The Utilitarian Ethics of R. B. Brandt

ANDREW MOORE

University of Otago

This collection of influential ethical papers by one of the twentieth century’s foremost American moral philosophers is greatly to be welcomed. The essays reprinted here were first published between 1965 and 1988, and were previously scattered far and wide. Many of them appeared originally in top journals, but some would previously have been missed by most cash-strapped university libraries. The essays are grouped under five section headings, prefaced by sixteen very helpful pages of ‘Introductory comments’. The sections are: Metaethical Preliminaries (five essays); Normative Ethics: Utilitarianism (three essays); Utilitarianism and Rights (two essays); Determinism, Excuses, and Character (four essays); and Implications of Utilitarian Theory (four essays). This study focuses almost exclusively on Brandt’s core writings on utilitarianism, and it fails to do full justice to the richness of his work even on this topic. Amongst the many excellent papers otherwise passed over here, it is perhaps fair to single out for mention Brandt’s ‘Morality and its Critics’, a superb drawing together of many of his views on ethics, in the context of a critique of Bernard Williams’ moral scepticism.

Morality, Utilitarianism, and Rights is by no means a comprehensive collection of Brandt’s writings on ethics, and some regrets were inevitable, whichever selection was made. More serious than the embarrassment of riches in selection is the omission from this volume of any bibliography of Brandt’s ethical writings. It is true that a full bibliography up to 1978 may be consulted in the Values and Morals volume edited by Alvin Goldman and Jaegwon Kim, but Brandt has published many important papers in ethics since then.

Any assessment of Richard Brandt’s contribution to normative ethics depends in large measure on an evaluation of rule utilitarianism. This theory has fallen on hard times in the last thirty or so years. Some have made influential objections, others have voiced what has become a widely shared scepticism about the independent existence of

1 A Critical Study of Richard B. Brandt, Morality, Utilitarianism, and Rights, Cambridge, 1992. All page references to the volume are contained in brackets in the main text.
any rule theory,\(^3\) and an often poorly formulated series of debates between rule and act utilitarianism has now largely been abandoned.

In the midst of all this, Richard Brandt and John Harsanyi are the two major philosophers who have kept the rule utilitarian flame burning. Brandt’s two most important essays in defence of his ideal form of the theory are reprinted here (‘Some Merits of One Form of Rule-Utilitarianism’ (1965), and ‘Fairness to Indirect Optimific Theories’ (1988)), though his valuable ‘Problems of Contemporary Utilitarianism: Real and Alleged’ is also well worth consulting.\(^4\) Utilitarians will need little encouragement to attend to these papers, but anti-utilitarians should do so too, for they will want to oppose themselves to the best theory in their rival’s camp. To determine whether rule utilitarianism is that theory, a close look at Brandt’s work is mandatory.\(^5\)

In ‘Some Merits’, Brandt puts his view thus:

an action is right if and only if it would not be prohibited by the moral code ideal for the society in which it occurs, where a moral code is taken to be ‘ideal’ if and only if its currency would produce at least as much good per person as the currency of any other moral code. (124–5)

‘Fairness to Indirect Optimific Theories’ offers a lightly revised account:

any other-person-involving act is morally permissible if it would be best for the moral motivations of (roughly) all agents to permit acts of that sort in its circumstances, and ... an other-person-involving act is an agent’s moral duty if it would be best for the moral motivations of (roughly) all agents to require acts of that type in those circumstances. (138)

Some of the revisions are noteworthy. The first formulation does not, whilst the second does, distinguish moral permissions from moral duties. The earlier formulation ranges over actions in general, whilst the later account deals only with other-person-involving acts, and thereby suggests that purely self-involving acts do not fall within the moral domain. The commitment to averaging utilitarianism in the earlier account is absent from the later formulation, and from the later paper. Finally, a maximizing or optimific form of welfarism and a societal constituency for the code of moral rules are prominent in the

\(^3\) See, for example, the ‘rule worship’ objection in J. C. Smart’s influential critique, variously reprinted in revised forms, ‘Extreme and Restricted Utilitarianism’, *Philosophical Quarterly*, vi (1956), 344–54. For the influential claim that rule utilitarianism collapses into extensional equivalence with act utilitarianism, see David Lyons *Forms and Limits of Utilitarianism*, Oxford, 1965.

