

Michigan Law Review

Volume 82 | Issue 4

1984

Intuition and Security in Moral Philosophy

Stephen R. Munzer

University of California, Los Angeles

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Law and Philosophy Commons](#)

Recommended Citation

Stephen R. Munzer, *Intuition and Security in Moral Philosophy*, 82 MICH. L. REV. 740 (1984).

Available at: <https://repository.law.umich.edu/mlr/vol82/iss4/11>

This Review is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

INTUITION AND SECURITY IN MORAL PHILOSOPHY†

*Stephen R. Munzer**

MORAL THINKING: ITS LEVELS, METHOD, AND POINT. By *R.M. Hare*. Oxford: Clarendon Press. 1981. Pp. viii, 242. Cloth, \$22.50; paper, \$8.95.

I. UNIVERSAL PRESCRIPTIVISM AND UTILITARIANISM

In *Moral Thinking* Professor R.M. Hare refines and extends a metaethical theory that he began to develop over thirty years ago.¹ The theory is universal prescriptivism. It holds that moral judgments are universalizable, prescribe rather than either describe facts or merely express feelings, and override other kinds of judgments, such as aesthetic or prudential judgments, in cases of conflict. In his new book Hare discusses how to think about moral problems and tries to show why sound thinking about them leads to utilitarianism. The book is self-contained and therefore accessible to the reader unacquainted with Hare's earlier works. Although Hare's views are now somewhat out of fashion, the new book demonstrates that he still has much to say and says it as compactly, lucidly, and wittily as anyone writing moral philosophy today. It would be a mistake for his opponents to ignore it.

To think well about moral problems, Hare maintains, requires that we distinguish two *levels* of moral thinking — intuitive and critical. At the intuitive level we accept and employ *prima facie* moral principles for deciding what we ought to do. Typically, these principles stem from our upbringing. If, as sometimes happens, these principles conflict, we must move to the critical level. Critical thinking resolves conflicts and in general tries to select the best set of intuitive principles. Both levels are necessary. Intuitive principles are usually simple and general — for example, "Do not kill," "Tell the truth," "Keep your promises." They must be so in order to be easy to learn and difficult to manipulate for our own purposes or situation. But because they are usually simple, they can conflict, and we need critical thinking to resolve the conflicts. At the critical level we may rely

† I wish to thank David H. Dolinko, Dale Jamieson, James W. Nickel, Donald H. Regan, and Frederick Schauer for helpful comments and suggestions.

* Professor of Law, University of California, Los Angeles. B.A. 1966, University of Kansas; B.Phil. 1969, Oxford; J.D. 1972, Yale. — Ed.

1. Important earlier works include R.M. HARE, *THE LANGUAGE OF MORALS* (1952); R.M. HARE, *FREEDOM AND REASON* (1963).

on linguistic intuitions regarding the logical properties of moral words but not on intuitions of moral substance.

Hare then proposes a *method* for rational thinking about moral questions at the critical level. The method involves a close look at the facts. Because of the consequences of acting on prescriptions in concrete situations, all prescriptions, in order to be rational, require cognizance of the facts. The principle of universalizability requires that identical moral judgments be made in factual situations identical in their universal properties. It is the logical requirement to universalize our prescriptions that generates utilitarianism:

[T]he method of critical thinking which is imposed on us by the logical properties of the moral concepts requires us to pay attention to the satisfaction of the preferences of people (because moral judgements are prescriptive, and to have a preference is to accept a prescription); and to pay attention equally to the equal preferences of all those affected (because moral principles have to be universal and therefore cannot pick out individuals). [P. 91].

So long as we put to one side philosophical skepticism about the very existence of other minds, Hare believes that interpersonal comparisons will not trouble the utilitarian. Such comparisons occasion no difficulty for intuitive thinking; at the intuitive level we believe that we can rationally compare degrees or strengths of preferences of different people. At the critical level, the key move is "to reduce comparisons between other people's preferences to comparisons between our own" (p. 128). We do this by asking ourselves what *we* would prefer were we forthwith put into someone else's position with *his* preferences.

This approach to moral questions yields, Hare believes, a conclusive response to those who attempt to show utilitarianism unacceptable by bringing it into conflict with some received moral opinion or intuition regarding, for example, loyalty to one's family or matters of justice and rights. At the critical level such appeals to moral intuition are out of court. At the intuitive level they are, of course, allowed, but only pending their certification by critical thinking. Hare perceptively discusses some of the stock examples in the literature.² He contends that it is, at the least, extremely difficult for the anti-utilitarian really to prove that the received moral intuition is both correct and incompatible with utilitarianism.

