this means. It is not clear what is meant by "position" or "situation," but there are two possible ways of reading it. The first may be called the weaker, and the second the stronger, sense of "position." In the first, all that is meant by "my position" is my station and material circumstances. That I would endorse a maxim no matter what my situation in this sense, e.g. rich or poor, servant or master, seems to have little to do with the moral point of view. What is essential to the moral point of view is consideration of what Wiggins calls "the otherness of the subjectivity of others."[11] It is vital that one realize that others have different values and preferences, so that one considers not only what one would

28

want if one was in their position in the first sense considered, but also if one had their preferences. It is quite possible that if I was in the situation of the person now suffering because of my actions, I would not object to being acted on in the same way, because it would not cause me to suffer. But although I might not object to having loud music played next door, I must consider the case in which I have my neighbour's desire not to be blasted with loud music. It is because I realize that I would not want to be subjected to that noise if I was in his position with his preferences, that i realize that I must stop playing my music so loudly that it disturbed him. Hare's universalizability requirement involves imagining oneself in the other person's position in this sense.

As mentioned above, Hare does not maintain that what I ought to do is deduced directly from facts about the preferences of others. Because "ought" judgements are universalizable, that is they must be held in all relevantly similar situations in order to be held at all, my claim that I subscribe to an "ought" judgement is falsifiable by exceptions. Thus, if there is a situation in which I would not want an action to occur, I cannot logically hold that I ought to do that action in a relevantly similar situation.

Hare acknowledges, however, that all that the recognition that I would not like to have x done to me requires is that I no longer accept the prescription "I ought to do x". It does not require that I adopt the prescription "I ought not to do x". It is possible that I might do the action without thinking that I *ought* to do it:

29

perhaps some actions are morally neutral.[12] However, if I consider this to be a situation that requires a moral judgement other than one of indifference, then I must adopt some universal prescription about it, if not that I ought then that I ought not.[13] Hare goes on to argue that one cannot logically make judgements of moral indifference unless one makes *only* judgments of moral indifference. According to Hare, unless we take the path of the amoralist we must make moral judgements in all cases where our actions affect the preferences of others.[14]

If we accept the formulation of the universalizability requirement with the strong interpretation of "situation," there are two problems that arise, according to Wiggins. The first difficulty is that it is possible that I will be unable to arrive at any judgement at all.[15] This is a difficulty that will arise even if the second is overcome, and it will be discussed in detail in the next chapter. The second problem is the possibility that no matter how vividly I imagine to myself other desires that I could have in a situation, the desire I now have is so strong that I would choose to disregard those other desires.

Hare discusses this problem at length in *Freedom and Reason*,[16] but I will use the example Wiggins gives,[17] because it is unusual and interesting in itself, as well as displaying all the features Hare recognizes.

Wiggins mentions a story in which a dictator who wishes England to regain the look it had prior to 1840, orders one of his men, Lord Cavalcade, to tear down every building erected after that date. Lord Cavalcade comes to believe very

strongly in this project, sincerely placing this aesthetic ideal above every other interest he might have. This extends so far that he would agree to have his own home torn down if it had been built after 1840. Further, if he considers the preferences of those whose homes he is destroying and who are now living in the streets, and he asks himself what he would want to have happen to *him* if *he* was in their position with their preferences, he still believes so strongly in his project that he would want *his* home to be destroyed no matter what his preference was at the time.

Hare agrees that this is a logically possible position to take, but he argues that people who are so wedded to their ideals, fanatics as he calls them, are in fact very rare. When he considers the case of Nazis who believe that all Jews should be exterminated, he claims that most have taken this view because they believe false information, or they have not thought morally on the subject, i.e. they have not universalized their prescription. He maintains that if one were to convince a Nazi by showing him evidence that he was in fact of Jewish descent, he would in most cases give up his belief. That there may be some fanatics who would agree that they should be put to death in this case is not a problem for the moral theorist, according to Hare, as long as the theorist keeps his aims modest and is not claiming a watertight theory, but merely a way, given the world and people as they are, to mediate differences in interest.[18]

Although Hare and Wiggins agree that such a fanatic cannot be made moral by a requirement that he universalize his maxim by putting himself in the situation, in the strong sense, of those he is affecting, they disagree about the implications of this fact. Because Hare thinks that universalizability is all that is required by moral thinking, and the fanatic does universalize his maxim, for him the fanatic is beyond the reproach of morals. For Wiggins, the universalizability requirement in this sense is not all there is to moral thought. He maintains this because he holds that there is a difference between what I can will *for myself*, even in another person's situation including his preferences, and what I can will *as that person* in his situation with his preferences. Wiggins maintains that it is possible that Lord Cavalcade, while holding on to his ideal and standing by his judgement that his own home be torn down regardless of his preferences at the time, may come to realize that he is doing wrongly by those he is evicting through his awareness that those people do not want to be living in the streets.[19]

So it is not enough that Lord Cavalcade consider how he would accept being treated *even against his will*, but he must make his decision according to how the people with those different desires want to be treated. Universalizability now requires that he consider how the *others* would consent to be treated, not just how *he* would consent to be treated if he had their desires. The universalizability requirement now takes on its third interpretation: as Wiggins puts it, I must "never endorse a judgement unless [I] will the corresponding maxim on the basis of

32

simultaneous hypothetical consideration of all positions, doing the best for all taken together."[20]

Hare arrives at just this interpretation in his later formulation of his theory. He does so by tightening up the requirements for moral thought. One refinement of Hare's theory, which comes in as he makes it more overtly utilitarian, is the distinction he makes between different levels of moral thought. There are two levels, the intuitive and the critical, both of which are concerned with moral questions of substance.[21] A person thinking morally at the intuitive level is simply applying moral rules that he has learned to a situation where they seem to apply. It is entirely possible, however, that in some situations these rules will conflict, because the rules that are employed at the intuitive level are very general. At the critical level, however, each situation is judged individually according to the logic of moral reasoning, and in each case there will be only one best thing to do, and so only one duty. In some instances, the need to reconcile two conflicting rules may result in one or the other being modified permanently; in most cases, however, one of them is simply overridden, which is to say that the other is acted upon without any modification of the former.

Some people, who see only the intuitive level, take the position that in situations where different moral rules prescribe incompatible actions as duties, whatever one chooses to do it is still the case that one ought to have done the other thing. My duty does not disappear just because I decide in favour of another

33

duty. The argument for this is that I feel remorse that I did not do something that it was my duty to do, which I would not if it had ceased to be my duty. Hare uses a simple example to illustrate this case. Two of the moral intuitions we tend to be raised to have are that we should keep promises and that we should be kind and helpful to our friends. In the example, a professor has promised to take his children on a picnic, but on that very afternoon a friend from Australia, who will be in town only for one day, asks the professor to show him around the university and town. Hare contends that even though the professor concludes, through critical thinking, that his real duty in this case is to give his friend a guided tour, he will still feel as though he is neglecting his duty to his children, even though there is no such duty in this case.[22] It may be that this example is a better illustration of how Professor Hare prefers to spend his Sunday afternoons than it is of moral reasoning, but we can overlook the details of the case and focus on the features essential for the argument. But if anyone thinks that critical thinking should decide in favour of the first duty, this may already indicate a problem with the theory. We must keep this in mind as we look to see how critical thinking is supposed to decide the issue.

Hare contends that what the professor feels because he has let his children down is regret rather than remorse. He further argues that if we are well brought up and have the proper moral intuitions, not only do we follow these intuitions, but we have strong feelings that we *should* follow them. Thus, we will feel regret if we go against these intuitions even in cases where an act that we normally have a duty

34

to perform is not required. The stronger our intuition about what we ought to do in normal cases, the stronger our regret if we do not do it in a specific case.[23]

Principles at the intuitive level, what Hare calls "prima facie principles," should not be too complex, for several reasons. First, they have to be short enough to be learned easily. They must be firmly instilled in each person, so that they can be acted on automatically in cases where critical thinking cannot be done. Second, they have to be general enough to apply to situations that are not identical in every detail. It is because of this necessary simplicity and generality that prima facie principles are overridable. These principles, when properly formulated, give us a better chance of doing the right thing at a given time than if we had to calculate every possible consequence for ourselves. A principle may have unfavourable results in some very unusual circumstances, but it is better to have it fail sometimes than not to have it at all if it works well for most cases.

Besides resolving conflicts between intuitive principles, critical thinking also has the task of deciding what the principles at the intuitive level should be. The method used for both functions is the same, but it is applied at different levels of generality in each case. Decisions about which prima facie principle to act on in a given circumstance need only consider cases that are exactly similar in their universal properties. At the critical level, one disregards one's intuitions and examines a particular case on its own merits, making one's decision based on the specific details of that case. In selecting prima facie principles, critical thinking looks at more

general features of action, and chooses principles whose application will have more good results than not applying them would.

Hare sees the distinction between, and the necessity of, both levels of moral thought as bringing together rule- and act-utilitarianism in one moral theory. Prima facie principles at the intuitive level are such as would be given by rule-utilitarianism, whereas act-utilitarianism at the critical level decides what these principles should be, and resolves any conflicts between them.

Hare demonstrates the function of, and relation between, both levels of moral thought by characterizing the beings who would use only one of the levels. An archangel with superhuman knowledge and powers of thought would always think at the critical level. For every situation he would be able to predict the outcomes of alternate actions. Since he is completely impartial, with respect to himself and to others, he will always endorse a maxim that he would accept in *any* of the positions affected by that maxim. Thus he only selects maxims that are fully prescriptive and universalizable.

The "prole" (as Hare calls him), on the other hand, would never on his own arrive at a maxim that is universalizable, and so needs to acquire, through education, the prima facie principles that will allow him to act correctly most of the time. Thus, his moral principles are merely intuitive. Hare is not claiming that anyone answers to either of these descriptions, but that we are all a bit of both, most of the time relying on intuitive principles, but sometimes thinking critically.

36

The separation of levels explains why many of our moral judgements appear as intuitions, and why these sometimes conflict; it also provides a way of resolving these conflicts. Critical thinking selects principles and decides between conflicting ones by reasoning according to the logical properties of moral terms, prescriptivity and universalizability. Critical thinking is rational, which means, at least, that it is done in light of the facts. All prescriptions, even singular ones, require knowledge of the facts of the situation and the consequences of the prescription if they are to be rational.[24] We could not rationally prescribe something without knowing what it entailed. But universalizable prescriptions require us to consider even more facts, because if the prescription is to hold in all situations, including one where I am in another's position, I must ascertain the facts about that other person, including her preferences.

