25

R. M. Hare (1919–)

WALTER SINNOTT-ARMSTRONG

Richard Mervyn Hare has written on a wide variety of topics, from Plato to the philosophy of language, religion, and education, as well as on applied ethics, but he is best known for his general moral theory. Hare's views on ethics have developed since his groundbreaking book, *The Language of Morals* (1952), but the main thrust of his position has remained fairly constant.

Definition of moral judgments

Hare defines the class of moral judgments to include any judgment that is prescriptive, universalizable, and overriding (1981: 53–7). Prescriptivity distinguishes moral judgments from judgments of natural science and history. Universalizability separates moral judgments from particular commands, such as by army sergeants, as well as from legal judgments (1963: 36). Overridingness divides moral judgments from aesthetic value judgments (1963: 139).

Hare's definition of moral judgments is formal in that it does not require any particular content. It is possible, on Hare's definition, to make a moral judgment that one should never step on cracks in the sidewalk, even if such steps harm nobody, break no promise or law, and so on. Some critics find this implication unpalatable, but Hare responds that his definition still captures one possible and useful specification of the term "moral." Many people seek to formulate and justify a system of judgments that are moral in Hare's sense.

Prescriptivism

Hare's moral theory, then, starts with prescriptivism, which is the view that moral and other value judgments are typically prescriptive. To call a judgment prescriptive is to say that it is used to prescribe action, that is, to perform some speech act in a large group that includes commanding, advising, encouraging, discouraging, and so on, with respect to some particular action or kind of action. The paradigm form of prescription is the imperative, so prescriptivism claims that a value judgment such as "Hondas are better than Toyotas" are used to perform a speech act similar to "Choose a Honda over a Toyota." Despite the analogies between value judgments and imperatives, Hare has

insisted from the start that "it is no part of [his] purpose to 'reduce' moral language to imperatives" (1952: 2).

Prescriptivism is best understood in contrast with its predecessors in metaethics. G. E. Moore criticized naturalism, which is the view that moral and other value judgments ascribe or deny evaluative properties, such as goodness or rightness, that are supposed to be reducible in some way to properties that are natural apparently in the sense that they can be studied by the methods of natural science. In place of naturalism, Moore proposed non-naturalism, which is the view that moral and other value judgments are about a wholly different kind of property that can be known only by intuition.

Hare accepts Moore's arguments against naturalism and adds new arguments of his own (1952: 79–93). However, Hare also rejects Moore's non-naturalism as unnecessarily mysterious. Moore's mistake, according to Hare, was to retain a basic assumption of naturalism, namely, descriptivism, which is the view that value judgments are used to describe values or evaluative properties. Hare argues that both naturalism and non-naturalism should be rejected, because value judgments are not used to describe in either way.

Descriptivism was rejected before Hare by emotivists, including A. J. Ayer and Charles Stevenson (see AYER and STEVENSON). Emotivists claimed that value judgments are used to express and arouse emotions. Against this view, Hare points out that one can make value judgments dispassionately, without having any feeling or emotion to express. Moreover, just as advice can be given but not followed, so a value judgment can succeed in its speech act even if it does not arouse any emotion or change any behavior in the audience. These points enable Hare to distinguish value judgments on his view from propaganda and to subject them to rational scrutiny of a kind that is often assumed to be inappropriate for emotions (1952: 9–16).

Hare's view, then, is that value judgments are typically used for a different speech act, namely, prescribing. He does admit that some value judgments are not used to prescribe. When someone says a building is good Gothic revival just to indicate that it is the kind of building that would be judged good by people who like Gothic revival buildings, Hare calls this an "inverted commas use" (1952: 124). Value judgments can also be used ironically and merely to pay lip service to conventions. Past-tense value judgments are explained in similar ways. When I say that it was morally wrong for Jefferson to hold slaves, I do not prescribe to Jefferson that he do anything, since it is too late for that; but, if I had been able to do so, then I would have prescribed that he not hold slaves. In such cases, value judgments are not used to prescribe, but such uses are still parasitic on, because explained in terms of, prescriptive uses by other people or in other circumstances. Hare's claim that value judgments are "typically" prescriptive seems to mean that all value judgments either are used to prescribe or can be explained in terms of other uses that are prescriptive (1963: 22n).