Lastly Hare comes to the *point* of his method: Why should we want to use it rather than some other way of thinking morally, or indeed rather than some other kind of practical thinking (say, pru-

2. Pp. 48-49, 132-40, 154-56. See, e.g., Smart, *An Outline of a System of Utilitarian Ethics*, in J. SMART & B. WILLIAMS, *UTILITARIANISM: FOR AND AGAINST* 3, 69-73 (1973) (sheriff "framing" and executing an innocent man in the general interest); Williams, *A Critique of Utilitarianism*, in *id.* at 77, 98-118 (one person killing an innocent man to avoid someone else's killing twenty persons).

dential thinking)? To take the second limb of the question first, Hare believes that the issue of "Why should we be moral?" resolves, in part at least, into the issue of whether there are prudential reasons — that is, reasons of perceived self-interest — for being moral. He responds that, while it is not *always* in our self-interest to be moral, prudential reasons support inculcating a wide range of prima facie moral principles. Resistance to moral thinking might also come from the "fanatic" and the "amoralist." These are terms of art in Hare's lexicon, and there are sub-types of each. Hare argues that the sort of pure fanatic who is able and willing to think critically but who holds opinions inconsistent with utilitarianism is not even a theoretical possibility. On the other hand, Hare acknowledges that to refrain from making moral judgments at all (universal amoralism) is a position that his theory cannot rule out. As to why we should use Hare's method rather than some other method of moral thinking, his answer is simply that it is the only way of reasoning securely about moral questions. For his method puts into play only ordinary facts and the logical properties of moral words. It requires no special "moral" facts known only by intuition, and thus avoids the irreconcilable conflicts that mark all forms of intuitionism.

This essay concentrates on the leading theme of Hare's book: that his moral philosophy is more secure than, and therefore superior to, all forms of reliance on moral intuitions. It pursues this theme through discussions of intuitive and critical thinking, the limits of universalization, and preferences and their alteration.³ I shall suggest that linguistic intuitions and methodology are less secure, and some moral intuitions are more secure, than Hare believes. Strong warrant exists, moreover, for holding that in spite of Hare's arguments utilitarianism rests on a moral intuition of substance. In the current state of moral philosophy it is hard to avoid the conclusion that there is a plurality of moral principles or considerations that sometimes conflict in ways that no priority rules can satisfactorily resolve.

3. As a result it is impossible to examine other aspects of Hare's book. Among them are his stimulating discussion of personal identity, pp. 96-99, 119-21, 221-23, his device of archangelic moral thinking, pp. 44-64 and *passim*, his acute analysis of prudence and supererogation, pp. 188-205, and the details of his case against anti-utilitarians, pp. 130-68. I also leave aside the important issue of whether Hare has made his system of moral reasoning so inspissated that he falls into naturalism — roughly, the doctrine that the meanings of moral words derive from the nonmoral properties of acts. Hare has been one of the most vigorous and persuasive critics of naturalism and is anxious to show that his new views do not commit him to it. *See* pp. 68-71, 186-87, 218-28. My view, briefly, is that Hare comes close to but does not embrace naturalism because of his insistence that people are free, first, to refrain from making moral judgments altogether, *see* pp. 182-87, 219-20, 228, and second, to prefer what they prefer subject to universalizability, *see* pp. 225-28. I would add, however, that the more Hare stresses the liberty-to-prefer escape route, the more trouble his theory is likely to have in rationally evaluating and altering preferences (*see* Part IV *infra*).

II. INTUITIVE AND CRITICAL THINKING

It is well to consider what motivates Hare's distinction between the two levels of moral thinking.⁴ In part the motivation is the same as that for rule-utilitarianism. To avoid problems of bias, prejudice, and insufficient time from always having to appeal directly to the principle of utility, the rule-utilitarian distinguishes between that principle and the set of secondary moral principles intended for everyday use which, if accepted, would best promote utility.⁵ Garden-variety rule-utilitarianism uses the secondary principles in day-to-day living and employs the principle of utility only to assess these secondary principles. Similarly, Hare's intuitive *prima facie* moral principles are for daily use, with critical thinking employed, when possible, to resolve conflicts and generally to select the best intuitive principles.⁶

An important additional motivation for Hare's distinction between the intuitive and critical levels is its use in the battle against "intuitionism." This term has received different definitions. For Hare it means the doctrine that one may properly appeal to one's opinions, feelings, convictions, or considered judgments in matters of moral substance.⁷ In his eyes John Rawls is an intuitionist,⁸ and though Rawls is the *bête noire* of the book, he is decorously kept mainly off-stage.⁹ The problem, in Hare's view, is that the intuitionist "has in fact nothing to appeal to but the feelings" (p. 77). Hare castigates arguments resting on "undefended moral intuitions of sub-

4. He also recognizes a third "metaethical" level, which concerns "the meanings of the moral words and the logic of moral reasoning." P. 26. Thus Hare's book is a metaethical treatment of what he considers to be the two levels of substantive ("substantial" in Hare's British English, see pp. 4-5, 16, and *passim*) moral thinking — intuitive and critical. Hare also identifies some sub-levels, not pertinent here, within the intuitive level. See p. 203.