Because critical thinking requires us to consider the facts of a situation, and it is not possible for us to consider them all, we must decide which facts it is relevant for us to consider. One type of feature that is likely to be relevant for a critical moral principle is the effects of our actions on people, ourselves and others. The method used in critical thinking, which is determined by the logical properties of moral terms, requires us to pay attention to preferences "because moral judgements are prescriptive and to have a preference is to accept a prescription,"[25] and to those of all the people affected because moral judgements are universal and so cannot just consider particular people.

37

If probable effects on the satisfactions of preferences are to be relevant features of actions, then we need to know how our actions will affect the satisfaction of preferences of others as to what experiences they should have. So the question is "what is it like to be those people in that situation?" and rationality in critical thinking requires that we discover the answer. Specifically, we need to know what they would prefer and with what strength.

Although only an archangel can have this information, and so always make the correct moral judgement, ordinary people can still approach correctness to the extent to which they accurately represent to themselves the preferences of others. It is with respect to this that Hare has tightened up his theory. His revised requirements for putting oneself in another's position makes this equivalent to taking account of her actual preferences in deciding whether to endorse a maxim.

When I regard the case of someone else suffering, I cannot say that I know how she feels unless I have an aversion to being in her position with her preferences. Knowing what it is like to have that experience with those preferences involves having the same "motivations with respect to possible similar situations, were we in them."[26]

Hare further explicates this by studying the case of two related statements:

(1) I now prefer with strength S that if I were in that situation x should happen rather than not

(2) If I were in that situation, I would prefer with strength S that x should happen rather than not.

He claims that although they are not identical statements, "I cannot know that (2), and what that would be like, without (1) being true."[27]

There are some problems with this claim. It seems that it is quite possible to know how much I would prefer something in another situation without now having a preference of the same strength.

Bernard Williams objects that I can know that if my house were on fire I would have an overwhelming preference to get everyone out as quickly as possible. But, although I do now have a preference that everyone get out safely should the case arise, strong enough that I now buy a smoke detector and make sure that there are clear exits from everywhere in the house, I recognize that this preference is not as strong as the one I would have if a fire broke out.[28] The problem can be seen more clearly, however, through another example. If I am a drug addict who is trying to quit, I can very well know how strongly I will prefer to be given drugs in a few hours, without now having a preference of any strength to be given drugs in that situation. My knowledge of what my preference will *then* be, is based on my knowledge of what my preferences *have* been in that situation, and not on any preference that I *now* have.

Hare says that (1) must be true in order to know (2) "in the sense of 'know' that moral thinking demands."[29] Williams says that there is no special meaning of "know" in moral thinking,[30] but we can see what Hare means. According to him, moral thinking requires that we identify with other people's preferences, not merely be aware of them. I cannot be said to identify with the other person's preferences unless I now have the same preference. I can only claim to identify with another person to the extent that I now have the same aversion to being in his situation, with his preferences, that he now has. If I say that I know what it is like for him but that I would not mind, then I do not know what it is like for him. But if knowing requires identifying, another problem arises: even if I do now have the same preference as I will have at some future time, I cannot know this until that time. In the case of adopting the preferences of others, I can never know whether or not (1) is true.

Hare needs for his definition to work because the method that he proposes for dealing with conflicting prescriptions involves deciding between them as though they were our own. This solution follows logically from the nature of moral thought - and this is what Hare claims for his theory - only if knowing the other person's preferences *necessarily* requires a complete identification with them, such that I now have the very same preferences myself.

Even if we allow Hare his definition for the sake of argument, there are problems with achieving this necessary identification. First, what sense can be made

40

of identifying with someone in this radical fashion? Second, because of the method used for solving conflicts, we need to be able to compare different strengths of preference.

First, leaving aside scepticism about other minds as a problem for all of philosophy, Hare addresses the question of how one can meaningfully speak of being in someone else's shoes. In what sense is it *me* that I am imagining in her situation, with her characteristics and preferences? Hare's answer, a view put forth by Zeno Vendler,[31] is that "I" has no essence. He takes the case of two different people, Smith and Jones, and concedes that it would be a contradiction to say Jones could be in Smith's shoes, because then he would no longer be Jones. Hare claims that this does not make it contradictory for Jones to imagine that *he* could be in Smith's shoes. Although it may be claimed that Jones has an essence such that Jones could not be Smith. "I", even when spoken by Jones, does not have an essence and can therefore refer to anybody. Hare also suggests that "I" is prescriptive in the sense that identifying with someone, i. e. calling that person "I", is to accept that person's prescriptions. Thus to say "if I were in that position" is to accept the prescriptions of the person in that position.[32]

Even if we admit that it is necessary - for moral thought - and possible to identify with others, we still need to ask how we can come to know what the prescriptions are of the people we identify with. How can we know enough about the experiences of others to enable us to make the proper moral judgements? Hare

argues that the problem of accurately representing to ourselves the preferences of others is a special case of the problem of representing to ourselves any absent experience, including our own past and future ones. His answer is that we do, and with confidence, represent to ourselves our past preferences. We may not be justified in so doing, but, if we assume that these sceptical problems have a solution, we may proceed as though we are. When we represent to ourselves absent states of mind, for past ones we rely on memory; for future ones on a combination of memory and induction; and for those of other people, by analogy with our own experiences in similar situations.

Hare needs not only to be able to discover what preferences people have, but he needs to be able to compare *"degrees or strengths of preference."*[11] He claims that this is not just a problem for utilitarians, but must be faced by any theory that includes a duty to beneficence. We need to know which act would do most good in order to know whether or not we are fulfilling the duty. But a duty of beneficence is a duty to do some good, not necessarily the most good. In any case, according to Hare it is not a problem for anyone, as he claims that is it possible to measure preferences, at least to the extent that we can say that one preference is stronger than another. However, that judgements about strengths of preferences are not indisputable can be seen if we look back at Hare's example of the children, the picnic, and the visiting friend. It can be argued, as I would be inclined to do,

that the children's preference is in fact much stronger. The problem, however, is that there is nothing that we can refer to in order to decide the matter.

If we grant, for the sake of argument, that we can accurately represent to ourselves the preferences of others, then if we admit that to have knowledge of someone else's motivations is to have the same motivations with respect to being in his position, we "reduce comparisons between other people's preferences to comparisons between our own...[I]n so far as I fully represent to myself the strengths of other people's preferences, I have preferences, myself now, regarding what should happen to me were I in their positions with their preferences."[34]

Hare uses the problem of deciding between conflicting intrapersonal preferences as the model for deciding between conflicting interpersonal preferences. There is a conflict of preferences if what I say I ought to do to someone is something that he does not want to have happen to him. I solve the conflict by first fully representing to myself his preference, which requires that I now have his motivations as to what should occur. I can only claim to identify with him to the extent to which I now have the same aversion to being in his situation, with his preferences, that he has. It is this aversion, or preference that something not happen, that is conflicting with my original preference that that very thing happen. I then decide between the two preferences as I would between any two conflicting preferences of my own. The same applies in cases where the preferences of more than one other person are at stake. Hare believes that we can accommodate quite

43

a large number of conflicting prescriptions, as we are able to deal with a great many conflicting prescriptions of our own.

Prudence involves deciding what to do when my preferences conflict, including when my present ones conflict with my future ones. Hare suggests that prudence dictates that I ask myself what would best satisfy my preferences for what happens at the times of the preferences, that is, that I maximize my happiness. So, I must decide between now-for-now and then-for-then preferences. Hare claims that once I have subjected my preferences to the facts and logic, I will acquire a now-for-then preference equal to my then-for-then preference, so that I am choosing between now-for-now and surrogate then-for-then preferences. Once I have acquired present preferences corresponding to my future ones, it is just a matter of seeing which preference is stronger, keeping in mind that the new now-for-then preference has the same strength as the original then-for-then preference.

Hare claims that in all cases the stronger preference wins, which shows that universalizability leads to utilitarianism.[35] He further claims that the stronger preference wins no matter whose it is, so that even the one who has the weaker preference sees this: thus universalizability leads to unanimity.[36]

Hare acknowledges that an objection can be made at this point. It is that talk of preferences is not the same as talk of morality: it is possible to have a weaker preference for what it is morally required that I do. And so talk of prefe-

rences, instead of grounding morality, ignores it. Wiggins discusses this problem in detail, and we will look at his arguments in the next chapter.

This concludes our chapter on Hare's moral theory. It should be noted that his final formulation of the universalizability requirement, that one must decide on a course of action based on the preferences of all those concerned, is different from the first formulation of it, that one must decide on a course of action by looking at what one would accept if one were in the place of those affected by one's actions, with their preferences. The requirement that we adopt the preferences of others is made in order to avoid a problem that Hare saw with his earlier formulation, which is that someone, who Hare calls a fanatic, can realize that he would not like to be in the other person's shoes, but, although being perfectly rational, not let that affect his actions. Hare has effectively defined the problem away by saying that if the fanatic's preferences are indeed stronger, he is not a fanatic; and if they are not then he is being irrational. In the next chapter I will look at arguments by Wiggins against Hare's modified theory.

<u>Notes</u>

1. R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981) p. 7.

2. Ibid, p. 10.

3. Ibid., p. 21.

4. Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge: Harvard University Press, 1985) p. 125.

5. Hare, p. 22.

6. Ibid., p. 24.

7. R. M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963) pp. 87-88

8. Hare, *Moral Thinking*, p. 107.

9. David Wiggins, "Universalizability, Impartiality, Truth," p. 70.

10. Ibid., p. 69.

11. Ibid., p. 70.

12. Hare, *Moral Thinking*, p. 112.

13. Ibid., p. 184.

14. Ibid., p. 186.

15. Wiggins, pp. 70-71.

16. Hare, *Freedom and Reason*, Chapter 9, pp. 157-185.

17. Wiggins, pp. 71-72.

18. Hare, *Freedom and Reason*, pp. 184-185.

19. Wiggins, p. 72.

20. Ibid., p. 74.

21. Hare, *Moral Thinking*, pp. 25-26.

22. Ibid., p. 27.

23. Ibid., p. 30.

24. Ibid., p. 89.

25. Ibid., p. 91.

26. Ibid., p. 95.

27. Ibid., p. 96.

28. Williams, *Ethics and the Limits of Philosophy*, p. 90.

29. Hare, *Moral Thinking*, p. 96.

30. Williams, *Ethics and the Limits of Philosophy*, p. 91.

31. Zeno Vendler, "A Note to the Paralogisms," quoted in Hare, *Moral Thinking*, p. 119.

32. Ibid., pp. 96-97.

33. Ibid., p. 117.

34. Hare, *Moral Thinking*, p. 128.

35. Ibid., p. 103.

36. Ibid., p. 111.

Chapter Three

So far we have had a couple of encounters with Wiggins. The first was when he concluded that Kant's contradiction in willing test was not tenable in Kant's terms because it both required and forbade empirical content. He argued that Kant gets around this by illegitimately importing teleological principles which violate his stated criteria for a basic moral principle. Wiggins argued that the only way to save the principle of universalizability is to allow it to have an empirical content. His suggestion was that actions be decided upon by asking how I would like to be acted on in the way in which I am now proposing to act, if I were "in any relevantly similar situation, including situations where my own position is quite different" - in particular, if it is the position of the person now being acted on.[1]

Wiggins objected that what is meant by "position" is not clear, and he distinguished two different interpretations. It cannot just include what we might want to call the person's external situation, but must involve his desires and values. This is what we have called the strong sense of "position."

