Moral agents

Moral agents are those agents expected to meet the demands of morality. Not all agents are moral agents. Young children and animals, being capable of performing actions, may be agents in the way that stones, plants and cars are not. But though they are agents they are not automatically considered moral agents. For a moral agent must also be capable of conforming to at least some of the demands of morality.

This requirement can be interpreted in different ways. On the weakest interpretation it will suffice if the agent has the capacity to conform to some of the external requirements of morality. So if certain agents can obey moral laws such as 'Murder is wrong' or 'Stealing is wrong', then they are moral agents, even if they respond only to prudential reasons such as fear of punishment and even if they are incapable of acting for the sake of moral considerations. According to the strong version, the Kantian version, it is also essential that the agents should have the capacity to rise above their feelings and passions and act for the sake of the moral law. There is also a position in between which claims that it will suffice if the agent can perform the relevant act out of altruistic impulses. Other suggested conditions of moral agency are that agents should have: an enduring self with free will and an inner life; understanding of the relevant facts as well as moral understanding; and moral sentiments, such as capacity for remorse and concern for others.

Philosophers often disagree about which of these and other conditions are vital; the term moral agency is used with different degrees of stringency depending upon what one regards as its qualifying conditions. The Kantian sense is the most stringent. Since there are different senses of moral agency, answers to questions like 'Are collectives moral agents?' depend upon which sense is being used. From the Kantian standpoint, agents such as psychopaths, rational egoists, collectives and robots are at best only quasi-moral, for they do not fulfil some of the essential conditions of moral agency.

1 Agents versus recipients

Moral agents should be distinguished from moral recipients (see Moral standing). Moral agents are those who are morally accountable for at least some of their conduct. They are subject to moral duties and obligations, and, therefore, to moral praise and blame. Moral recipients are those who are owed moral consideration for their own sakes (see Respect for persons). On certain views moral agents and moral recipients are coextensive. Thus according to Kantians persons are the only moral agents and the only moral recipients (see Kant 1785). We should not be cruel to animals, not because animals are owed anything for their own sakes, but because cruelty to animals may indirectly harm persons who are the only ends in themselves (see Animals and ethics). According to utilitarians all sentient beings are owed consideration to the extent that they have feelings; so on this view there are moral recipients who are not moral agents - for instance, animals, or at least their feelings. Can there be moral agents who are not moral recipients? It would seem not, unless perhaps one uses moral agency in a weak sense which includes among moral agents nonsentient entities such as robots and corporations.

2 Understanding

The view that moral agents must have the capacity to conform to some of the external demands of morality is consistent with the view that there are parts of morality that they cannot conform to. Thus kleptomaniacs do not have the capacity to conform to certain moral requirements about not stealing, but it does not follow that they are not moral agents or that they should not be held morally responsible for murders that they might commit (see Responsibility). Moral agents must be morally responsible for some of their conduct, not necessarily for all. What, then, are the conditions of moral agency?

One essential condition is that the agent must have the relevant understanding (or capacity for understanding) of what the external requirements of morality are (see Moral knowledge §1). Thus in the case of murder one must understand that murder is wrong and that a particular act is an instance of murder. Exactly how much understanding is required is not easy to specify. There are plenty of borderline cases, but there are clear cases on both sides of the line. A baby does not have the relevant understanding of any of the requirements of morality, while an average adult citizen does at least sometimes. Of course even average citizens are very ignorant in many matters, but that at most is relevant to assessing their responsibility in these matters; it does not prevent them being

moral agents.

Some existentialist philosophers insist that moral agency requires the ability to create and choose one's own values, unconstrained by objective or rational considerations (see Existentialist ethics). It is objected that such creation involves a capricious freedom, since the agents have no guide as to how they should choose their values. Charles Taylor (1982) attempts to overcome this problem by suggesting that moral agency requires that one should have the capacity to choose one's values, after reflection, in accordance with one's deepest and most authentic nature (see Taylor, C. §§4-5). Does this requirement provide the necessary guidance for the agent? Critics would point out that this just shifts the problem. In what sense are we responsible for our deepest nature?