Such prescriptivism provides a natural explanation of many features of evaluative language. One common dictum claims that "ought" implies "can." If a judgment that an agent ought to do an act is used to prescribe that the agent do the act, and there is something wrong with prescribing that an agent do an act when the agent cannot do it, then this explains why there is something wrong with saying that an agent ought to do an act that she cannot do (1963: 51–66).

A more often questioned implication of Hare's prescriptivism is that one cannot fully think that one ought to do an act and yet not do it, if one can do it and now is the time to do it. This seems to rule out weakness of will, but Hare explains apparent cases of weakness of will as cases where agents do not really think at the time that they ought to do relevant particular acts, use parasitic senses of "ought," psychologically cannot bring themselves to do what they think they ought to do, and so on (1963: 67–85). If such reinterpretations can adequately explain away all apparent cases of weakness of will, then Hare's prescriptivism is harder to refute than his critics assume.

Universalizability

Despite their prescriptivity, value judgments differ from singular imperatives in a crucial respect, according to Hare. When a drill sergeant commands, "To the left, march!," he needs no reason for choosing left over right, and nothing goes wrong if, when the same circumstances arise again later, he then commands, "To the right, march!" In contrast, Hare argues that there is a logical inconsistency between saying that one agent ought to do an act and denying that another agent ought to do an act with relevantly similar properties in relevantly similar circumstances (1952: 81). Hare's explanation of this linguistic rule is that "all value judgments are covertly universal in character, which is the same as to say that they refer to, and express acceptance of, a standard which has application to other similar instances" (1952: 129), that is, they presuppose a general principle. This implicit universality makes it legitimate to ask for a reason why the agent ought to do the act, or a property that makes the thing good. It also enables value judgments to be useful for public teaching of standards (1952: 134). In this respect, value judgments are like descriptive judgments, according to Hare (1963: 10–14).

This doctrine of universalizability might seem empty without limits on which properties are relevant. Hare does claim that references to individuals, as in proper names or indexicals, must be irrelevant in order for the presupposed standard or principle to be universal. However, he insists that no additional limits are part of the shared meaning of evaluative terms. In his view, we cannot determine which properties are morally relevant in advance of determining which moral principles are defensible (1981: 62–4). This makes Hare's doctrine of universalizability much weaker than that of his predecessor Kant on most interpretations.

Nonetheless, Hare's thesis of universalizability is criticized by particularists. Some argue that even an agent's individual identity or particular spatiotemporal location might be morally relevant; or at least that this is not excluded by language alone, so to assume otherwise is to adopt a substantive position. Other particularists argue that the reasons why one act is morally wrong might not make another act morally wrong, because the force of each morally relevant feature varies with the circumstances in ways that cannot be specified in any general principles. Hare would respond by arguing that some feature of the person or time or circumstances must be available to explain why one act is wrong when another is not. We might and need not be able in practice to specify fully the relevant properties or underlying principles, but some specification must be possible in theory for our value judgments to be logically consistent, in Hare's view (1963: 18–20).

Rationality

The next prong of Hare's theory is a particular view of rationality. The linguistic theses of prescriptivism and universalizability lay down some limits on consistency and, hence, rationality. Hare also assumes that "any rational thinking about [moral questions] has to be done in the light of facts" (1981: 87). At times, Hare also seems to assume that logical consistency and knowledge of facts is all there is to rationality.

This account of rationality might not seem very controversial, but Hare adds that one fact that needs to be considered for moral judgments to be rational is "what it is like" for people who are affected by our actions (1981: 92). In particular, if they suffer, it is not enough for me to know that they suffer. I need to know what their suffering is like to them.