Wiggins sees two problems even with the strong interpretation of the universalizability requirement. First, the possibility exists that no maxim will be found that can be endorsed. Second, it is possible that no matter how vividly I imagine to myself other desires that I could have in a situation, the desire I now have is so strong that I would chose to disregard my alternate desires. We

discussed the second problem in Chapter Two. It is now time to come back to the first problem.

The revised interpretation of universalizability that we are now considering is the requirement that we act only on "a maxim that we are prepared to endorse *on the basis of* the consideration of all positions."[2] The universalizer arrives at moral principles by "finding mutual accommodation for desires, interests or preferences as pre-morally conceived"[3] and, because these are bound to conflict, "doing the best one can for all taken together."[4] The question, then, is of how to do this.

The first possibility that Wiggins considers, based on a suggestion by Adam Smith, is that in order to judge between two conflicting interests we must look at both positions from the point of view of a third person who can judge impartially between them.[5] Wiggins claims that in order to do so, the impartial judge must have some criteria by which to judge, and these will have to include his moral views. Judging with impartiality means judging without taking into account who the people are whose interests one is mediating, but presumably it also means judging fairly. And to judge fairly means to judge according to some idea of the merit, or desert, of the people, or the worthiness of the interests in question. Thus, we must judge impartially between conflicting interests on the basis of concepts that are given by morality itself. This, of course, is not an option for someone who wants to generate the whole of morality from universalizability.

In order to get around the problem that one must judge according to some standard, Wiggins suggests a method for deciding between preferences that is similar but involves only the points of view of those whose interests are in question. The person who is making the decision in this case moves back and forth in his imagination between each position. As he sees that each person's original maxim will not be endorsed by the others, he reformulates it to be more acceptable by them, until he arrives at a maxim that each would endorse once he realized that his first choice would not be endorsed by the others.[6] This could be done, Wiggins suggests, by taking "account of others' attitudes and preferences, but no more account than they would take of their own attitudes and preferences if they were taking account those of others."[7] Thus, in searching for a maxim, one starts from the point of one's interests, but is willing to give them up to some extent in favour of finding a maxim, and one also sees others' interests in the same light.

This brings us back to the first problem that Wiggins mentions in connection with his first reformulation of universalizability (that which required putting oneself in the position, in the strong sense, of the person affected.) The problem, Wiggins claims, is that there is no reason to think that a maxim will be found, or that only one, or mutually compatible maxims will be found.

Also, if each person finds a maxim he can universalize, but each finds a different one, the differences might be due to, as Wiggins puts it

...their different conceptions of what is forbidden and what is not, from their different evaluations of the virtues, from the divergency of their ways of sorting attitudes, preferences, interests by their moral quality, or from the different emphases they put upon claims of need, claims of desert, claims of formal entitlement, claims based on prior undertaking, etc...[8]

If such prior evaluations are disallowed, then how are we to judge between conflicting interests? We are faced with the possibility of total indeterminacy.

Wiggins recognizes that Hare *does* say how we are to resolve conflicts of interest without recourse to antecedently given moral values, but he does not think that this method can accomplish the task that he sets it. The whole point of the requirement that we come to have preferences ourselves that are identical to the preferences of the people who are affected by our actions, which Hare argues is necessary if we are to know what these preferences are, is to make the method for resolving interpersonal conflicts of preference the same as that for resolving intrapersonal ones, that is, by applying prudence. Hare claims that this method requires that we give equal weight to equal interests, taking into account only the different strengths with which they are held.

Wiggins considers this method as a way in which the threat of indeterminacy can be removed, but claims that it involves illegitimately importing a moral standard. Wiggins argues that it does not follow from the theory, which only

requires that we do the best for all interests taken together, but does not itself say how. Hare,[9] J.L. Mackie,[10] Don Locke,[11] and Bernard Williams[12] all agree that adopting other's preferences as our own leads to utilitarianism. Wiggins, however, claims that they are making two transitions in getting from one point to the other. The first is from a principle that we do the best for all taken together to a principle of evaluation of the claims of others, i.e. to give them the same weight we give to our own. Wiggins claims, however, that we can do the best for all taken together by evaluating claims according to "agent-centred or self-referentially altruistic principles,"[13] i.e. principles that select those to whom the agent owes altruistic behaviour according to some relation that they have to the agent. The second transition is from this evaluation to a principle of allocation, i.e. to each in proportion to the strength of his interests.[14] Wiggins objects that even if we disregard *whose* interests we are considering, we will still want to distinguish between needs, desires, preferences, etc., and how these different types of interest are to "count in favour of this or that allocation."[15]

In response to Hare's claim that these transitions *do* follow if we apply prudence to interests that we treat as our own, Wiggins further argues that in cases where many conflicting preferences are involved prudence might dictate, not that I try to discover which preference is the strongest, but that I find some way of discriminating between them. So, applying prudence to my adopted preferences does not necessarily result in equal weight being given to equal interests. Further, in

making these discriminations, prudence might do so on moral grounds. If, as is likely, it does, then this method cannot serve as a way of generating moral principles.[16]

It seems inevitable that we judge preferences on grounds other than strength alone, and it can be argued that Hare himself slips and does this. In his example of the professor who chooses to guide his out of town friend around the college rather than go on a picnic with his children, his choice cannot easily be seen as having been made based on strength of preference; things tend to mean more to children, and their preferences must be combined. It makes more sense as a decision based on the consideration that the friend lives half-way around the world and will be in town only for one day, and that the children can have their picnic next week. And this is as it should be: it does not look like a moral proposition to say that if one wants something badly enough, one deserves it.

Wiggins argues that if we did decide between conflicting preferences by granting the most weight to the strongest ones, whether or not this utilitarian solution followed from the theory itself, the principles thus generated would not be recognizably moral. He claims that the maxims that would be generated would be "extremely strange candidates for the moral point of view."[17] It seems to us that some interests are more worthy than others, and so deserve special consideration independently of the strength with which they are held. If they are granted the same weight as other interests, then the resulting decision will not reflect that worth.

Wiggins claims that because he cannot distinguish between worthy and unworthy interests, a utilitarian, despite protests to the contrary, would have to allow some people to be sacrificed if the numbers showed that more and stronger preferences would be satisfied in that case.[18]

Hare, however, maintains that if the results of applying his theory go against our intuitions about what is moral, this will happen only in exceptional cases. Our intuitions are a faithful guide only in normal cases, which is all that is required of them. Strange cases are likely to yield what appear to be strange results because our intuitions were not made to cover them.

Hare answers all such objections by referring to the distinction between the levels of moral thought. Critical thinking has two roles, that of deciding what intuitive principles are best adopted, and that of deciding what to do in particular cases where our intuitions conflict. To take an example, each time Lord Cavalcade goes to evict someone, he should ask which preference is stronger, his to have a beautiful city, or the family's not to live in the street. Each time he finds that their preference is stronger, he should refrain from evicting them. If it turns out that this is most often what happens, then as a matter of policy, which can be thought of as a principle at the intuitive level, people should not be evicted from their homes in the interests of an aesthetic enterprise. The principle at this level must be very general in this way. Now, consider the case where a family does not very much mind living in the street. Their preference to keep their home is outweighed by

54

Lord Cavalcade's preference to evict them. Our intuitions tell us that it would be wrong to evict them because of the general principle that we have internalized. However, at the critical level, if the preferences do measure up in this way, the family is evicted and rightly so. Hare maintains that such cases will never actually occur: in the world we live in, fanatical preferences will always be shown to be alterable in light of the facts and logic. The preferences that we think ought to be acted on will always be the stronger ones. A Nazi will never have a stronger preference for killing a Jewish person than the latter will for staying alive.

Even if Hare is right and the sums always come out the way we think they should, there is still a problem. Bernard Williams raises the objection that in deciding what to do, the Nazi's preferences should not even figure in the calculation.[19] Further, there is no independent reason for thinking that the purpose of morals is to ensure that the greatest preferences are satisfied. Hare reaches this conclusion because he needs - in order to justify his claim to be showing what is morally required by showing what follows from the meanings of moral terms - a way of mediating between preferences that does not involve prior moral judgements. But if prudence does not choose merely on the basis of strength, how can morals? And it must be remembered that we have gone this distance with Hare's theory only under protest: his argument failed when he equated *knowing* someone's preference with *identifying with* it. If his results had been worth saving, it would have been worth trying to come up with another argument for making this identification.

However, it seems clear now that measuring preferences will not enable us to make moral judgements.

Yet universalizability still seems to be an inescapable aspect of moral judgements. It is illegitimate to make a moral judgement in one case and refuse to make it in a relevantly similar case. This is the premise that Hare starts with, and one that Wiggins grants, except that he does not believe that it can serve as the foundation of the whole of ethics; it has a role to play only within an existing ethical system. Next, I will look at the role that Wiggins proposes for universalizability, and then I will look at what an ethical system must be like, according to Wiggins, in which it is to play this role.

Wiggins claims that if we are to have universalizability at all it cannot be in the business of generating a moral system but only of clarifying one that we already have, so that the universalizer has a moral base from which to judge conflicting interests. The universalizer takes prospective maxims and subjects them to public scrutiny by seeing how well each would sit with the other people involved. But these maxims are evaluated according to "a pre-existing understanding of the virtues...of the difference between vital interests and mere desires...of what distinguishes the more morally admirable from the less morally admirable..."[20] The universalizer's job is to "straighten out, to correct and to extend an existing corpus of judgments about a subject-matter already anecdotally and experimentally known, using only an extension of the *same methods* by which the original judgment(sic) were arrived

at."[21] He does this by figuring out what each person in a dispute would agree to, given that others would not agree to his first choice nor he with theirs.

