The modern Aristotelian, less inclined to discount inferiors and outsiders than Aristotle himself, can fight back. She can say that such cases need sustaining by rationalizations, and these rationalizations will mainly consist in lies the privileged tell themselves. And we already conceded that a life lived amidst lies, or in a fool's paradise, is not a flourishing life. So the ingredients are there to suggest that *real* flourishing or *true* human health implies justice. It implies removing the oppression, and living so that we can look other people, even outsiders, in the eye.

However, this need for rationalizations is itself not a given. Sometimes, as we go our careless ways, we do not even seem to need lies to sustain us. Our generation may flourish by consuming all the world's resources, and letting the future go hang. We do not tell ourselves a story according to which the generations to come are inferior to us and deserve to inherit a deadened world. We just don't think about it. It is only when we have to have a conversation with the dispossessed that we scramble for rationalizations.

Are we being 'unreasonable' as we discount or forget about dispossessed outsiders? We are certainly failing in benevolence, and we may be failing in justice (more on this below). But even if we concede much to the Aristotelian argument, we might remain pessimistic about its effect. Insofar as it works by 'pumping up' what is required for a life of reason or a life of *true* flourishing, we will find people perfectly ready to settle for a good fake. Better to buy the cheap running shoes and not to think too much about how they were made. To unsettle such people we will need, eventually, to look further at the motivation to justice.

18. The categorical imperative

Hume's challenge to Reasons (section 16) was taken up by Immanuel Kant. We can approach Kant's views by thinking of a common gambit in practical discussion. When we try to stop people acting in some way, a good question is often: 'What if

everybody did that?' The test is sometimes called a 'universalization' test. If the answer is that something would go especially wrong if everybody did that, then we are supposed to feel badly about doing it. Perhaps, for instance, we would be claiming an exemption for ourselves that we couldn't allow to people in general.

Kant picked up the universalization test and ran with it. In his hands it became not only a particular argument *within* ethics – a device, as it were, for making people think twice, or feel guilty – but the indispensable basis *for* ethics. It became the foundation stone for ethics, grounding ethics in reason alone. It gives us Reasons, even in the domain of prescriptions or imperatives. He unveils the way this happens in his short masterpiece, the *Groundwork of the Metaphysics of Morals* of 1785, a work that has probably inspired more love and hatred, and more passionate commentary, than any other in the history of moral philosophy.

The universalization test can sound like a version of the Golden Rule: 'Do as you would be done by' – a rule sometimes claimed by Christianity as its own, but found in some form in almost every ethical tradition, including that of Confucius (551–479 BC). Kant denies that his idea is just that of the Golden Rule. It is supposed to have more meat. He points out, for example, that the Golden Rule can be misapplied. A criminal can throw it at a judge, asking him how he would like it if he were being sentenced – yet the sentence may be just, for all that. A person in good circumstances may gladly agree that others should not benefit him, if he could be excused from benefiting them. He apparently abides by the Golden Rule. So something with more structure is needed.

Kant starts by distinguishing what he wants to talk about from what he calls 'talents of the mind', such as understanding, wit, or judgement, and from advantages of temperament, such as courage or perseverance or even benevolence. He also distinguishes it from gifts of fortune, happiness, and even admirable qualities such as moderation. None of these are 'good in themselves'. For all of them

can be misused, or can be lamented. Even happiness is not admirable, if it is the happiness of a villain. Benevolence may lead us astray, letting other people enjoy what they have no right to enjoy, for example. And 'the very coolness of a scoundrel makes him not only far more dangerous but also immediately more abominable in our eyes than we would have taken him to be without it'.

The only thing good in itself, then, is a good will. Even if the agent with the good will is handicapped, 'by a special disfavour of destiny or by the niggardly endowment of stepmotherly nature', from actually doing much good in the world, still, if he has a good will, it will 'shine like a jewel for its own sake'.