This kind of knowledge, Hare argues, requires me to have a certain motivation or preference. The following two claims are distinct for any situation:

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. (1981: 95)

(1) is about my current preferences regarding a counterfactual situation, whereas (2) is about a counterfactual situation in which I would have preferences that I do not currently have. Although Hare admits that these claims are distinct, he argues that "I cannot know that (2), and what that would be like, without (1) being true, and . . . this is a conceptual truth, in the sense of 'know' that moral thinking demands" (1981: 96). This transfer principle, as I will call it, implies that one must have certain preferences in order to be rational.

The master argument

By combining prescriptivism, universalizability, and rationality, Hare claims to be able to derive a kind of utilitarianism (1981: 109–11). The argument starts with an example, whose facts I have modified somewhat to clarify the logic of the argument:

1 B cannot park B's car in this parking place unless I move my bicycle.
2 B has a strong preference to park B's car in this parking place.
3 I have only a weaker preference not to move my bicycle (or to leave it there).

Utilitarianism suggests that I morally ought to move my bike, so Hare's goal is to derive an absurdity from the supposition that an opposite belief is rational:

4 I rationally believe that I morally ought to leave my bicycle there.

This supposition on Hare's view of rationality implies

5 I know all of the relevant facts about leaving my bicycle there, including what it would be like for B if I left my bicycle there.

Assuming that roles include preferences, so that, when we switch roles, we also switch preferences, (2) and (5) imply

6   I know that, if B and I switched roles, I would strongly prefer that B not leave the bicycle there.

Hare's transfer principle plus (6) yields

7   I now strongly prefer that, if B and I switched roles, B would not leave the bicycle there.

Assuming that this is the strongest preference between these alternatives in this situation, then, since preferences are prescriptive,

8   I accept the prescription: "if B and I switched roles, let B not leave the bicycle there."

But the universalizability of moral judgments plus (4) imply

9   I believe that, if B and I switched roles, B morally ought to leave the bicycle there.

This plus the prescriptivity of moral judgments implies

10   I accept the prescription: "if B and I switched roles, let B leave the bicycle there."

The prescriptions in (8) and (10) are inconsistent, and inconsistency excludes rationality, so (4) is refuted. This means that

11   If I am rational, I cannot believe that I morally ought to leave the bicycle there.

Once this belief is ruled out as irrational, the options are limited:

12   If I believe that I morally either ought or ought not to leave the bicycle there, and if I am rational, then I must believe that I morally ought not to leave the bicycle there.

Assuming that this example is not different in any relevant respect from any other conflict involving only two people, we can generalize:

13   In every two-person conflict, if I believe that I morally either ought or ought not to an act, and if I am rational, then I must believe that I morally ought not to do whatever frustrates stronger preferences than it satisfies.

What is believed in the consequent of (13) is just what preference utilitarianism says about such two-person conflicts.

Hare next extends his argument to choices that affect any number of people. Then, to make a rational choice, an agent needs to know what it is like for each person who is affected. That knowledge, according to Hare's transfer principle, requires the agent to have current preferences for or against the act being done if he were to switch roles with each of them and take on their positive or negative preferences. Indeed, the agent must have preferences with the same strengths as those of the people affected. Thus, an agent morally ought to do what he would prefer to be done if he had to occupy successively each of the roles of each person affected. This conclusion is a general version of preference utilitarianism.

Hare's argument could be criticized from many angles. First, amoralists might refuse to make any positive moral judgments at all. Then they do not deny (12) or (13), but

they do block any derivation of their consequents. Hare admits that there is nothing "logically inconsistent in this position" (1981: 186), but he gives "reasons of a non-moral sort why it should not be chosen" (p. 190).