But if this is an extension of the same methods that were used to arrive at the moral notions that we already have, this suggests that, historically, consensus is arrived at by a similar process of bargaining between actual individuals. When the rules in force in society are unacceptable to some, there will be pressure for them to change, and the result (aimed for) will be a set of rules acceptable to all. But this acceptability need not be based on pre-given moral notions, and indeed if we are talking of how these notions came to be, then it cannot depend on them. The universalizer mimics this process in order to decide on his own what would be decided by the parties involved if they worked things out fairly, but there is no fixed procedure for arriving at this decision. The procedure is one of "tatonnement" for which there are only general guidelines. Conflicting moral claims or judgements can be mediated by subjecting them to public scrutiny, where what is accepted from all positions is seen as objectively, or at least intersubjectively, acceptable.[22]

Wiggins suggests that we draw on our experiences in three different roles - those of agent, patient, and spectator - in order to decide how to act, or what judgements to make. He suggests that the Golden Rule appeals to our role as patient in deciding how to act, and that Pittakos of Mytilene's precept - that one not do oneself what one disapproves of in others - appeals to our role of spectator.[23] Wiggins claims that this is what Confucius had in mind when he said

57

"Do not to others what you would not like yourself." Wiggins looks at this rule in the contexts of other recommendations Confucius gives along with it: "Behave when away from home as though you were in the presence of an honoured guest. Deal with the common people as though you were officiating at an important sacrifice."[24] I see this as suggesting that we always be on our best behaviour, that we never behave toward anyone in a way that we would not behave toward, or in front of, someone we respect. Wiggins, on the other hand, sees it as suggesting that one refer ones actions to a public perspective, where I judge what I do according to both my reaction to being acted on in this way, and my reaction to seeing others act and be acted on in this way.

This interpretation makes sense in connection with the precept of Pittakos of Mytilene, which calls for our feelings as spectators to inform our actions, and which indeed seems required on pains of hypocrisy, if not contradiction. However, it cannot generate moral judgements. If one disapproves of the drinking of alcohol in others, it seems that one should refrain from drinking; but no moral judgment about drinking can be legitimately derived from mere disapproval. Many people do draw conclusions about morality from their own feelings toward certain activities, but this is usually considered narrow-mindedness rather than legitimate moral argument.

The public perspective to which Wiggins refers our moral judgements involves not only considering our own response in various public roles, but also

considering the responses of others to the same events, in the same roles. Wiggins claims that universalizing has the role of helping us to discover what the consensus is that should be arrived at with respect to any moral judgement.

Wiggins is led to a consideration of Hare's theory because of a suggestion from Hume that Wiggins finds appealing. Hume says that when someone adopts the moral point of view, he "refers his actions, feelings, evaluations, complaints and exhortations to the point of view that is 'common to him with others.' Departing from his 'private and particular situation', he must try to speak and think not only on behalf of himself but as if on behalf of others too."[25] If he sees that not everyone would agree, then he must either try to convince them or he must revise his claim.

Wiggins notes that consensus does not serve as the "source of moral information"[26] We do not arrive at a moral conviction by asking what others believe; it is not in this sense that we turn to others in order to verify our conclusions. The consensus to which we appeal is not actual: I do not verify my judgements by asking what it is that others as a matter of fact say, but by asking whether those judgements are ones upon which consensus could be reached. Each moral judgement is a claim about what consensus could be reached. and therefore about what the moral facts are. This consensus is not one that actually exists - if it were, then Wiggins' claim could easily be refuted by showing that there is no such thing - but "one which it is natural for human beings living together in society to arrive at."[27]

Wiggins claims that "on the level of content, consensus appears somehow intrinsic to what judgments of value, moral obligation and moral necessity actually say." Moral judgements are about how certain things affect us, not each of us individually but all of us in virtue of a shared tendency to be affected in that way. From a personal point of view one might say "I do not like ι." But if one says "ι is a bad person", one is assuming that everyone would, or at least should, agree. Wiggins quotes from Hume:

[When making a moral judgement, a man must] depart from his private and particular situation and must choose a point of view common to him with others; he must move some universal principle of the human frame and touch a string to which all mankind have an accord and symphony.[28]

It seems to me that there are three ways in which all people may share the same idea of something: the first is if the thing in question is given in perception using sense organs common to all people, e.g. all people (except, of course, for those with identifiable physical deficiencies in this respect) share the same abilities to make colour discriminations, and consensus about what colour an object is is grounded in human biology.

The second is if certain objects or events arouse certain emotions or feelings in everyone in virtue either of being human or of living in society. In this case

either people have an inherent disposition for feeling, for example, pity when faced with someone else's misery, or else this disposition is learned but without reference to any reason for feeling this way. We do not feel badly because something is wrong, we simply feel badly, and so say that it is wrong.

The third way in which people may come to have the same idea about something is if it is an idea that is rationally reached, and rationality is something that all people have in common. In this case something can be said to be wrong for a reason, and this can be a reason for everybody.

Wiggins claims that the consensus involved in moral judgements is "derived from the shared propensity to feel the various feelings that are presupposed to the fixing of the sense of predicates that occur essentially in judgments expressive of morality."[29] This propensity is one that all human beings have, but in different cultures these feelings may be elicited by different things. Very loosely, the reason for this is that there is a historical process during which it is decided that some of the things that elicit a response are really such as *should* elicit that response, and some are not. In this way we come to change our responses to things, but there is no point at which the responses we *do* have are, in virtue of our having them, just those that we *should* have, and so there is always room for criticism and change.

It is not clear to me, however, if moral judgements are based on feelings that we have - although subject to changing ideas about what feelings we *should have* - how we can ever justify our moral judgements. Either the changes we make in

the range of what may elicit a particular response are made according to reason, in which case, if reason - as I believe - is not culturally relative, then all cultures should arrive at pretty much the same consensus. If, on the other hand, these changes have no basis in reason, then it is difficult to see in virtue of *what* these changes will be made, and even more difficult to see in virtue of *what* arguments about what the consensus should *be* are to be evaluated.

It is in general possible to explain any feeling that it is natural to find in common among people living together in society by showing that the state of affairs that is promoted by that feeling is one that it is beneficial for people living in that society to promote, e.g. the feeling that stealing is wrong promotes security in a society whose citizens have private property. Unless the moral feelings that arise in a society are arbitrary, we should be able to see why they have arisen. If moral judgements are amenable to this type of explanation, then we can show that certain "moral" feelings are wrong by showing that they have arisen for bad reasons.

Wiggins justifies the claim that moral judgements involve consensus by appeal to what moral judgements say, but he cannot justify the content of that consensus, nor, therefore, particular moral judgements. Although consensus is an important aspect of moral judgements, and perhaps even a *necessary* aspect, such consensus about what our moral judgements should be must be based on something more solid than sentiment, even shared sentiment, if we are to able to criticize the judgements we do make by reference to the judgements that we should make.

Although there can be rational discussion arguing from shared responses to moral judgements, these responses cannot provide a rational foundation for such conclusions, as they are at some point simply given. A more solid foundation for ethics would be a rationally acceptable view of the world, and our place in it, that could serve as a true premise in rational moral argument.

It was the idea of consensus, and that this involves seeing things from an impersonal standpoint, that lead Wiggins to consider Hare's claim that seeing things impersonally entails giving equal weight to equal interests. However, he rejects Hare's theory because it is not faithful to our actual moral experience. In order to make clear the problems that he sees with Hare's theory, Wiggins sets out a list of "the background facts... that help to make the moral point of view what it is."[30] These facts are a combination of reasons why we need a moral point of view (because of interests that we have); ways of ensuring conformity to it (by being favourable to interests that we have); and the need to be unaware that we are adopting moral principles for any of these reasons, but to see them as good in themselves and not as means to ends.

As Wiggins points out, a moral system has to be possible, that is, it must be able to be embodied in the practices of an actual society. He lists seven points that he thinks are true of any moral system:

1- In living together human beings must see conflicts of interests as requiring arbitration by agreed procedures whose outcomes, at least in theory, are seen as fair.

2- Human beings need norms of reciprocity and cooperation that can counteract the settled tendency of things to turn out badly rather than well.

3- These norms must not only determine expectations, but be such that can take on a life of their own by being seen as enshrining values and virtuous concerns that have a "moral beauty" all of their own.

4- Whatever the origin of morality, unless it can become an end in itself to act according to its demands, e.g. reciprocity, loyalty, veracity, and unless this can help fill out the picture of what it is constitutively for things to go well, the prospects are not good for a standard that requires these things. It cannot survive the idea that such values are mere means to other ends.

Furthei, a social morality that is embodied in institutions and practices that are consented to by those they govern will be regulated by two other factors:

5- A social morality must safeguard and uphold arrangements that specially protect whatever is vitally or centrally important to the happiness or welfare of individual agents; and it must carefully proportion the protection it provides for an individual's interests to how central these interests are for the individual-as an incentive for people to participate in it.

6- Further, a social morality must require of its participants some however minimal and delimited respect towards other participants--at least an acceptance of them as *equally participants*. Each participant must recognize every other participant as equally a bearer of rights and duties who must himself recognize any other participant as equally the bearer of rights and duties...

Lastly, he refers to the notion of consensus in that

7- a point of view that can be shared between members of an actual society must give expression to a potentially enduring and transmissible shared sensibility. To adopt the moral point of view is to see one's thoughts, feelings and actions as answerable to the findings of such a shared sensibility.[31]

These points are meant to show what a morality must be like if it is to exist at all, not to make any claim - in spite of the first two points - even with respect to preserving interests, about the purpose of morality. So far there is no mention of what particular ideas are to be adopted in virtue of adopting the moral point of view. Wiggins claims that a social morality and the actual mores and institutions that embody it can be mutually scrutinized and adjusted, as well as sustained and perpetuated.[32] But this only means that the rules can be brought into better conformity with whatever the shared sensibility dictates, not that they can be made better, or more moral, by reference to any idea of what morality is supposed to be or supposed to be for.

Wiggins claims that morality cannot be seen as a tool for human welfare: if that is how it arose in the first place it has transcended its origins. It may not be a tool but a necessary condition for human welfare, or else why do we care to have morals in the first place - what would be lost by abandoning them?