5. See, e.g., Brandt, *Some Merits of One Form of Rule Utilitarianism*, 3 U. OF COLO. STUD. SERIES IN PHIL. 39 (1967).

6. Hare offers, in fact, an acute combination of rule-utilitarianism and act-utilitarianism. "[T]he critical thinker considers cases in an act-utilitarian or specific rule-utilitarian way, and on the basis of these he selects . . . general *prima facie* principles for use, in a general rule-utilitarian way, at the intuitive level." P. 43.

7. See, e.g., pp. 40, 75, 172, 217. I shall use the term "intuitions" to cover all of the terms in this family.

8. See p. 75. Rawls defines intuitionism differently as any of a number of ethical theories having two features: "first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another: we are simply to strike a balance by intuition, by what seems to us most nearly right." J. RAWLS, *A THEORY OF JUSTICE* 34 (1971); accord Urmson, *A Defence of Intuitionism*, 75 PROC. OF THE ARISTOTELIAN SOC. 111 (1975). Rawls is not an intuitionist in his own sense of the term because, though he has a plurality of principles of justice, he gives priority rules for resolving conflicts between them. See J. RAWLS, *supra*, at 302-03. Hare complains that Rawls and Urmson divert "intuitionism" to a new usage. P. 75. For a sensitive discussion see Smith, *Rawls and Intuitionism*, in NEW ESSAYS IN CONTRACT THEORY, CAN. J. PHIL. (Supp. vol. III) 163 (1977).

9. Cf. p. 12 (passage criticizes a position akin to that of Rawls without mentioning him by name, see J. RAWLS, *supra* note 8, at 20-21, 48-51).

stance," which involve nothing but the "mere moral opinions of the authors" (p. 76). For moral thinking to be secure, he considers it essential that there be a critical level which can appraise and certify intuitive *prima facie* moral principles. Critical moral thinking relies only on "linguistic intuitions" (p. 9) and the "ordinary procedures" (p. 7) of philosophical logic. It is not wholly clear what Hare means by "philosophical logic," but it has much more to do with a careful linguistic and philosophical methodology than with formal or mathematical logic.

Although some distinction between intuitive and critical moral principles is both possible and necessary, Hare's use of this distinction against intuitionism is not free from difficulty. Assume for the moment that we can accurately separate linguistic from moral intuitions and thus critical from intuitive thinking. Even so, we have powerful artillery against intuitionism, as Hare speaks of it, only if "philosophical logic" is highly secure and any intuitive appeal to moral substance is quite fragile. Perhaps neither is the case.

Today no unanimity exists on which logical and linguistic methods are best. It did not exist, I dare say, even in the 1950's and 1960's when Hare developed the main lines of his moral philosophy. Hare's methodology is sophisticated;¹⁰ but it belongs to the mainstream of post-war Oxford linguistic analysis¹¹ and little reflects the work of Chomsky, Davidson, Kripke, Putnam, and others.¹² I am not saying that Hare's methods are suspect because they are unfashionable. Nor am I saying that they are wholly incompatible with the methods of other philosophers. But there are important methodological differences — for example, on the analytic-synthetic distinction¹³ — which Hare does not explore and on which the views of other philosophers have some plausibility. Thus, while Hare could reply that later methods are correct where and only where they do not conflict with his own, this response helps little, given the current state of the philosophy of logic and language, in convincing the reader of the security of critical thinking that relies solely on linguistics.

10. See pp. 7-10, 80-82.

11. Hare makes some minor adjustments, at least in his way of putting things, in deference to papers by A.N. Prior, see pp. 9-10, and W.V.O. Quine, see pp. 77-78. See Prior, *The Runabout Inference-Ticket*, 21 ANALYSIS 38 (1960); W. QUINE, *Two Dogmas of Empiricism*, in FROM A LOGICAL POINT OF VIEW 20 (2d ed. 1961).

12. See, e.g., N. CHOMSKY, ASPECTS OF THE THEORY OF SYNTAX (1965); S. KRIPKE, NAMING AND NECESSITY (1980); H. PUTNAM, MIND, LANGUAGE AND REALITY — PHILOSOPHICAL PAPERS, VOLUME 2 (1975); Davidson, *Truth and Meaning*, 17 SYNTHESIS 304 (1967).