But what is a good will? Kant considers cases of people doing good things, things that might even be their duty, not, however, from a sense of duty, but from other inclinations, such as self-interest, or even benevolence, or a sense of vanity. A salient example is a shopkeeper who does not overcharge an inexperienced customer, but only because his self-interest is served by not doing so. Perhaps he calculates that the customer is more likely to return, or that his shop will profit from a good reputation. The shopkeeper behaves honestly enough, but not because he has the right feeling that he *ought* to do so. There is no jewel shining by itself here. This is not the good will in operation. So what is?

The shape of the answer becomes clear from such examples. The good will is one acting from a particular good motive. It is one acting out of a sense of law or duty. 'Duty is the necessity of an action from respect for law.' We are able to represent laws of action in ourselves, and a good will is one that acts in accordance with that representation. The core of morality, then, lies not in what we do, but in our motives in doing it: 'When moral worth is at issue, what counts is not actions, which one sees, but those inner principles of action that one does not see.'

This is all very well, we might say. Kant seems to be praising up the

conscientious agent, or the agent of principle or righteousness or rectitude. This is a person who, once he thinks 'Such-and-such is a duty' is strong-minded or principled enough not to be deflected from doing it. This is in some respects an admirable psychology, although it is also one that can do a lot of harm, since people's consciences can be as perverted as anything else. One wonders why righteousness in this sense is exempt from the criticism levelled at benevolence and the rest, that it can be a Bad Thing.

Some writers also remind us that in many of life's situations, rectitude is not what we want. We often want people to act out of love or gratitude, not out of duty. Good parents take their child to an entertainment because they enjoy the child's pleasure; a parent who takes the child out of a sense of duty is to that extent lacking. A lover who kisses out of a sense of duty is due for the boot. But this is not a fundamental criticism of Kant. He can, and does, allow dimensions in which the good-hearted parent or lover or benefactor scores highly. It is just that these are not, for him, the *moral* dimensions. Moral excellence is found only in the strength of the sense of duty.

There is a more fundamental difficulty. Kant's answer seems to demand that certain things got onto a list of duties *in the first place*. It is no good saying 'Act from a sense of duty' if when asked the question 'And what is my duty?', the only reply is 'To act from a sense of duty!'

We have to break out of the circle somewhere, and so far we do not know how. So how is it all going to get us nearer to the foundations Kant promises? His move is breathtaking, both in its speed and its result:

> *But what kind of law can that be, the representation of which must determine the will, even without regard for the effect expected from it, in order for the will to be called good absolutely and without limitation? Since I have deprived the will of every impulse that*

could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, I ought never to act except in such a way that I could also will that my maxim should become a universal law.

This is the famous Categorical Imperative, or more accurately, the Categorical Imperative in its first form, the so-called Formula of Universal Law. Later on Kant glosses it in other ways. One is 'Act as if the maxim of your action were to become by your will a *universal law of nature*' (the Formula of the Law of Nature). Another, possibly the most influential, is 'So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means' (the Formula of Humanity). It is not at all clear that these different versions can be derived one from the other, but Kant regarded them as somehow equivalent.

The promise is that we have here both quite substantial moral principles, or versions of the one principle, and principles that have been proved by reason alone. This last claim is hard to make good, but perhaps the idea goes like this.

As Hume illustrates, we might suppose that there are no Reasons in the area of ethics – just the desires or wills of particular persons, not necessarily shared or respected by anyone else. But Kant replies that the very formal nature of the Categorical Imperative gives it a universal authority. You cannot flout it and defend your principle in doing so. If you do flout it, you declare yourself to be un-Reasonable. If this is right, we have the required foundation: ethics comes from Reasons alone.

Unfortunately, when it comes to applications of the principle, things become a little stickier. The most persuasive examples of the Categorical Imperative doing some real work are cases where there is an institution whose existence depends on sufficient performance

by a sufficient number of people. Suppose, as is plausible, that our ability to give and receive promises depends upon general compliance with the principle of keeping promises. Were we to break them sufficiently often, or were promise-breaking to become a 'law of nature', then there would be no such thing as promise-giving or promise-breaking, because no words could any longer have the required force. So, Kant considers somebody whose principle of action is, 'Let me, when hard pressed, make a promise with the intention not to keep it.' Then, says Kant, I could will the lie, but I could *not* will the universal law to lie, for in accordance with such a law there would be no promises at all. It would be willing a kind of contradiction. So we have a Reason against the lying promise.