It seems that moral accountability requires that the agents should have an objective basis for choosing their moral values. They could then be held morally accountable to the extent that they have the capacity to find out what the relevant moral requirements are and, to the extent that they have the capacity, to conform to such requirements in the relevant ways. People who cannot reason properly (such as the severely mentally ill) or those who lack certain volitional abilities lack the capacity to conform to the relevant moral requirements.

In order to be morally accountable, an agent does not always have to know or even have the correct opinion about what the moral requirements are. The capacity for finding out such things can be enough. For instance, some Nazis who persecuted Jews may have thought sincerely that they were doing the right thing; but if they could and should have known better then they can be censured for moral negligence. Had they thought things through, which they could and should have done, they would have realized how wrong such acts were. Or so it is believed by those of us who want to hold them morally responsible.

To have the capacity to find out that something is morally wrong does not necessarily involve having the capacity to know why. It is plausible to distinguish having right opinions on moral matters from knowledge of them. The person who has knowledge in moral matters not only has the right opinions but also has them for the right reasons. Even after a study of moral philosophy many people do not know why things like stealing and murder are morally wrong; they do not understand the grounds for such judgments. True, theories have been advanced to answer such problems, but there is no general agreement on correct answers. An ideal moral agent, who exists only in the imagination of philosophers, might have knowledge of all moral matters, but ordinary moral agents have only opinions (for example, that stealing is wrong) and the capacity to find out such opinions in areas where they are held morally responsible.

According to some philosophers it is not enough to have the intellectual ability to tell right from wrong. There are psychopaths who are quite intelligent in general, and can even talk intelligently about morality. They might be able to tell us what things are wrong and even why, but they lack moral sentiments, such as remorse and consideration for others, and are unable to act for the sake of moral considerations. Many would say that they are not moral agents, and therefore not subject to moral condemnation, nor to punishment in so far as punishment presupposes moral condemnation. Bradley (1894) thought that psychopathic killers are not moral agents and so we cannot morally condemn them or punish them in the way we punish moral agents, but we have the right to use social surgery, even to kill them if that is necessary for social welfare, somewhat as we have a right to kill dangerous animals; considerations of justice do not apply to those who are not moral agents. Bradley forgot that considerations of humanity may still apply to them.

Bradley's view has been endorsed by Jeffrie Murphy (1972) but with two important qualifications. First, he points out that there is the danger of abuse in such a system. If we were permitted to go in for social surgery against nonmoral agents, moral agents might sometimes be wrongly diagnosed as nonmoral agents. Second, many people become psychopaths because of bad social conditions, and not of their own free will. Murphy rejects Aristotle's view that psychopathic wickedness is like a disease that people are responsible for acquiring by their voluntary conduct. He argues that psychopaths are those whose potential moral agency has been destroyed by society, and that they should therefore not be treated too harshly.

John Rawls (1971) maintains that only those who can give justice are owed justice (see Rawls, J.). And he stresses the importance for moral personality of acting from the sense of justice. On this view those who cannot act from a sense of justice would not be moral agents and so would not be owed duties of justice. But there is a weaker thesis according to which it will suffice to be owed justice if people can give justice even if they are not motivated by the

sense of justice. In the case of potential criminals, even if they are incapable of acting out of a sense of justice, or for the sake of moral considerations, we can apply considerations of justice to them and respect their rights if they respect the rights of others, even if their reasons for doing so are egoistic.

Indeed some Hobbesian philosophers contend that rational egoists can set up a just society, without the aid of a moral sense or a sense of justice (see Contractarianism §§2-3). Similarly, one could operate with a sense of moral agency, according to which agents have a capacity to conform to some of the external requirements of morality, but may lack the capacity to act for the sake of the moral law. They could sustain something like a morality. From a pragmatic point of view it would not matter too much why people conformed to moral requirements as long as they continued to do so. Jeremy Bentham (1817) thought that human beings were primarily egoistic, and that basic human nature was unalterable. He suggested that the setting up of the right institutions, such as representative democracy, and sanctions, such as punishment, would lead all members of the public, including the rulers, to see that contributing to the common good would be in their own best interests; thus rulers who acted against the public interest would be unlikely to be re-elected (see Bentham, J.). But would such rational egoists be moral agents? This is partly a verbal matter. They would satisfy the requirement that people should have the capacity to conform to some of the external requirements of morality, but they would not meet the Kantian requirement about being able to act for the sake of moral considerations.