13. On pages 77-78 Hare finesses the criticisms of the distinction offered by Quine. See note 11 *supra*. It may be that Quine's position is overstated and that we can at least sometimes distinguish between analytic and synthetic statements. See Grice & Strawson, *In Defense of a Dogma*, 65 PHIL. REV. 141 (1965), cited by Hare on page 81. Even so, the distinction may not escape the more cautious criticism that it is of little philosophical importance. See Putnam, *The Analytic and the Synthetic*, in H. PUTNAM, *supra* note 12, at 33. On the relevance of all this to Hare's moral philosophy, see text accompanying notes 23-24 *infra*.

tic intuitions and what Hare regards as "ordinary" methods of "philosophical logic."

On the other side, perhaps what Hare calls intuitionism need not be so indefensible. It will help here to distinguish between crude intuitionism, which relies straightway on "intuitions" in the sense of mere opinions or feelings, and refined intuitionism, which relies on intuitions only after they have been subjected to some procedure for eliminating intuitions that are apt to rest on bias, prejudice, poor empirical information, and the like. Hare seems to understand "intuitionism" to cover both the crude¹⁴ and refined¹⁵ varieties. Few moral philosophers would defend crude intuitionism. Neither shall I. But perhaps there are forms of refined intuitionism that offer a more cautious and plausible way of invoking moral intuitions.

Rawls provides an example. His principles of justice are those selected in the original position in a state of "reflective equilibrium." This requires the parties to match a favored description of the original position with their most considered judgments or convictions ("intuitions"). Rawls contemplates that they will go back and forth, with adjustments made to both the description and the considered judgments until they are consistent. The equilibrium thus achieved need not be stable; either or both may demand revision upon being examined more closely.¹⁶ The parties manifestly do *not* take considered judgments as they find them. They may need to smooth out irregularities and distortions after they have placed their judgments against abstract principles. As an ideal, reflective equilibrium involves being "presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them."¹⁷ True enough, intuitions about moral substance are involved; Rawls rejects the idea that definitions or analyses of meaning either are sufficient or have a special status.¹⁸ Nevertheless, the method of reflective equilibrium is considerably more sophisticated than a simple appeal to substantive moral intuitions and offers a way of filtering out some such intuitions.

The issue at stake does not reduce to a wrangle between Hare and Rawls; there are attempts, different from Rawls', to use intuitions in a cautious and plausible way. A notable example is the ideal speaker theory of M.B.E. Smith.¹⁹ His program does not rely

14. See, e.g., pp. 76, 77, 217.

15. See p. 12. Some passages, e.g., pp. 40, 75, straddle the two varieties.

16. See J. RAWLS, *supra* note 8, at 20-21. Some similarity exists between these oscillations and Hare's procedure for determining which facts are relevant in making a moral judgment. See pp. 89-90.

17. J. RAWLS, *supra* note 8, at 49. See generally *id.* at 48-51, 120, 432.

18. See *id.* at 51.

19. Smith, *Ethical Intuitionism and Naturalism: A Reconciliation*, 9 CAN. J. PHIL. 609 (1979).

on all moral intuitions or judgments. Smith stresses that moral judgments are a function of both our moral principles and our factual beliefs, and hence that we cannot expect moral unanimity until we free ourselves of false empirical beliefs, personal interests, class or cultural biases, and the like. Reliable moral intuitions are just those that we would hold under ideal conditions — an ideal reflective morality. To implement his intuitionist program, Smith hypothesizes an ideal speaker who tries to determine when the sentences that express moral judgments are true. To do that the ideal speaker uses some semantic theory which has linguistic rules for determining reference, as well as an account of linguistic competence, such as that of Chomsky,²⁰ for investigating linguistic rules. The ideal speaker, who is also devoid of personal interests and class and cultural associations, endeavors to attain substantive moral agreement by pruning and filtering out deviant moral judgments. Some difficulties exist with this theory. Among them are objections that Hare himself would raise, such as that it has difficulty in accounting for the motivational or prescriptive character of moral beliefs. Whether Smith has dealt with these difficulties satisfactorily cannot be taken up here.²¹ But surely his theory handles moral intuitions or judgments carefully, offers a way to remove those that are likely to be incorrect, and at least holds open the prospect of moral agreement. It thus represents a hardier variety of intuitionism than Hare addresses.