Wiggins claims that morality is something above and beyond a way of assuring human well-being described pre-theoretically. The well-being that morality is supposed to assure is one that can only be defined in terms of morality itself. However, he does allow that there are some goods and evils that can be given independently of social morality, and that they play a role in the foundation and maintenance of morality, although he thinks that there are important ones that cannot. It would be helpful here if he gave some specifics as it would help to clarify his position. If we knew what things he takes to be good and bad pre-morally and morally, we can ask how he makes the distinction, and whether we can have an idea of the latter that is not based on the former.

Wiggins considers the claim that universalizability is constitutive of the moral point of view. He then argues for a conception of the moral point of view that he finds appealing and shows that it does not depend on the notion of universalizability. He claims that universalizability and the Golden Rule cannot have the combination of "self-sufficiency and moral and conceptual primacy,"[11] that they claim, and that therefore they cannot serve as the basic concepts from which all morality is derived.

Wiggins argues that universalizability can serve a function within a pre-established moral system. Its function would be to make clear what is required by that system by judging conflicting claims according to the criteria provided. T h i s differs from Hare's theory in that the differences that may arise are not clashes of brute preferences, but of interests that may already have a moral value. The difference may be made clearer by using an example of a type that Hare discusses. If during a fire I can save either a great doctor who, if he survives, will go on to save the lives of many others, or my own child, what shall I do? According to Hare, preferences weigh in favour of saving the doctor. But this result is not implied by the idea of consensus: it may be the case that prevailing attitudes are such that a consensus could be reached that I should save my child. Wiggins complains that according to Hare's utilitarian version of universalizability, feelings such as a mother has for her child can be seen as good only because they are beneficial. Wiggins wants to claim that such feelings are good in themselves. (I will argue in chapter 4 that, although this may well be so, no *moral* judgement is involved in such cases. People who do not save people close to them are seen as deficient in some way, but not morally deficient.)

If Wiggins' arguments are decisive against Hare's utilitarian interpretation of universalizability, the idea that consideration of particular empirical circumstances is necessary in order to make sense of the requirement that moral judgements be universalized has led nowhere.

If we accept some of what Wiggins says about the requirements of morality in general, but do not accept that the moral point of view is a matter of shared *feelings*, nor that it is the only job of universalizability to mediate between interests, even post-morally conceived, then we might want to reconsider our options. The main question is where do these moral notions come from, that the universalizer mediates between? For example, to take one that is pretty uncontroversial, where does the idea come from that it is wrong to lie? For each generation, of course, it comes from parents, teachers, and others; but if the question is raised "why not?" we should be able to give answer other than "because that's the way it is" or even "it's not nice." Morality is something that each person should be able to rediscover for herself. If this is not to be done by reference to some goal of morality, which it cannot because the results of pursuing such a goal would not be recognized as morality, then it must be answered with reference to something that it follows from.

Wiggins claims that although the consensus aimed at by morality is that warranted by objective truth, it is not clear that these methods will result in any such thing.[34] It seems to me that a method that is more likely to succeed is that of looking for a consensus whose basis is reason rather than feeling, and in chapter 4, I will consider this possibility. We can look at morality as it is to see what underlies it and then use these presuppositions to bring our actions in line with what morality requires. If morals can be seen as rational, then we can discover through reason what, specifically, we ought to do. This, I will argue, is what Kant's

Categorical Imperative does. But before considering how it reflects our pre-theoretical conception of morality, and how far this reflection provides a justification for using it, we must consider whether it can be used as a guide to moral action.

Notes

1. Wiggins, "Universalizability, Impartiality, Truth," p. 69.

2. Ibid., p. 73.

3. Ibid., p. 73.

4. Ibid., p. 74.

5. Ibid., p. 74.

6. Ibid., p. 75.

7. Ibid., p. 75.

8. Ibid., p. 76.

9. Hare, *Moral Thinking*, p. 111.

10. J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin Books, 1977) p. 93.

11. Don Locke, "The Principle of Equal Interests," *Philosophical Review* vol. 90, 1981, pp. 531 and 555.

12. Williams, *Ethics and the Limits of Philosophy*, p. 83.

13. Wiggins, p. 85.

14. Ibid., p. 85.

15. Ibid., p. 85.

16. Ibid., p. 86.

17. Ibid., p. 76.

18. Ibid., p. 77.

19. Williams, *Ethics and the Limits of Philosophy*, p. 87.

20. Ibid, 78.

21. Ibid., p. 79.

22. ibid., p. 82.

23. Ibid., p. 83.

24. Confucius, *Analects* II.2, quoted in Wiggins, p. 82.

25. David Hume, *Inquiry concerning the Principles of Morals*, section IX, quoted in Wiggins, p. 59.

26. Wiggins, p. 60.

27. Ibid., p. 62.

28. Hume, *Enquiry concerning the Principles of Morals*, section IX, quoted in Wiggins, p. 60.

29. Wiggins, p. 62.

30. Ibid., p. 64.

31. Ibid, p. 64-66.

32. Ibid, p. 67.

33. Ibid, p. 63.

34. Ibid, p. 84.

Chapter Four

If we are not satisfied with the account of morality Wiggins offers, we may want to question his arguments against alternatives. If his criticisms of Hare are valid, then the notion of universalizability that concerns ethics cannot involve giving empirical content in the form of contingent desires a role to play in the testing of judgements. But this is not the only way of applying the universalizability requirement. Wiggins dismisses Kant's own version rather quickly, and perhaps mistakenly. If there is a way of reading Kant that makes his claims more plausible it is worth investigating its merits before confining universalizability to a much lesser role in moral thought.

Kant claims that the fundamental ethical principle is the Categorical Imperative: "Act only according to that maxim by which you can at the same time will that it should become a universal law."[1] The agent sees whether his maxim passes this test by seeing whether it is possible to will that it be a universal law. A maxim can fail this test in one of two ways. An agent cannot rationally will a maxim (1) whose universalization involves an inner contradiction, nor one that is (2) incompatible with what every rational agent *must* will. The test for the first of these

possibilities is called the "contradiction in conception" test, and for the second is the "contradiction in willing" test.*

Wiggins admits that the CC test gives "some impressive results," but contends that it gives some strange ones as well.[2] It would classify 'releasing one's debtor from his debt as an act of generosity' as a forbidden act. Wiggins finds this result counterintuitive, but as we argued in Chapter 1, if we treat the debtor as fully rational, then the result of the test makes sense. This is exactly the kind of counter-example Paul Dietrichson claims would undermine the CC test,[3] and it may be possible to come up with one that does so, but I cannot think of any. It should be expected that applying the CC test might sometimes give results that seem strange at first, but make sense on further reflection: if our first intuitions about a case are taken as absolute data against which to test the results of applying the Categorical Imperative, there is no point in looking for a principle that can guide our decisions and show us whether our intuitions are correct.

Next, let us examine the problem Wiggins sees with the CW test. He takes the example that Kant gives and asks what happens when the agent tries to universalize the maxim "to neglect someone in need." The contradiction here occurs between the universalized maxim that no one help anyone and the will of the agent to be helped when in need. Wiggins claims that the latter is a merely contingent

---

* In the interests of brevity, these tests will henceforth be respectively abbreviated the CC test and the CW test.

desire of the agent and that this is in fact the on'y kind of consideration that can be brought into opposition with the universalized maxim. It is this line of thought that leads him to consider Hare's theory. If the contradiction is not to depend on the agent's contingent desires, then there must be at least some end, e.g. to be helped when in need, that all rational agents necessarily have. The problem here is to defend this claim without making any teleological assumptions.

Onora Nell gives an interpretation of Kant's theory that concludes that his test for maxims is effective for discovering which actions are morally worthy, morally unworthy, or without moral worth. It is also effective for discovering which maxims it is permissible, obligatory, or forbidden to act on. If she is right, and if her interpretation does not involve any teleological claims, then universalizability is redeemed, so in this chapter I will look at her arguments and see how far we might want to agree with her. It is outside of the scope of this thesis to examine whether her interpretation of Kant is justified by his writings. Whether or not she accurately represents Kant's thought, if she gives a reasonable account of a universalizability requirement that does everything we could want it to, then she has done enough.

Nell wants a moral theory to answer the question "what shall I do?" For every proposed action the question arises of whether or not it is legitimate to do it. In order to answer this question, one must be able to describe the act in question. In a theory that holds that this question is answered by seeing whether the proposal that one act in a certain way is universalizable, we need to know which

74

description of the act to test. As any act can have infinitely many descriptions, the question of which one is relevant is a big one. Nell claims that Kant has answered this problem by making the action's maxim, i.e. the principle that the agent takes himself to be acting on, the subject of the universalizability test.[4]

Now that we know what it is that we are testing, we must see how the test works. First Nell addresses the question of what form a maxim that is willed as a universal law will have. She characterizes maxims with a schema of the form:

To - - - - if....

where ---- is an act description and .... is an agent description, e.g. respectively "make a false promise" and "one is in trouble with only this avenue of escape open to him."[5]

The universalized version of this is:

Everyone to ---- if....

This, however, is still merely a practical principle and not a law.[6] The notion of a universal law must be made concrete so that it can be used, and Kant suggests that this be done by considering an analogous natural law. The formulation of a law of nature that corresponds to the universalized maxim is:

75

Everyone will - - - - if....

Nell calls this the universalized typified counterpart, or UTC, of the maxim, because

the universalized maxim is typified by a corresponding law of nature. It is this UTC

that we test for contradiction.[7]

Once we have established the UTC of a maxim, we must determine whether

a contradiction exists. There is some dispute, however, about where this

contradiction is supposed to occur. Some commentators, e.g. Dietrichson and J.

Kemp, hold that the contradiction must occur *within* the law itself; Nell maintains

that the contradiction must be between the law and the maxim. Both sides take as

their justification Kant's reference to an "inner contradiction."[8] Nell, however, also

takes account of the injunction to will the law "through the maxim" and takes it to

mean within this act of willing which is at the same time of the maxim and of the

UTC.[9]

Nell argues against the Dietrichson/Kemp interpretation by claiming that

there is no contradiction within the law taken by itself. The contradiction does not

arise within the law, but between the law and some set of arbitrary facts about

human life. In the case of the law "Everybody will make false promises when in

financial difficulty," the contradiction occurs only if it is the case that people learn

from experience, and this fact is not part of the law that is being tested.[10] Because

some facts beyond what is given in the law itself are required in order to derive a contradiction, we need a way of non-arbitrarily determining which facts to consider.