\(^4\) In N. E. Bowie, ed., *Ethical Theory in the Last Quarter of the Twentieth Century*, Indianapolis, 1983.

\(^5\) On the other hand, some anti-utilitarians will be disappointed that Brandt sets aside objections, such as those from justice, which have been urged against all forms of utilitarianism (136).
earlier formulation, and although absent from the later formulation, they remain in the later paper. Where any of these differences matter in what follows, I shall take it that the later formulation gives us Brandt’s current position.

Brandt’s theory is a form of ideal rule utilitarianism. Moral permission and moral duty are determined by an ideal set of moral motivations or rules, by way of a two-stage process. We are to imagine that more or less everyone in our society accepts whichever moral code would, at this high level of acceptance or currency, maximize welfare. Brandt suggests, but is not wedded to, a minimum of 90 per cent of adults (120, 9). By reference to the rules and principles contained in that ideal code, we then determine whether any given act in our non-ideal circumstances is morally prohibited, permitted or required. I may favour my children over others, for example, only if the moral code which would be welfare optimific under conditions of ideal acceptance would permit this sort of thing.

The now-familiar objection that rule utilitarianism collapses into a form of act utilitarianism is nicely discussed by Brandt. In the earlier paper he argues that there is no reason to expect there to be either a first-order or a conflict-resolving act utilitarian rule in the ideal moral code (130). In the later paper he addresses the standard worry that rule utilitarianism must always make allowance for utility maximizing exceptions to its rules, and that the ensuing endless series of rule refinements collapses the theory into extensional equivalence with act utilitarianism. Brandt has a good reply. He points out that rules in the ideal code must ‘be simple enough to be interiorized’ (9; see also 143, 150f), for otherwise they could be neither adopted nor maintained. A rule-set extensionally equivalent to act utilitarianism would be utterly indigestible, however, so Brandt’s theory is invulnerable to death by a thousand refinements.

Brandt’s discussions of act utilitarianism in these papers seem to me less successful. His main reason for rejecting the theory is the pragmatic one that it would be a self-defeating method of moral calculation, would undermine co-operation, and would have various other pragmatic defects (142f, 153). To the reply that act utilitarianism need only be a criterion of right action, not a decision procedure or mode of moral thinking, Brandt makes rather puzzling appeal to a contractualist meta-ethical constraint on the justification of normative theories (140).

6 In his later paper, Brandt calls his view an ‘indirect optimific theory’, on the grounds that he assesses actions not directly, in terms of their own consequences for welfare, but indirectly, in terms of the consequences for welfare of rules which permit or require them. Many other writers use ‘indirect’ to refer instead to those theories with utilitarian criteria of right action which allow moral thought and motivation to be non-utilitarian. I much prefer Brandt’s usage, but I shall side-step possible confusion here by using only Brandt’s earlier term, ‘ideal rule utilitarianism’. 
He goes on to suggest 'some degree of publicity as one of the prerequisites for an optimal moral system' (144), and he understands this to require each principle in that system to be learnable, and to be recognized by most adults to be part of the shared code (120, 136). Leaving aside the interesting question whether this makes Brandt fundamentally a contractualist, his argument does seem to miss its target. Act utilitarianism too has a moral code, and it is learnable, public, and very similar in content to a rule utilitarian code, as Brandt himself observes (150). It might arguably be chosen by informed rational persons as the code for their society. From a theoretical viewpoint, its rules are only rules of thumb, but it is not clear why that should matter. The act utilitarian criterion of rightness is not part of its code, of course, but neither is the rule utilitarian criterion part of its code. Nor does learning, teaching or applying the act utilitarian criterion of right action seem to generate any problems more formidable than those arising for Brandt's rule utilitarianism, with its very complex counterfactual deliberations about the content of an ideal code, and the acts which it would approve. Indeed, if I have correctly understood his later remarks on this subject (153-4), Brandt himself now has a rather sceptical view about his case against act utilitarianism.