In any case, Hare's theory requires the assumption, mentioned earlier but not considered, that we can accurately separate linguistic from moral intuitions and hence critical from intuitive thinking (pp. 7-15). But perhaps we cannot. One aspect of this problem — whether certain words like "courageous" or "cruel" or "rude" incorporate moral substance or embrace some other sort of evaluative standard — has been debated in the literature for years,²² and I shall not pursue it here. A related but nevertheless isolable part of the same problem is how people are to separate the linguistic and moral components of their intuitions. After all, Hare hopes that most people will be capable of critical moral thinking some of the time; that hope is clouded if no one but the keenest philosopher would have a secure way of determining whether he is relying solely on his linguistic intuitions and logic. And perhaps even the keenest philosopher will have trouble. Here Hare's relative neglect of criticism of the analytic-synthetic distinction²³ comes back to haunt him. His project

20. N. CHOMSKÝ, *supra* note 12.

21. See Smith, *supra* note 19, at 625-29. Notice that Smith, using linguistic methods, embraces a form of naturalism, while Hare, using different but related linguistic methods, must struggle to keep naturalism at bay. See note 3 *supra*.

22. Compare R.M. HARE, FREEDOM AND REASON 22-25, 187-96 (1963), with Foot, *Moral Arguments*, 67 MIND 502, 507-10 (1958). Hare's new ripostes are on pp. 17-18, 71-75.

23. See note 13 *supra* and accompanying text.

of separating linguistic intuitions from intuitions of moral substance is parallel to the project of separating analytic statements, whose truth supposedly hinges on the meanings of words used to make them, from synthetic statements, whose truth or falsity depends on empirical facts. If some of the objections to the analytic-synthetic distinction are sound, then, by analogy, it seems possible that some commitments on matters of moral substance are built into our language. If so, Hare's project is dubious.

Hare's now claiming that utilitarianism is rooted in the universal and prescriptive character of our moral concepts points up how nettlesome the matter is.²⁴ At even second blush, it would seem that utilitarianism, like any other normative moral theory, involves a commitment on moral substance. Hare himself acknowledges it to have a substantive element (pp. 4-5). Yet now he tells us that, despite his continued essential adherence to Hume's Law ("No 'ought' from an 'is'") (pp. 16, 186-87), utilitarianism arises solely from linguistic intuitions about our moral words as extracted by the methods of philosophical logic.

III. LIMITS OF UNIVERSALIZABILITY

Can utilitarianism be derived in the way Hare claims? There is reason to suspect that the argument must contain a gap. If Hume's Law applies, if only factual and logico-linguistic premises are in play, if universalizability is a logical or conceptual affair, and if, finally, utilitarianism is a substantive moral doctrine, then it is hard to see how utilitarianism can be generated without some additional substantive moral premise or intuition. Hare constructs his argument in chapters 5 and 6 with great care and skill. Still, it seems to contain some open space.

A possible gap lies in the move from preferences as *candidates* to be the relevant features of actions for purposes of universalizability to preferences as *rightful holders* of that office. The location of the hiatus within the general framework of Hare's argument is as follows. Moral principles pertain to actions and situations. Given the requirement of universalizability, we must identify some feature of actions or situations that counts as a universal property and thus can figure in hypothetical situations that are similar in all relevant respects. Hare proposes "probable effects on preference-satisfactions as *candidates* for the status of relevant features of actions."²⁵ But in the very next section (pp. 94-96) he moves to talking of universalization in terms of preferences and in effect treats preferences as the *rightful holders* of the office. To universalize, I must imagine myself in some-

24. See pp. 108-09, 111, 115-16, 176-77.

25. P. 91 (emphasis in original). On page 90 Hare puts the matter only a bit differently: the "likely effects of possible actions" on people's "experiences" are "obvious candidates."

one else's situation, with *his* preferences, and ask whether I would not have the same aversion that he does to what I propose to do.

If a gap exists here, it may stem from the following shortcoming of Hare's argument in these sections.²⁶ Hare neglects to defend why we must select the candidate-features he proposes, human preferences. As an alternative, consider the religious proposal that what matters about actions is their acceptability to God. This is a genuine alternative; the standard is what God can accept, not a human preference to do only what God can accept. It will not do to reply that probable effect on human preferences is a more sensible way of testing moral prescriptions than divine acceptability. For the criterion must, in Hare's view, involve the logical properties of the moral concepts rather than a substantive moral intuition of sensibility. I do not see how we are compelled as a matter of logic, because of the universalizability and prescriptivity of moral judgments, to choose probable effect on human preferences as the sole, or even one, relevant feature of actions and situations — however convenient that may be for utilitarianism. If no such logical compulsion exists, then preference-satisfaction utilitarianism rests on an unacknowledged substantive moral intuition which Hare, to be consistent with Hume's Law, must not use.