Nell's interpretation does just this by seeing the contradiction as arising between two *intentions*, one expressed by the maxim, the other by the law. Nell asks what can be meant by willing a universal law. It cannot mean "want" as I may want many incompatible things. She suggests that the agent *intends* both to act on the maxim and for it to be a universal law of nature. These intentions are contradictory if they have as their objects two sets of circumstances, one of which makes the other impossible. This accounts for the specification that the agent must intend the law *at the same time* as the maxim, because it is only if they are simultaneous that two such intentions are incompatible.[11]

It this is to make sense at all, we need to make sense of the idea of intending something to be a universal law of nature. As the agent does not have the power to make such a law, how do we ascribe to him the intention to do so? Nell suggests that we do so if the agent *would have* that intention if he were able to make universal law through an act of the will. This is supported by the formulation of the Categorical Imperative: Act *as if* the maxim of your action were to become through your will a universal law of nature.[12]

The claim that it is intentions that we are testing for compatibility allows Nell to say what empirical content it is legitimate to consider in applying the test. To intend to do something involves intending some set of conditions sufficient for that

77

act to occur, and also the natural and predictable results of that act.[11] This gives her a principled way of including facts that are not explicitly mentioned in either the maxim or the law and are required in order to derive a contradiction between them.

Now I would like to look at Nell's interpretation of the CW test. This test depends on showing that there are certain maxims that a person cannot rationally will as universal laws, even though doing so involves no "inner" contradiction. They cannot be willed without contradiction because their UTC's conflict with what every rational person must necessarily will. What needs to be shown, therefore, is that there are necessary ends that everyone must have and that these can be established without recourse to any concept of a proper end for persons, in order to counter Wiggins' objection that Kant appeals to teleological considerations in his argument.

Most ends are objects of desire, and, being empirically determined, cannot be the ground of any moral law. If obligatory ends are not to be excluded from moral argument, they must be shown not to depend on the faculty of desire. According to Kant, it is also possible to *choose* ends. In order for there to be obligatory ends that can play a role in moral thought, it must be possible to have ends that can be chosen without being desired, although, of course, it is possible to desire an end that one has chosen.[14]

Nell schematizes the maxims that are tested in the CW test differently from those in the CC test. Maxims that make it this far have already been found to be

consistent by the CC test. If any contradiction arises here, it will be due to the addition of the purposiveness of the act to the maxim. Therefore the maxim has the form

> To - - - - if .... in order to ----.

Because it is only necessary to consider the end in looking for a contradiction, the maxim may be given in incomplete form

> To do/omit what is needed in order to ----.

The UTC of this is

> Everyone will do/omit what is needed in order to ----.[15]

The example Kant gives is of not helping someone in need, without specifying the act. Nell renders this maxim as "To neglect everything needed to help the needy."[16] It is this purpose that the agent has in acting, or neglecting to act, that is to be tested. Two premises are necessary in order to derive a contradiction in willing. These premises are legitimately imported because they are prerequisite for willing in general. The main premise required is that persons have ends. This is not likely

to be disputed, but if it is one can point out that anyone testing a maxim has at least one end. The other premise is the Principle of Hypothetical Imperatives, which Kant sees as analytic.[17] This principle states that if one wills an end, one wills some set of conditions sufficient for that end. Thus, if a person cannot achieve his end unaided, he must will that others help him. Because an agent necessarily wills these ends, a contradiction arises when he wills the UTC of not helping when needed, i.e. Everyone will neglect everything needed to help the needy. If it is a law of nature that everybody refrain from helping when needed, then the agent will not be helped when needy. Since he must will to be helped, he cannot without contradiction will the UTC.[18] The same thing applies to the maxim of not developing one's talents. The agent must will some set of circumstances sufficient to his ends, and his abilities will be part of that set. In this case, however, it is not necessary to universalize the maxim. Only the agent can develop his talents, so it is not necessary that everyone neglect his talents in order for there to be a contradiction, but merely that the agent propose to neglect his own.[19]

Next Nell looks at what follows if a maxim of ends cannot be universalized. She has shown, at least, that it is forbidden to take as maxims "to do nothing to help others" and "to do nothing to develop my talents." In order to make clear what follows from this, she sets up a square of opposition. She wants to show that the results of the CW do not entail unreasonable demands on people: she wants to refute charges that if it is forbidden to neglect to help everyone then it must be

obligatory to help everyone. Nell claims that there are four possible maxims with respect to an end.

A - To do some of what is needed to achieve x.

B - To do some of what is needed to prevent x.

C - To neglect everything needed to prevent x.

D - To neglect everything needed to achieve x.

Nell claims that these maxims stand in certain relations to one another, such that an agent must adopt one and only one of A and D, and of B and C. Further, he may hold both A and B, but never both C and D. She also claims that A follows from C, and B from D.[20] If this were so, it would depend on the assumption that not doing anything to prevent something is tantamount to doing something to achieve it. It does not allow an agent to be neutral with respect to any end. If he is not actively contributing to the eradication of world hunger, he is contributing to its continuance. This is an extreme view that is not likely to be generally accepted.

She arrives at this square by claiming that there can be no maxims of the form "Doing everything..." or "Neglecting something..." because the former is impossible and the latter trivial for the same reason: we must always omit some of what is needed for an end because some means required by an end are mutually exclusive.[21] For example, I cannot pay my debt both in cash and by check. This is

a strange way of understanding "needed." It seems, rather, that all that is needed in order to pay my debt is to pay it *either* by cash *or* by check (for simplicity I limit the possibilities) and that by doing one or the other I have done everything needed. It cannot be held that what is needed is that I pay by check *and* in cash. This would be impossible, as Nell claims, but this is not the situation described by saying "everything that is needed." If I perform some act sufficient to promote an end, then I have done what is needed to promote that end. No other act can be needed beyond this, even if it was an alternate sufficient act. If we lift the ban on these forms of maxims, we can set up a square of opposition that sets up the relations that Nell wants to show between maxims. I will discuss below what it is that leads Nell to adopt the position that she does.

A - To do some of what is needed to achieve *x*.

B - To omit some of what is needed to achieve *x*.

C - To do everything that is needed to achieve *x*.

D - To omit everything that is needed to achieve *x*.

Notice that B and C have changed, and that these are the forms of maxims that Nell labelled, respectively, trivial and impossible. We have tidied up her square, and now we can address her claim that A follows from C, and B from D, and that A and D are contradictories, i.e. such that one and only one of them is held at any

82

time, and that, therefore, if D is forbidden, then A must be adopted. The problem here is that we are dealing with maxims and not statements. A maxim is a principle that an agent actually ac's and acts on, so although by adopting C one is committed to acting in ways that include those that would follow from adopting A, there is no reason to think that maxim A will actually *be* adopted and acted on by those who act on C.

The same observation holds true for the relationship between A and D. The CW test shows that certain ends are forbidden because they conflict with other ends that every person must have to be rational. If A and D were true contradictories, then the negation of D *would* entail A. However, the fact that it is forbidden to act on maxim D does not entail that I must act on maxim A because it is possible not to adopt any maxim involving x. If I am forbidden to act on the maxim of never helping those in need, it need not be the case that I ever adopt the maxim to help some of those in need: there are infinitely many maxims that are compatible with not acting on D. However, as Nell points out, "[a]ny agent *choosing a policy with respect to* x must either adopt D or A as a maxim."[22] It will then be necessary to judge case by case whether I am required to adopt some such policy. For example, in a case where I am making a promise, I must either intend to keep it or not.

Nell's strange square of opposition is made necessary because of her argument that the CC test can show which maxims are obligatory and which are

83

merely permissible,* as well as which are forbidden. Forbidden acts are those whose maxims fail the tests and whose contraries pass, and obligatory ones are those who pass the test and whose contraries fail. And those of which neither they nor their contrary fails are permissible.[23] (She also argues that those of which both they and their contrary fail are also permissible. This strange conclusion is suggested because of a perceived problem with non-reciprocal action that we will look at below.) She does not claim that this method of discovering obligatory and permissible maxims is restricted to maxims of action, and this method involves testing the *contrary* of a forbidden maxim, which if not itself forbidden is then obligatory. In the case of maxims of ends, the contrary of maxims of the form D is one of form (revised) C. Therefore if, for example, omitting to help everyone is forbidden, then helping everyone is obligatory - and this is a conclusion that Nell is anxious to avoid. Her formulation of C, however, does not require that one adopt the maxim of helping everybody. On the other hand, as we have seen, this version of C does not entail adopting a maxim of the form A. She further confuses the issue by calling A the contrary of D.[24]

Let us see whether we can clear this up. All that follows from the discovery that it is forbidden to act on the maxim "To x" is that one must omit acting on that maxim. The descriptions "Acting on the maxim 'To x'" and "Omitting to act on the

---

* For ease in exposition I will refer to the maxims of forbidden, obligatory, and permissible acts as, respectively, forbidden, obligatory, and permissible maxims.

maxim 'To x'" are contradictories, and thus one and only one can be applied to an act. What does not follow is that I must adopt the maxim "To not-x." Being forbidden to act on "To make false promises" does not entail that I must act on "To make true promises": there is no need for me to make any promises at all. These are contrary maxims: they cannot both be acted upon but one can omit acting on either. There exists the possibility adopting no policy with respect to x. Therefore, as was mentioned above, determining whether the contrary of a forbidden maxim is obligatory depends on determining whether some policy *must* be adopted with respect to x. Only in this case will the contrary of a forbidden maxim be obligatory if it can be consistently universalized.

Does the same analysis apply to the case of maxims of ends that are forbidden by the CW test? If it is forbidden to act on the maxim "To do nothing to promote x" what follows? The contradictory of the description "Acting on the maxim 'To do nothing to promote x'" is "Omitting to act on the maxim 'To do nothing to promote x.'" This is what follows from the maxim being forbidden, but what follows from this? One can either act on the maxim 'To do something to promote x,' the maxim 'To do everything to promote x,' or adopt no maxim at all with respect to x. This has the result that Nell wants to assure, i.e. that it does not entail that it is obligatory to adopt the maxim "To do everything to promote x." Again each case must be examined individually in order to determine whether

85

some policy with respect to r is required. Further, if some policy is required, it need not be more than that the maxim 'To do something...' be adopted.