3 The inner life and the Kantian view

The substantial question is whether the Kantian requirement is legitimate. Different moral theories give different answers to this question. Critics of the Kantian view might ask: if members of your family rescue you from a burning house, would you not think better of them if they did this out of affection for you rather than out of a sense of duty? Kantians would reply that if they did it out of affection, this would not at that time involve an exercise of their moral agency. This is consistent with the view that there may have been some exercise of moral agency in the past if the agent cultivated the right feelings and dispositions out of a sense of duty. Some people through good luck have better feelings and dispositions than others. Their qualities may even be wonderful, but do they deserve *moral* credit? Wittgenstein thought that G.E. Moore's lack of vanity and his innocence generally might be loveable but Moore did not deserve any moral credit, for he was not 'talking of the innocence a man has fought for, but of an innocence which comes from a natural absence of a temptation' (see Malcolm 1966: 80). But what of the family members who fight against a selfish temptation to run away from the burning house where you are, and overcome it not out of a sense of duty but because of their love for you?

Persons who, through no fault of their own, have temptations to commit serious crimes are at a serious disadvantage compared to those of us who are lucky and do not have these temptations (see Moral luck). The former may deserve moral credit for conquering their temptations and for attempting to cultivate the right dispositions. Moral agency in one important sense of the term is to be contrasted with what happens as a result of luck. The conduct of individual agents is produced by a combination of factors: heredity, environment and free will. On the Kantian view, since personal choice is the only contribution to conduct made by agents themselves, it is only for this last factor that they are to be held morally accountable (see Free will).

The Kantian view of moral agency presupposes an enduring self that has the power of acting freely in a strong libertarian (or nondeterministic) sense. The enduring self on this view is different from an enduring character. The self is that which has the character and is autonomous in the sense that it has the power after reflection to change or not to change the character to some degree (see Autonomy, ethical). It is only to the extent that it has this power that it is held morally accountable for actions that issue from it or its character. If there is no enduring self, if the later self is a different self from the earlier one, then the later self cannot be morally accountable for the acts of its predecessor any more than a child can be for the deeds of its parents. Critics point out that the free will and the enduring self that are presupposed are incoherent. Either the character is determined by various factors that are ultimately outside the agent's control or, if there is a break in the causal chain, then the act is a chance or random event and so again there can be no freedom.

Kantians reply that this objection loses force once we acknowledge the existence of the inner standpoint of the moral agent. C.A. Campbell (1957) has pointed out that from the internal standpoint we can make sense of a free act that is neither determined by factors outside our control nor a random event. Campbell appeals to our phenomenological experience of moral effort in the face of moral temptation, in which we can make sense of the

creative agency of the self, when the self has the power to go beyond its formed character. On this view what is really admirable about moral agents is that they can obey the moral law, rising above their feelings and passions by efforts of will made for the sake of the moral law. So the existence of the inner life is considered, at any rate on the Kantian view, to be another essential requirement of moral agency. It is this requirement that is not met by robots, corporations, states and other groups. Even if they instantiate rational systems or functional systems such that it makes sense to attribute actions (in a functionalist sense) to them, they do not have an irreducible inner phenomenology (see Functionalism). Thus a corporation or a state is not joyous and does not suffer (in the phenomenological sense) except in the sense that is reducible to the suffering and joys of its members. To say this is consistent with the view that such entities are extremely important in the influence they have over their members and that their behaviour and the laws governing them are not reducible to the behaviour and the laws governing the behaviour of their members.