One cannot close the gap by defining utility as, say, pleasure or happiness or welfare, rather than preference-satisfaction. None of these other potential candidates has a logical claim to being the sole relevant feature of human actions or situations for the purposes of universalization. Moreover, some aspects of Hare's position make preference-satisfaction the only serviceable candidate.²⁷ Given his prescriptivism, he needs preferences, since prescriptions "are their expressions in language."²⁸ He also uses preference-satisfaction as the touchstone of utility in disposing of an objection to utilitarianism. The objection is that utilitarianism is unacceptable because, if there were a machine that could dispense unalloyed pleasure, it would best promote utility to have everyone connected to it rather than live other sorts of lives, for the machine would produce the most pleasure — a counterintuitive consequence.²⁹ Hare replies that while the machine might yield the most pleasure, most people would not *prefer* being hooked up to it over other sorts of lives available to them (pp. 142-44).

26. Sections 5.2 and 5.3 (pp. 90-96).

27. Hare allows that, if happiness were equivalent to the satisfaction of a restricted class of preferences, his theory would be a happiness theory of that kind. Pp. 103-04.

28. P. 107. The connection may not be quite so tight. Prescribing is one sort of speech act that might be used to express a preference, but there are many others — requesting, ordering, imploring, entreating, refusing, rejecting, and so on.

29. For the example and another response, see Smart, *supra* note 2, at 8-25.

IV. PREFERENCES AND THEIR ALTERATION

We must now look more closely at Hare's discussion of preferences and how one should evaluate and, possibly, alter them. It will be easiest to see the force of our critique if we first suspend, and only later reinstate, some of the arguments in Parts II and III. Assume, then, that we can accurately separate linguistic and moral intuitions. Assume also that utilitarianism arises solely from linguistic intuitions about our moral words as extracted by philosophical logic. Retain, nevertheless, the criticism that moral intuitions are sounder and linguistic intuitions and methodology are less secure than Hare claims. Given this background for inquiry, what reasons are available for evaluating and perhaps changing people's preferences?

To appreciate the bearing of this question on Hare's version of utilitarianism it is necessary to know a bit more of his account of preferences. For him the principle of utility requires maximizing the satisfaction of one's current preferences ("now-for-now" preferences) and the preferences one will later have for what should happen then ("then-for-then" preferences).³⁰ He stresses that, subject to universalizability, we are free to prefer whatever we prefer (pp. 225, 226, 228). Moreover, in comparing preferences of different persons, we may look only to the respective strengths, not the content, of the preferences. "[E]qual preferences count equally, whatever their content" (p. 144). Put more fully, at the critical level no cognizable difference exists, save possibly in strength, between, say, the Marquis de Sade's preference to torture people and Mother Teresa's preference to help the poor, even if at the intuitive level we should discourage the former and encourage the latter (pp. 141-42).

It is clear that Hare believes preferences to be subject to rational evaluation and change. "Preferences are certainly often alterable," he writes, "and this fact has very wide implications for utilitarian theory" (p. 180). In discussing egalitarianism and a hypothetical society whose members are content with existing inequalities, he observes that, nevertheless, "it may well be right to seek to change social attitudes, if change would be for the best" (p. 159). The reasons that Hare's utilitarian can supply, of course, are reasons of utility. They must appeal, as his discussion of the push-pin/poetry problem makes clear, to the greater preference-satisfaction that *would result* if people were exposed to the reasons and in consequence changed their preferences.³¹ Or, to put the matter more cau-

30. See pp. 101-06, especially p. 105. This does not include the satisfaction of preferences that one currently has for what should happen later ("now-for-then" preferences). See *id.*

31. Pp. 144-46. Compare J. BENTHAM, *The Rationale of Reward*, in 2 THE WORKS OF JEREMY BENTHAM 189, 253-54 (J. Bowring ed. 1843) (if the pleasure is the same, push-pin is as valuable as poetry), with J.S. MILL, UTILITARIANISM 9-15 (O. Piest ed. 1957) (higher pleasures are more valuable than lower pleasures).

tiously, people would change, if not their preferences alone, some combination of their preferences and the world.

Notice that Hare's utilitarian must invoke on the whole *rational argument* here rather than, say, coercive persuasion ("brain-washing"), or subliminal, chemical, or surgical transformation of preferences, or even specious argument. The reason is partly to forestall the abuses, and eventual disutilities, that would stem from such nonrational modes of altering preferences. But it is also to hew to the highly rationalistic spirit of Hare's theory. Even if greater pleasure would come from the judicious use of nonrational ways of changing preferences, it might be said that a higher-order preference is in play. Whether we prefer to satisfy a certain preference may depend in part on the genesis of the preference; and on the whole we might prefer not to have to satisfy preferences produced by nonrational means.