This analysis also allows us to deal with the problem that Nell sees with non-reciprocal action. It is this problem that leads her to conclude that maxims such that they and their contrary are both forbidden are permissible. Laws such as "everybody will buy lettuce and not sell it" are clearly self-contradictory, and so are forbidden. Nell then argues that the contrary of this UTC, which she takes to be "everybody will sell lettuce and not buy it" is also self-contradictory. Our intuitions, however, tell us that neither maxim is forbidden but that both are permissible. However, she seems to be clearly mistaken in her claim of which maxim is the contrary of "To buy lettuce and not sell it." The real contrary must be "Not to buy lettuce and not sell it" and its UTC "no one will buy lettuce and not sell it" is not self contradictory. But now, according to her earlier interpretation, acting on the maxim "Not to buy lettuce and not sell it" is obligatory. Thus, if one buys lettuce, one must also sell it.[25] That this is not in fact the case can be seen from our reinterpretation of the test for obligatory maxims. It is not an automatic procedure, and each case must be examined in order to see whether some policy must be adopted with respect to adopting a maxim. All that is required by the maxim 'To buy lettuce and not sell it' being forbidden is that I not act on it. In most circumstances I will have no need to adopt a single policy governing both buying

and selling lettuce, but I can merely act on the permissible maxims 'To buy lettuce' and 'To sell lettuce.'[26]

Once we have established whether acts and ends are obligatory, forbidden or permissible, we must ask whether they are morally worthy, unworthy, or lacking in moral worth. This depends on the motive with which it is done, i.e. the agent's end in so acting. Thus if an agent performs an act that is determined to be obligatory by the CC test, he may be doing it because he wants to, or because it is conducive to something else that he wants. Or he may do it purely because of his recognition that it is obligatory: he may do it out of duty. It is only in this case that the act has moral worth. The same considerations apply to acts that conform to the rational ends of all agents: the acts are morally worthy only if they are motivated by these ends. According to Kant, moral worth depends on the agent's motives, regardless of the results of his act, and such an intention must be based on a rational recognition of what is required and not a contingent, empirically given inclination, as this can never provide a solid foundation for right action.

So far we have looked at Nell's interpretation of Kant's universalizability tests. If we include certain reasonable presuppositions, i.e. that intending a maxim or a law entails intending its natural and predictable results, and some set of circumstances that allow it to be acted on, then the CC test works. If we allow that persons have ends, and that the Principle of Hypothetical Imperatives holds, then the CW test works too. Both tests are effective in that they are easily applicable

and they give acceptable results. Even if other commentators argue for a different interpretation of Kant's writings, the tests as Nell presents them work. They do not lead us into any of the difficulties Wiggins criticizes in Kant and in Hare.

Kant's goal in the *Groundwork* is to explicate the common notion of morality. As was shown in Chapter 1, the conception of a duty that is binding irrespective of contingent desires or circumstances is the given that Kant explores. What Kant does is formalize this notion. From the form that ethical duty has, he derives a formula that allows us to determine which acts conform to this duty. It can show what actions are perfect duties, such that specific actions or omissions are always required; and it can also show that we have duties to adopt certain ends, such as benevolence, although it does not say precisely how to act according to these imperfect duties. However, there is one thing that Kant's account so far has not done, and that is to justify the Categorical Imperative itself. Although we may agree that the Categorical Imperative is derived from the common concept of ethical duty, there is no argument that our concept of ethical duty has any basis, rational or other. Although the results generated by the Categorical Imperative generally conform to our moral intuitions, this is not a justification for accepting it as a guide to moral action. After all, if we take our intuitions as the authority in morals we do not need to check them by reference to any fundamental principle.

Kant himself argues that moral thinking arises necessarily from the faculty of practical reason in its pure application, and that the reason that everyday moral

thought conforms so closely to what is discovered by philosophical inquiry into its nature is that practical reason *is* practical and therefore actually used. I want to enquire into the possibility of justifying the Categorical Imperative on different grounds, grounds that can take into account what we have learned from Wiggins about the nature of moral systems.

The starting point for any ethical inquiry is to ask what it is that we are doing when we engage in moral discourse. It is from a description of what morality as we know it *is* that we can come to answer the more specific questions about ethics that are often asked. And as was mentioned, Kant, in the *Groundwork*, does just this. He examines the common notion of morality, formalizes it, and thus shows us how to act in accordance with justice and virtue. I want to argue that his theory can be seen as a formalization of the concept that Thomas Nagel argues is the basis of altruism, i.e. the idea that each person has of being one among many others who are equally persons.[27] I want to suggest, however, that this concept can provide the basis for all moral principles.

This idea can most easily be seen as expressed in Kant's version of the supreme principle of morals that enjoins us to treat others as ends. This of course raises the point that the conception of oneself as one person among others can lead to our treating others in certain ways only if we see ourselves as deserving of that treatment *qua* persons, or *qua* persons under some description that applies to others as well. Kant makes this claim when he says that "man necessarily thinks of his own

existence in this way," i.e. as an end in itself, but as this has a rational ground that each person recognizes in his own case, the claim that each person is an end in himself can be seen as objective by others.[28]

The bare conception of oneself as one among others can be seen in the requirement that maxims be universalizable. The idea that everyone is equally a person leads to the requirement that whatever one does must be able to be done by everybody who is relevantly similar.* And the criterion of relevant similarity is whatever characteristics are mentioned in the maxim. Thus maxims that cannot be acted on by everybody who falls under that description cannot be acted on by anybody, because there is no principled way of distinguishing between those who may and those who may not.

The recognition of each person of being one person among others who are all equally persons seems to me to be fundamental to ethics. It also provides an independent justification of any ethical principle that is legitimately derived from it because, if we follow Hare and dismiss solipsism as a problem for all philosophy, it is an undeniable proposition.

Seeing others as equally persons is not to say that we give to their interests a concern equal to that which we give to our own; that would be the case only if

---

* Paul Dietrichson notes that the awareness that everyone must be able to act on a maxim that I am proposing to act on is given to me by my reason ("When is a maxim fully universalizable?", p.156.) The conception of oneself as one among many supplies the missing premise.

we expected the same from them in virtue of our being persons. What it does show is that we owe them the same respect or non-interference or fair dealing that we expect from them in virtue of our being persons, that is, more or less according to our nearness or distance from them. This is the meaning of "do unto others as you would have them do unto you." You would not expect a stranger to buy you a gift for your birthday, but you would expect him not to run you down in the street. The Golden Rule has two parts and attention is usually focused on the second. However, the first part of the rule must also be considered. After all, in deciding how you would be treated you are formulating a rule of conduct that you must then apply to yourself. This provides a check on the possibility of requiring too much of others, as the same will be required of you. It does not require that equal weight be given equal interests, nor does it rule out self-referential altruism as morally acceptable. This is seen clearly in connection with Kant's duty of benevolence: there are no constraints on how or to whom we are to be benevolent.

The CC test forbids acting on maxims that could not be acted on if everybody acted on them. Those proposing to act on such maxims require that a system be in place such that they are exceptions to this system. So Kant shows, for example, that given the institution of promising it is not permissible to make oneself an exception, but does not show that there must be such an institution. Because of this, different results will arise for different cultures that have different institutions. For example, there will be no rule against taking things in a society where there is

no private property. Different moral codes can then be seen to follow from applying the fundamental moral idea to the world as it is at different times and places.

Kant's rule for morally worthy acts requires that they be done for the sake of acting according to the moral rule, and not for some subjective end. The notion of being one person among many is not only an idea that must necessarily occur to anyone thinking rationally about the world and her place in it, it is also a view of the world such that morally worthy acts follow from it, independently of any end that such acts may achieve. Therefore, if people come to see moral principles as good - and they do - they can do so without reference to any ends that acting on such principles can bring about. This conforms with Wiggins' stipulation that for a moral theory to work it must be seen as good in itself. However, it is also compatible with the idea that morality does have a purpose in the maintenance of a system in which individual ends are most easily realised.

We have seen that this conception can be the basis of an answer to "what, morally, shall I do?" and that it can also provide an answer to the question "Why should I act morally?" Now we shall see if acting according to it can serve the purpose of morality. For utilitarians the answers to these three questions are the same, which is part of their problem, as Wiggins points out.

According to Wiggins, morality, in order to work, cannot be seen as having a purpose. People refer their moral judgements to a consensus that they see as possible in the responses people have when faced with various situations, and

92

change their principles if they come to see that their responses should be other than what they are. It is not clear to me, however, if there is no rational basis of morality, in terms of *what* an existing morality would be revised. Further, how do we distinguish specifically moral behaviour from other types of consensual social behaviour unless simply because it so called. There must be a reason for classifying some social rules as moral and others as, for example, manners.

J.L. Mackie, although he sees morality as just what *is* accepted as such in a society, sees it as a consensus about what rules must be in place in order to counteract people's naturally selfish tendencies so that social life, and the benefits to each person that are unattainable otherwise, is possible.

Even if we take th's to be the purpose of morals, in making individual moral judgements we do not look forward to some end that morality is supposed to achieve in order to answer the question of what to do in terms of what the best way is to achieve that end. We decide what to do by referring to the conception we each have of being one among other persons, and this naturally leads to a state of affairs that is at the same time fundamentally good for each person, i.e. that of being considered equally a person by everyone else who considers herself a person, and one in which each person can flourish and best achieve whatever ends she sets for herself. There is a parallel here with Kant's insistence that although one can feel wonderful by acting from duty, one can never act from duty in order to feel wonderful.[29]

93

The questions that are answered by the Categorical Imperative are those of moral obligation and not those of what might be called moral value. Mackie divides morality along these lines into narrow and broad morality.[30] * Narrow morality is concerned with the specific rules of conduct that are meant to "counteract limited sympathies" in our conduct toward others, so that our own interests and inclinations do not run away with us. If we already had such regard for other people, there would be no need for moral rules.[31] Broad morality concerns values that guide action in general. There are an indeterminate number of equally valid ways of living, and choices must be made independently as there are no hard and fast rules for most of the decisions that we have to make.

The arguments in this chapter apply to narrow morality. We do, however, need to point out that some problems that are commonly considered moral belong to broad morality and so cannot be decided by the tests we have set out. The decision to save my child, mentioned in chapter three, is a moral decision in the broad sense. The feelings such as a mother has for her child are valued in our society, but they are beyond the domain of narrow morality. This is what Kant sees when he says that something done out of inclination has no moral worth: there is no point in talking about a duty to do what you most want to do. Bernard Williams fails to see this point: saving a loved one is neither a morally worthy act nor an

---

* Bernard Williams makes a similar distinction between what he calls ethics and morality in *Ethics and the Limits of Philosophy*.

obligatory one. It is certainly permissible and conforms with our duty of beneficence, but it is not done out of a moral motive but because of feelings that one *contingently* has. These are feelings that Mackie calls "self-referentially altruistic." He takes these feelings to be a fundamental aspect of human life, and sees morals, in the narrow sense, as a way of counteracting, among other things, the negative effects that sometimes arise as a result of this tendency. It seems to me that this is why it is important, as Kant insists, that morality depend on reason and not on inclination. This is true when establishing the supreme principle of morals; when applying it, i.e. when judging what is morally *right*; and when establishing the agent's motive, i.e. when judging what is morally *worthy*.