William James (1907) pointed out that what gives significance to human life is that we can set ourselves ideals or goals and then pursue them with zest, overcoming obstacles in the way (see James, W. §3). If there were no struggle in human endeavour there would be nothing heroic about us. When we admire individuals for their struggle against temptation, or against disease, or their heroic attempt at conquering mountains or solving mathematical problems, we are appealing to an inner life. If a computer solves a mathematical problem, however ingenious its solution, there is nothing heroic about it.

It might be objected that one can understand the duty versus temptations battle without appealing to an inner life; one may give a dispositional analysis of temptation as well as of overcoming it. But Kantians would reply that we can only make sense of moral effort and free will if we assume an internal point of view. If we look at things from a purely objective point of view neither the presence of determinism nor its absence can make sense of our free will and of our moral agency.

4 Collectives and moral agency

There is sometimes disagreement about whether such categories as the insane, children, robots and collectives are moral agents. This disagreement can partly reflect different standards of moral agency being used. It can also be due to different views about the facts. Thus some of us may deny that robots are moral agents on the ground that they lack an inner life. Others may disagree with us on the grounds that an inner life is not essential to moral agency; and some may argue that even though an inner life is essential, robots of the sophisticated variety might be constructed in the future who have inner lives in the sense that human beings do.

People often wonder whether collectives such as nations, states and corporations are moral agents. On the assumption that they lack an irreducible inner life or that they lack a permanent or enduring self, they will not fulfil an essential condition for moral agency in the Kantian sense. But if we use moral agency in a weaker sense, then they too could be moral agents. Kantians would say that they are at best quasi-moral agents, but critics of the Kantian approach would complain that the Kantian view is too stringent; it has presuppositions that are unverifiable and controversial. Some of these critics prefer a practical solution where terms like free will and moral agency are used in a weaker sense.

On a Humean view people can be given moral praise if they meet the external requirement of morality by promoting social welfare, provided their conduct reflects their character (see Hume, D. §4). Similarly, they can be apportioned moral blame and punished if they produce social misery, provided their conduct stems from their character (rather than by accident). No permanent or enduring self in any deep sense is required. What is required is that the individual being blamed has the same character traits (in the relevant respects) as the one who committed the act. We are free to the extent that our conduct reflects our character. Deep problems like whether we are responsible for acquiring our characters are bypassed on this view. It implies that collectives too could be moral agents in the way that individual persons are.

Indeed Hume explicitly compared the person's self to a republic or a commonwealth. So the argument that groups are less eligible for moral agency than individual human beings collapses on this view. Individual human beings can have character traits that persist; but so can groups. And though groups do not have an irreducible inner life, neither do individual human beings on the Humean view, according to which the ultimate constituents are the items of experience. The person does not persist over time in any deeper sense than groups do. The person's

relation to the individual items of experience is like the relation of the republic to individual members who compose it (see Mind, bundle theory of). So on this view an individual can be morally accountable in much the same sense as collectives; this is one of the corollaries of his view that Hume did not notice.

Much communal violence presupposes ideas of collective moral agency. For instance, during the partition of India, when some Muslims in one part of the country persecuted Hindus, another group of Hindus would take retributive actions against some other Muslims in another part of India. The Hindus would often justify such conduct on the grounds that the people they attacked shared common characteristics with their coreligionists who committed the evil deeds elsewhere. The Muslims would use a similar justification against the Hindus and the vicious circle would continue. Jonathan Edwards (1758) thought that human beings are now morally accountable for the sins of Adam, for they belong to the same collective, humanity, and share the same fallen nature as Adam (see Edwards, J.). On the Kantian view (as well as on some other individualist views) this is not enough. Persons now are separate moral agents and have a separate centre of consciousness from Adam, even if they have similar character traits; the self is different from and transcends the character. Such requirements are often violated by those who mete out retribution at a collective level.

But we do seem to make moral demands upon collectives. Thus one might say that the IMF (International Monetary Fund) ought to provide more facilities for the poorer countries. And it has been claimed that some of these demands are not reducible to demands upon individuals. If this claim is correct, it would follow that at least a part of our moral language is addressed to those who are moral agents in the weak sense but not in the Kantian sense. This can also be seen in the case of young children, with whom we sometimes use moral language before they have become moral agents in the full Kantian sense. Indeed it is partly by participating in such use of language that children gradually acquire moral sentiments.