If, however, we can essentially make only rational arguments invoking preference-satisfaction, at least two factors constrain the reasons that we can give for changing preferences. One factor is existing preferences. Consider a social setting, as perhaps in ancient Rome, where countless citizens prefer watching gladiators locked in mortal combat to any other available activity and the combatants have no strong objection (it is, after all, a job). If it is wrong for gladiators to kill and be killed in order to satisfy the bloodlust of spectators, the utilitarian must show how the rational evaluation of preferences arrives at the same verdict. Hare suggests that, at the critical level, this happens as follows:

The right thing to have done from the utilitarian point of view would have been to have chariot races or football games or other less atrocious sports; modern experience shows that they can generate just as much excitement. [P. 142].

This answer is unsatisfactory because it does not take account of the social psychological setting in which preferences exist. Let it be granted that athletics today generate passionate enthusiasm. But the issue is whether people in the time of a hypothetical Diocletian could be convinced to prefer some sport to gladiatorial combat. It would perhaps be necessary to transform a good deal of the social context and frame of mind to effect such a change in preferences. For all we know, football might have as much spectator appeal for our hypothetical Romans as lawn bowling does for most of us. Perhaps it will be suggested that we should deprive one imagined Roman generation of their gladiators in order to produce a robust preference in succeeding generations for football. It is unclear to me whether this suggestion comports with Hare's account of then-for-then preferences. A given generation might not have so much sympathy for future generations that its members have then-for-then preferences that football supplant combat. In any case, the difficulty remains that

utilitarians always confront an existing set of preferences, and it is hard to show that they can always maximize preference-satisfaction by moving from the existing to a different set of preferences that they see as more innocuous.

A second factor is the cost of altering preferences. It consumes resources — time, energy, perhaps money — to exchange some existing preferences (combat, bullfighting, pop music) for new preferences (football, gardening, classical music) that Hare and most of his readers would find more acceptable. If we focus solely on the respective sets of preferences, we may find that more satisfaction would result from new preferences. But people also have preferences regarding the use of resources. Hence they may prefer that a given quantity of resources be employed in some way other than to move from an existing set to a new set of preferences, even though, if the move consumed no resources, they would derive more satisfaction with the new preferences. In short, the cost of altering preferences will block some exchanges of preferences that, were the exchanges costless, otherwise would be made.

It would mistake the purport of my argument for a defender of Hare to invoke against it his standard response to anti-utilitarians. In several places³² Hare responds to arguments which hold that a thoroughgoing utilitarianism at the critical level leads to the selection of principles which run counter to our moral intuitions. Hare's responsive strategy relies on his distinction between the intuitive level, where substantive moral intuitions regarding the content of preferences are allowed, and the critical level, where they are not. It then places the burden of proof on anti-utilitarians to show that, in the actual world to which intuitive principles apply, critical thinking judges some preference to be both evil *and* promotive of utility. Thus, Hare challenges anti-utilitarians to produce an example of such a preference in the real world to which our moral intuitions are suited, rather than fanciful worlds where, perhaps, sports fans crave blood and athletes willingly expose themselves to the risk of being killed. Now it might be questioned why the burden of proof should be on anti-utilitarians, and why, if utilitarianism and some moral intuition are inconsistent, we should eliminate the inconsistency by discarding the intuition rather than abandoning or limiting utilitarianism. It might also be questioned whether the gladiator example only belongs to a fanciful world rather than an earlier period of our own world; to pursue that issue would require a solid knowledge of Roman social history. But even if we resolve these issues in Hare's favor, his standard response to anti-utilitarians would, if directed to my argument, miss the point.

My aim is not to refute utilitarianism but simply to argue that,

32. See pp. 131-35, 180-82.

because of the twin factors of existing preferences and the cost of changing them, serious constraints may exist on available utilitarian reasons for altering preferences. If so, then it makes sense to see whether other reasons might be adduced. In Part II, I argued that moral intuitions, suitably purified and pruned, might be a much sounder and more secure guide to action than Hare allows. I venture, for instance, that virtually all readers of this essay, were they to follow some appropriate nonutilitarian method for sifting moral intuitions, would find themselves in possession of the intuition that any preference to watch gladiators fight to the death should be heavily discounted if not disregarded entirely.