A different problem is raised by fanatics, who cling to ideals even when they know that their moral judgments run contrary to utilitarianism. Hare responds that the fanatic's moral judgments are in effect a new moralistic kind of preference, so fanatics face a dilemma (1981: 181). On one horn of this dilemma, the fanatic claims that his moralistic preferences are strong enough to outweigh all of the preferences of everyone else affected by the action. This claim would be implausible, and, even if it were true, a preference utilitarian could just grant that this moralistic preference should be fulfilled, since it is stronger than the rest put together. Thus, if the fanatic's position is to remain contrary to utilitarianism, the fanatic must grab another horn of Hare's dilemma. The only alternative for the fanatic is to admit that his moralistic preference is not strong enough to outweigh all of the preferences of everyone else affected by the action. But then, if the fanatic is rational, he must know what it is like for those people's preferences to be frustrated. By the transfer principle, he must then come to have corresponding preferences with the same strengths. These preferences will then make him abandon his fanaticism. What he cannot do is remain a fanatic and also remain rational, according to Hare.

The transfer principle itself faces problems, however. Consider someone who is trying to kill me so that they can take my place on the basketball team. Should I stop them? For my answer to be rational, I need to know that, if we switched places (so I wanted to kill them in order to get on the team), then I would strongly prefer that they not stop me from killing them. According to the transfer principle, I cannot know what that would be like unless I *now* strongly prefer that if I tried to kill them to get on the team, then they would not stop me. However, this current preference does not seem necessary for me to know all that is going on. I can know why they want to kill me, and what it is like for them, even if my current moral beliefs keep me from having any current preference that they not stop me if I try to kill them. More generally, preferences that we see as immoral seem to be one kind of preference that rationality does not require us to take over from others in order to know what it is like for them to have that preference frustrated.

The same problem arises when the transfer principle is used against non-utilitarian moral theories. Suppose a retributivist believes that the government morally ought to execute a convicted murderer, even if this execution would not maximize preference satisfaction. This position is contrary to utilitarianism, so Hare would label this retributivist a fanatic. Whatever he is called, such a retributivist can know that, if he were in the convicted murderer's situation, then he would prefer strongly that the execution not happen. He can even know what it would be like to be executed, as much as anybody else can. But the retributivist still need not *currently* have any preference that, if he were in the same situation as the convicted murderer, he should not be executed. In that counterfactual situation, he himself would deserve to be executed, according to his current retributivist views; so, if he sticks with his views, he would prefer that, if he himself had committed such a heinous murder, then he himself should be punished like any other murderer. Such retributivism seems conceptually coherent, consistent with both prescriptivism and universalizability, and also common. Thus, Hare's

transfer principle seems to fail in the very kind of case where he needs it. If so, his argument cannot prove his version of preference utilitarianism.

Utilitarianism

Even if Hare's argument does not prove his version of utilitarianism, that conclusion still might be correct and defensible. Hare is basically an act utilitarian with a preference-based theory of value. Such views have been subjected to a great deal of criticism and even scorn, but Hare's version has distinctive features that make it harder to refute than many others.

The most important innovation is Hare's distinction between two levels of moral thinking. This distinction is motivated by a discussion of apparent moral conflicts. In his main example (1981: 27), a lifelong friend visits unexpectedly from Australia and asks to be shown around Oxford on the very day when Hare promised to take his children for a picnic on the river. It seems initially that Hare morally ought to keep his promise, and he also morally ought to show his friend around Oxford. The problem is that Hare cannot do both, so the fact that he morally ought to keep his promise implies that he morally ought not to show his friend around Oxford. Since moral judgments are prescriptive and overriding, so "ought" implies "can," Hare denies the possibility that he morally both ought and ought not to show his friend around Oxford. So he also denies that such moral conflicts are ever possible in the end.