Mackie further claims, however, that (narrow) morality is made and not discovered, yet says that the obligation that people perform, in Hobbes' terms, their covenants made is "an eternal and immutable fragment of morality."[52] How can this be if this obligation is simply made and not in some way required by the notion of morality itself, and thus in some sense discovered? Perhaps there is some set of principles that do best what morality is supposed to do, and would be accepted by everyone, and this would then be what morality *is* regardless of what is considered moral by a given society. This is close to what Wiggins understands by the consensus that grounds morality, but because our notion of consensus is based on reason we have been able to set out the method for discovering what that consensus would consist in.

Once a person finds herself in society, with certain interests and a notion of what is due to her *qua* person, she must rationally consider other persons in the same light. What follows from this is respect, justice, benevolence; and what follows from these are conditions under which communal, and therefore individual, human life can flourish. I suggest that on a charitable interpretation, the Golden Rule can be understood as a reminder that one must treat others as one would be treated *qua* person. It can therefore serve as a moral rule of thumb in everyday life, where people are not likely to actively apply the Categorical Imperative.

Now we have seen that taking the concept of oneself as one among others as the fundamental concept of morals can do three things: it answers the question "why be moral?" because it is a concept that reason imposes on us; it answers the question "what, morally, shall I do?" when it is applied formally via the Categorical Imperative; and it fills the purpose that morals can be seen as having because it makes it possible for individuals to live together in spite of divergent, and even conflicting, aims.

Notes

1. Kant, p. 421.

2. Wiggins, p. 68.

3. Paul Dietrichson, "When is a Maxim Fully Universalizable?" *Kant-Studien*, 55 (1964) p. 159.

4. Nell, p. 13.

5. Ibid, p. 35.

6. Ibid, p. 61.

7. Ibid, p. 62.

8. Kant, p. 424.

9. Nell, p. 69.

10. Ibid, p. 66.

11. Ibid, p. 69.

12. Ibid, p. 69.

13. Ibid, p. 70.

14. Ibid, p. 83.

15. Ibid, p. 85.

16. Ibid, p. 87.

17. Kant, p. 417.

18. Nell, p. 87.

19. Ibid, p. 88.

20. Ibid, p. 86.

21. Ibid, p. 86.

22. Nell, p. 86; my italics.

23. Ibid, p. 75.

24. Ibid, p. 88.

25. D. Davies, "Nell on Maxims of Nonreciprocal Action" (Unpublished) pp. 9-10.

26. Ibid., p. 12.

27. Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970) p. 88.

28. Kant, p. 429.

29. Paul Dietrichson, "What does Kant mean by 'Acting from Duty'?" *Kant: A Collection of Critical Essays* (Garden City: Doubleday and Company, 1967) p. 326.

30. J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin Books, 1977) p. 106.

31. Mackie, p.110.

32. Ibid, p. 123.

Conclusion

When thinking philosophically about morality, there are several different though related questions that can be asked. We can ask what the point or purpose of morals is. Or we can ask what it is to think morally or what it is we do when we make moral claims. We can also ask why we should be moral. Or we can ask what it is specifically that we are morally obliged to do. Those who think that there is an answer to the last question often seek a unifying principle or test that allows us to discover what actions are permitted, forbidden, or obligatory. Further, in order to answer this last question, we need to take into account answers to the first two. If we are to discover which rules are moral rules, we need to know both what moral rules are for, and what features a moral rule must have. Ideally, along the way we will also find an answer to the question of why these rules are rules for us.

The recognition that part of what moral judgements say is a claim to being true, and being recognizable as true by anyone who thinks clearly about them, is what all moral theories strive to explain and express. The mystery is of how a judgement that is about something not "in the world" can make this claim of objectivity. This question has been variously answered by having it turn out that moral judgments *are* identical with statements about the world, e.g. moral sentiments, satisfaction of preferences, social mores, etc. Wiggins argues that this perceived objectivity can be explained by the intersubjectivity of moral judgements,

that is by their being based on subjective responses that are, or should be, common to all. For Hare talk of objectivity is better replaced by talk of rationality, and I agree. If everyone reasons rationally from the same premises, they will all arrive at the same answer. This can then be seen as objectivity, or consensus, or convergence. My contention is that the moral point of view stems from the acceptance of one basic premise that can be seen to underlie every moral judgement, in the sense that it can provide an intelligible answer to the question "why...?" This premise is that "I am one person among others who are all equally persons."

I have argued that Kant's Categorical Imperative provides a decision procedure for moral action. It is a formalization of the common sense conception of duty, and this can be seen as based on the premise mentioned above. The Categorical Imperative allows us to determine what specific maxims we may or may not act on because it shows which maxims it is not possible to will *everyone* to act on. If one sees oneself as having a reason for acting in a certain way *qua* person under a certain description, then the conception of oneself as one among others who are equally persons requires that one recognize that others have that same reason for acting. One cannot logically make an exception of anyone because there is no principled way of making a distinction, therefore if not everyone can act on that reason, then no one may.

The test that will show which maxims cannot be acted on by everybody must show a logical impossibility, allowing for consideration only the empirical

information referred to -explicitly or implicitly - in the maxim. Wiggins claims that this is inconsistent with Kant's goal of ascertaining specific duties without reference to contingent empirical concerns. Wiggins specifically refers, however, to people's preferences, desires, and inclinations, and although he is right in saying that to take account of these would subvert Kant's enterprise, these are not the empirical facts required by Kant's tests of maxims. It cannot appeal to particular inclinations in order to show that there is *one* circumstance under which I cannot will that it be acted on, and therefore cannot will that it be acted on in *every* like circumstance. This is the method that Hare tried and we have found wanting. Hare's theory fails because it does not adequately reflect what it is that we take moral thought to *be*, and he does not provide arguments compelling enough for us to change our view.

Basing the moral point of view on the conception of oneself as one among many who are equally persons answers Wiggins' concerns about morality being seen as good in itself and bound up with interests that each person has. It is well fitted to serve the purpose of making things go well and fairly, but anyone can come to accept these premises simply by reflecting on his situation and that of others, and come to see adopting this point of view as good in itself. It answers our intuitions about what morality is and does: we can see the difference between more and less moral people or societies as those to whom this conception of others is more or less real. And it can be seen as an answer to the problem of how people who live together in society can get along in a way that is satisfactory for each. I suggest that

101

the recognition each person has of being one among others who are equally persons can serve as the rational basis for an adequate theory of universalizability.

Bibliography

Allison, Henry E. *Kant's Theory of Freedom*. Cambridge; New York: Cambridge University Press, 1990.

Davies, D. "Nell on Maxims of Nonreciprocal Action" unpublished.

Dietrichson, P. "When is a Maxim Fully Universalizable?" *Kant-Studien*, 55 (1964), pp. 143-70.

Dietrichson, P. "What Does Kant Mean by 'Acting from Duty'? *Kant-Studien*, 53 (1962), reprinted in Wolff, pp. 314-330.

Foot, Philippa. "Utilitarianism and the Virtues." *Mind* vol. 94, 1985, pp. 196-209.

Gauthier, David. "Maximization constrained: the Rationality of Cooperation." *Rationality in Action: Contemporary Approaches*. Ed. Paul K. Moser. Cambridge: Cambridge University Press, 1990, pp. 315-334.

Gewirth, Alan. *Moral Rationality*. Lawrence: University of Kansas Press, 1976.

Gewirth, Alan. *Reason and Morality*. Chicago: Chicago University Press, 1978.

Gewirth, Alan. *Human Rights*. Chicago: University of Chicago Press, 1983.

Gewirth's Ethical Rationalism: Critical Essays with a Reply by Alan Gewirth. Ed. Edward Regis Jr. Chicago: University of Chicago Press, 1984.

Haezrahi, Pepita. "The Concept of Man as End-in-Himself," reprinted in Wolff, pp. 291-313.

Hare, R.M. *The Language of Morals*. New York: Oxford University Press, 1964

Hare, R. M. *Freedom and Reason*. New York: Oxford University Press, 1963.

Hare, R. M. *Moral Thinking: its Levels, Method, and Point*. Oxford: Clarendon Press, 1981.

Kant, Immanuel. *Foundations of the Metaphysics of Morals*. Trans: Lewis White Beck. New York: The Liberal Arts Press, 1959.

Kant: A Collection of Critical Essays. Ed. Paul Wolff. Garden City, New York: Doubleday. 1967.

Kemp, J. "Kant's Examples of the Categorical Imperative." *Philosophical Quarterly*, 8 (1958), reprinted in Wolff, pp.246-258.

Kolnai, Aurel. "Moral Consensus." *Proceedings of the Aristotelian Society*, 1969-1970. pp. 93-118.

Korner, S. *Kant*. Harmondsworth, Middlesex: Penguin Books Ltd., 1955.

Macdonald, Margaret. "Natural Rights." *Proceedings of the Aristotelian Society*, 1947-1948, pp. 40-60.

Mackie, J.L. *Ethics: Inventing Right and Wrong*. Harmondsworth, Middlesex: Penguin Books Ltd, 1977.

Midgley, Mary. "Trying Out One's New Sword." *Heart and Mind*. New York: St Martin's Press, 1981, pp. 159-165.

Mill, J.S. *Utilitarianism*. ed. George Sher. Indianapolis: Hackett Publishing Company, 1979.

Nagel, Thomas. *The Possibility of Altruism*. Oxford: Clarendon, 1970.

Nell, Onora. *Acting on Principle: an Essay on Kantian Ethics*. New York: Columbia University Press, 1975.

Wiggins, David. "Universalisability, Impartiality, Truth." *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Basil Blackwell, 1987, pp. 59-86.

Wiggins, David. "Truth, Invention, and the Meaning of Life." reprinted in Wiggins, 1987, pp.87-137.

Wiggins, David. "A Sensible Subjectivism." in Wiggins, 1987, pp. 185-214.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press, 1985.