If we now put the two strands of the argument together, it follows that we sometimes shall have greater confidence in such moral intuitions than in the capacity of utilitarian thinking to handle a given case properly. For the latter must, at the critical level, count equally all preferences of equal strength, no matter what their content. To argue critically that a preference should be changed can be a risky business. Such an argument must rely on complicated human reasoning which, as we well know, often proves faulty. It must sometimes embrace the hazards of trying to understand people who are very different from us. And it can require laborious empirical investigations. The investigations would be needed to ascertain existing preferences and their strengths, the satisfaction that would result from a new set of preferences, and the cost of moving from one to the other. Hare is aware that some practical difficulties exist (pp. 122, 127). Yet he does not seem aware of them all. He appears, in any event, to underestimate how fraught with difficulty, even in extraordinarily favorable circumstances, utilitarian thinking at the critical level will be.

We might enlarge on the point in this way. Hare in effect tries to make moral philosophy secure by resting it on what seem to him the most parsimonious and least vulnerable foundations. This procedure does not necessarily yield the best moral philosophy. For it carries with it the uncertainties and difficulties of taking preferences neutrally and resting a great deal on universalization. If what I have said about the soundness of some moral intuitions is correct, a satisfactory moral philosophy should be able to discount or eliminate some preferences because of their content. The preferences of the Marquis de Sade for inflicting sexual tortures (at least on unwilling victims) and of Roman citizenry for mortal gladiatorial combat should, and not only at the intuitive level, be stricken or discounted on the score of content, rather than counting them equally (with allowances for differences in strength) and seeing how the calculation of preference-satisfaction works out. It is, I believe, both difficult and unwise to offer a moral theory resting so much on preferences without an accompanying theory of value or human nature that can

help determine which preferences count and for how much. Such an accompanying theory would involve what Hare calls intuitions of moral substance. It could not count, therefore, as critical moral thinking as he understands it. Accordingly, a corollary of my point is that we should redefine the critical level to allow certain restricted appeals to substantive moral intuitions.

All the same, more can be said for Hare's position of taking preferences neutrally than many philosophers acknowledge. If the questioned preferences are voluntary and authentic, and if the available alternative preferences would yield less satisfaction, our justifiable confidence may be lessened in moral intuitions that would disallow or discount the questioned preferences. This is especially true where the intuitions may be the result of bias. A possible example is the moral intuition of many people that the sexual preferences of homosexuals should be rejected or discounted. So Hare's utilitarian neutrality may have its advantages. It also has its shortcomings. If we can arrive at a sound nonutilitarian procedure for certifying moral intuitions, then, as argued above, we can sometimes repose more trust at a critical level in those intuitions than in our ability to calculate preference-satisfaction.

V. CONCLUSION

So far I have suggested that some moral intuitions are sounder than Hare allows and that they can afford a way, at a critical level, of evaluating and altering preferences. This suggestion does not require the arguments from Parts II and III that we suspended earlier. But let us now reinstate those arguments and make out a more ambitious claim.

Assume, then, that we cannot always separate accurately linguistic from moral intuitions. Assume also that even Hare's utilitarianism rests on a substantive intuition of its own, namely, that probable effects on preference-satisfaction are morally relevant features of actions. On these assumptions the presence of intuitions of moral substance is, at the critical level, not only acceptable but nearly inevitable. If so, we have somewhat rehabilitated the place and security of certain intuitions in critical moral thinking. I do not, of course, mean that there is a special psychological faculty that intuits moral truths, or even that there necessarily is something that we can correctly regard as objective moral truth. Nor do I exclude from acceptable critical moral intuitions the intuition that probable effects on preference-satisfaction, or, more broadly, consequences of human actions, are extremely important in determining right and wrong. On the contrary, this intuition is likely to carry considerable weight in most cases.

Still, it is also likely not to be the only intuition suitable for use in

critical moral thinking. As I have argued, we shall sometimes properly have more confidence, at a critical level, in certain apparently nonutilitarian intuitions than in the central intuition of Hare's utilitarianism. Thus, we shall have an irreducible family of refined and critically acceptable intuitions. Elsewhere I have aired doubts that it will ever be possible to arrive at firm priority rules for resolving conflicts among them.³³ So if to defend intuitionism is to argue for the plausibility of this sort of critical moral thinking, then this has been a defense of intuitionism.

33. See Munzer, *Persons and Consequences: Observations on Fried's Right and Wrong* (Book Review), 77 MICH. L. REV. 421, 430-31, 444-45 (1979). Such doubts commit me to intuitionism in Rawls' sense as well as Hare's. See note 8 *supra*. For a similar view, see Urmson, *supra* note 8. Hare's book convinces me, however, that earlier I was too hasty in believing that a utilitarian perspective will fairly frequently prove unsatisfactory. Yet if my arguments here are correct, it will still sometimes be unserviceable at a critical level.