Nonetheless, Hare wants to explain why moral conflicts seem to be not only possible but common. To explain this appearance, Hare distinguishes two uses of "ought" that occur at two levels of moral thinking. On the intuitive level, Hare's thinking that he morally ought to show his friend around Oxford is "inseparable from" his feeling compunction before and guilt after his failure to show his friend around Oxford (1981: 30–1). Since a good person would also feel compunction before and guilt after failing to keep his promise to his children, Hare admits that conflicting moral beliefs at the intuitive level can both be justified by their practical usefulness. Intuitive moral beliefs facilitate moral teaching and reduce errors in moral judgment when information is scarce or time is short. That explains why we do and should sometimes think at the intuitive level, where moral conflicts seem to occur.

But this intuitive level of moral thinking is not enough by itself. We need to determine which intuitive moral principles should be taught and used in making everyday decisions. We also need some way to determine the morally right course of action when intuitive moral judgments conflict. Both of these problems, according to Hare, can be solved only at a different level of moral thinking, which he calls the critical level. At that critical level, to say that Hare morally ought to show his friend around Oxford is to make a prescriptive and overriding judgment, so such critical moral judgments cannot conflict. Moreover, critical moral thinking must conform to preference utilitarianism, according to Hare's master argument. Since incompatible alternatives cannot each frustrate stronger preferences than it satisfies, Hare's utilitarianism implies that an agent never morally ought to adopt each of two conflicting options. Such critical thinking in conformity with utilitarianism is what ultimately determines which act is morally right, according to Hare. That is why Hare can remain a utilitarian and yet still explain why moral conflicts do and should seem to occur.

This dichotomy between levels of moral thinking also enables Hare to respond to standard counterexamples that are supposed to refute utilitarianism by showing that it conflicts with common moral intuition: "Perhaps the [proverbial] sheriff should hang the innocent man in order to prevent the riot in which there will be many deaths, if he knows that the man's innocence will never be discovered and that the bad indirect effects will not outweigh the good direct effects; but in practice he will never know this" (1981: 164). Hare's point is that, in order to create any trouble for utilitarianism, the details of such examples must be so unrealistic that we have little or no reason to trust our moral intuitions about such circumstances. Our moral intuitions are justified because we will act better in most situations if we inculcate such moral intuitions deeply into our characters, but the same moral intuitions cease to be reliable guides in unrealistic circumstances for which they were never intended. Consequently, utilitarianism cannot be refuted by any appeal to moral intuition in this or any other unrealistic example, according to Hare (1981: 134).

This response will not convince every opponent, and many other objections could be raised to Hare's utilitarianism and to other aspects of his moral theory. Nonetheless, the range of Hare's views, his attention to detail, the clarity of his writing, and his ability to bring abstract moral theory to bear on concrete issues of great importance provide a model for all moral philosophers to follow.

Bibliography

Works by Hare

1952: *The Language of Morals*, Oxford: Clarendon Press.
1963: *Freedom and Reason*, New York: Oxford University Press.
1971a: *Essays on Philosophical Method*, London: Macmillan.
1971b: *Practical Inferences*, London: Macmillan.
1972a: *Applications of Moral Philosophy*, London: Macmillan.
1972b: *Essays on the Moral Concepts*, London: Macmillan.
1981: *Moral Thinking: Its Levels, Method, and Point*, New York and Oxford: Oxford University Press.
1982: *Plato*, New York and Oxford: Oxford University Press.
1989a: *Essays in Ethical Theory*, New York and Oxford: Oxford University Press.
1989b: *Essays on Political Morality*, New York and Oxford: Oxford University Press.
1992: *Essays on Religion and Education*, New York and Oxford: Oxford University Press.
1993: *Essays on Bioethics*, New York and Oxford: Oxford University Press.
1998: *Sorting Out Ethics*, New York and Oxford: Oxford University Press.
1999: *Objective Prescriptions and Other Essays*, New York and Oxford: Oxford University Press.

Work by other authors

Seanor, D. and Fotion, N. (eds.) (1990) *Hare and Critics: Essays on Moral Thinking*, Oxford: Clarendon Press.