That's all very well, but consider a person who is against the whole business of giving and receiving promises. Why shouldn't he try to undermine the institution from within: by giving false promises, with one of his aims being the breakdown of trust and cooperation? Of course, a nice or benevolent or even a prudent person wouldn't have that goal, but if Kant appeals to these virtues, the purely formal appearance of his theory begins to vanish. We only have a reason against giving the lying promise, not a Reason.

An example I like here is the institution of credit cards. These depend on enough people not paying them off each month in order to keep profits coming in to the issuing banks. So there is a kind of contradiction in imagining a world with credit cards, but where everybody pays them off each month. Suppose my principle is, 'Pay off your card whenever you feel like it.' Can I 'universalize' this, willing it to govern people in general? Surprisingly, perhaps, yes. Even in a world where people can always afford to pay off their cards, we might have it that everyone pays off their card when they feel like it. This could be true provided they don't often feel like it, for instance because for most people most of the time the urge to consume is greater than the urge to save. So on the rare occasions when someone feels like paying the card off in full, she can go ahead

and do so without falling foul of the Categorical Imperative. Similarly, then, a person can consistently adopt principles of the kind 'Lie/break a promise/steal/cheat on taxes whenever the situation is this serious', provided the situation isn't *very often* that serious. The institutions survive, and so do the possibilities for making exceptions.

A third limitation appears if we consider the man mentioned above, who misapplies the Golden Rule, saying that he does not mind others refraining from benefiting him, provided he can be excused from benefiting them. Kant's only argument that he fails the Categorical Imperative test is that he *might* get into dire straits in which he needs the assistance of others. But this evidently invites the all-too-human rejoinder that he might not, and is willing to take the risk. He can will that nobody help anybody else, because he can gamble on staying self-sufficient.

Kant descends somewhat from the abstract heights of the Formula of Universal Law version of the Categorical Imperative. He argues in effect that the capacity of human beings to act in accordance with the imperative – the jewel within – is itself a thing of absolute, unconditional value. It is true, he thinks, that we can never be sure that we are acting from our sense of duty alone, since our motives are often mixed and often hidden from us. But at least we can set ourselves to do so. We can distance ourselves from our mundane desires and wishes, and set ourselves to act as duty requires. This capacity itself gives us our fundamental title to respect and self-respect. We are proud of our reasonings – in fact, whenever we offer reasons we are showing how much we respect reason in ourselves. So it deserves respect wherever it is found, that is, within all rational agents.

This argument (or something like it: the texts are dense) takes Kant to the Formula of Humanity: 'So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.' It is not, of course, easy to

see exactly what this involves, but the general idea of remembering to respect each other is clearly attractive, and perhaps more practicable than remembering to love each other. Whether we deserve respect purely because of our capacity to make laws to ourselves is a good deal less certain. Perhaps we deserve respect from each other insofar as we are like each other in a whole mass of ways. The raiding party bent on enslaving a rival group has forgotten a shared humanity, which includes a shared capacity to love, and suffer, and hope, and fear, and remember. It hasn't *only* forgotten that the victims can reason according to general rules.

Many people think Kant offers the best possible attempt to find Reasons, and therefore to justify ethics on the basis of reason alone. Since many people want such an attempt to succeed, and fear the result if it does not, there are major intellectual industries of trying to find ever more complicated interpretations of the approach that make it work. It might be doubted whether this does much service to Kant: he was a great democrat, and believed that the necessity of the Categorical Imperative was easily visible to any reasoning creature.

19. Contracts and discourse

Some writers think that a descendant of Kant's approach, often called 'contractarianism', gives us a powerful foundation for ethics, or at least for the large part of ethics that concerns our rights and duties to each other. One formula at the centre of recent work is this, due to the contemporary American philosopher T. M. Scanlon:

> *an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement.*

As in Kant, there is a concern for the universal, and a concern for reason. A slightly different version occurs in the 'discourse ethics' of