The Kantian sense of moral agency is presupposed only by parts of our moral system, especially that part which is concerned with apportioning moral desert from the point of view of cosmic fairness. If certain wrongdoers are not Kantian moral agents, the view that they deserve to suffer for their conduct is undermined. We do say to a child of two that it ought not to kill. We may even punish the child for its wrongs if that does some good; but we do not think that if it did not suffer for its wrongs in this life, it would be fair if there were another world, a hell, where it would be made to suffer, in the way that it would be fair if Stalin (assuming that he was a Kantian moral agent) were made to suffer in hell. As for the IMF, even God would not be able to send it to hell or make sense of its deserving to suffer (in the irreducible phenomenological sense) as a collective.

See also: Action; Desert and merit; Kantian ethics; Moral justification; Moral motivation; Morality and identity; Praise and blame

VINIT HAKSAR

References and further reading

Although they involve intricate argument, none of these items is particularly technical. All except the Kant are fairly easy to follow.

Bentham, J. (1817) *Plan for Parliamentary Reform*, in J. Bowring (ed.) *The Works of Jeremy Bentham*, vol. 3, Edinburgh, 1843.(Proposes that the setting up of the right institutions, such as representative democracy, and sanctions, such as punishment, would lead all members of the public, including the rulers, to see that contributing to the common good would be in their own best interests, and therefore that rulers who acted against the public interest would be unlikely to be re-elected.)

Bradley, F.H. (1894) 'Some Remarks on Punishment', *International Journal of Ethics*; repr. in *Collected Essays*, Oxford: Oxford University Press, 1935, vol. 1, essay 7.(Advocates Darwinism and social surgery against nonmoral agents.)

Bradley, F.H. (1927) *Ethical Studies*, Oxford: Oxford University Press, 1927, ch. 1.(Outlines some of the conditions of moral responsibility.)

Campbell, C.A. (1957) *Selfhood and Godhead*, London: Allen & Unwin, 376-85. (A Kantian defence of free will and moral agency.)

Edwards. J. (1758) 'Original Sin', in C.A. Holbrook (ed.) *Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1970, part IV, ch. 3.(Argues in favour of collective moral accountability.)

Haksar, V. (1964) 'Aristotle and the Punishment of Psychopaths', *Philosophy* 39: 323-40; repr. in J. Walsh, and H. Shapiro (eds) *Aristotle's Ethics*, Belmont, CA: Wadsworth, 1967, 80-101.(Discusses whether psychopaths

are morally accountable for their actions and examines Aristotle's views on moral agency.)

Haksar, V. (1991) *Indivisible Selves and Moral Practice*, Edinburgh: Edinburgh University Press, chaps 2 (section 2), 9 and 10.(Examines the analogy between individuals and collectives.)

James, W. (1907) *Talks to Teachers on Psychology and to Students on Some of Life's Ideals*, London: Longman's Green and Co., 265-301.(Discussion of what makes life significant.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Advocates 'strong' interpretation of conditions required for moral agency.)

Malcolm, N. (1966) *L. Wittgenstein, A Memoir*, Oxford: Oxford University Press.(Includes Wittgenstein's views on G.E. Moore's 'innocence'.)

Murphy, J. (1972) 'Moral Death: A Kantian Essay on Psychopathy', *Ethics* 91: 284-98.(Discussion of psychopathy from a Kantian standpoint.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, ch. 8.(Expounds the view that a moral personality must have the capacity to act from a sense of justice.)

Taylor, C. (1976) 'Responsibility for Self', in A. Rorty (ed.) *The Identities of Persons*, Berkeley, CA: University of California Press, 281-99; repr. in G. Watson (ed.) *Free Will*, Oxford: Oxford University Press, 1982, 111-26. (Suggests moral agency consists in the capacity to choose one's values in accordance with one's authentic nature.)