Returning to Brandt's own theory, a vital issue arises over what his ideal moral code would include. He says it 'must contain "rules" dealing with recurring problem situations, . . . not leave too much to discretion, not ignore the costs of teaching or conforming, or impose unbearable demands' (9). The ideal code would obviously include various moral prohibitions, permissions, and requirements, and Brandt says these would vary in their 'strength' (126). It would also include a specification of moral sanctions and rewards, together with an account of exculpating and mitigating excuses (126). Amongst other things, it would require us to keep promises, to refrain from false swearing (154), to keep contracts, make restitution for injuries, help others in distress, and perhaps bring about a just distribution of goods (9).

Troubles arise for the ideal rule utilitarian code, however, as soon as one gets into detail about a particular rule, or a particular case. Reflection on the bindingness of promises and on their various releas-

---

1. Brandt claims that act utilitarianism imposes unbearable demands (9), and that it cannot give a proper account of the personal nature of morality (131). These matters have now been very widely discussed. Only very recently, however, have scholars seriously investigated Brandt's suggestion that rule utilitarianism can do much better on these matters. See Brad Hooker 'Rule Consequentialism', Mind, xcix (1990), 67-77, and 'Brink, Kagan, Utilitarianism and Self-Sacrifice', Utilitas, iii (1991), 263-73.

2. Brandt's 'A utilitarian theory of excuses' and 'A motivational theory of excuses in the criminal law', both reprinted here, set out and defend his detailed views on the subject.
ing conditions, for example, soon reveals extraordinary complexity. Brandt acknowledges this (149, 154), but the trouble is that very little complexity can be sustained in rules for the acceptance of almost everyone. By keeping complexity firmly out of the code, one might be able to secure or maintain acceptance by almost everyone. But it does seem that this would only be at the cost of moral mediocrity (157); that the morality of ‘What if everyone did the same?’ would thus be transmuted into the morality of ‘Only if everyone can do the same’, and thereby a code too low for humanity. On the other hand, Brandt’s theory may at the same time be too high for some of humanity. With their moral psychology in its present non-ideal state, it seems very likely that many individuals are not currently capable of the acts required of them by the ideal code.

Brandt candidly admits in his closing words on the subject that ‘the optimific indirect theory does run into complications when we try to work out the details in a realistic way’ (157). Another way into these complications is to ask how in detail rival candidates for the ideal moral code are to be assessed. In particular, should one include the costs and benefits of establishing the code, or should one include only code maintenance? Suppose we are to include code establishment. To say the least, there would be major costs here. It may reasonably be expected, for example, that communities with minority moral views would resist any drive to bring them into the moral majority. It is also to be expected that many counter-suggestible individuals would set about acquiring non-code and anti-code motivations, as a way of asserting their moral individuality in the face of pressure for unanimity. Stubbornness and contrariness would thus enormously raise the costs of getting any moral code whatsoever established to 90 per cent levels. One might call this the missionary problem. In morality as in religion, disappointment would almost certainly face any attempt to establish a single dominant view. In the light of this, one may reasonably doubt that there is any single moral code ideally establishable in our society. Yet if there is no such code, Brandt’s theory gives us nothing with which to determine moral right and wrong.

Suppose establishment costs are to be left out, and that all we need worry about is maintenance of the ideal moral code at 90 per cent levels of adult acceptance. In morality as in religion, however, there are strong and heterodox currents of both of individualist and communitarian sorts. One might refer here to the ‘orthodoxy problem’, and again one might reasonably suspect that there is no ideally sustainable moral...
code. Yet if there is not, Brandt's theory gives us nothing with which to determine moral right and wrong.

Various replies can be made to the 'no ideal code' objection. Ideal rule utilitarians might say they are interested in a code little different from the common sense morality which already enjoys wide currency. Let them then produce a formulation of common sense morality with which 90 per cent of our society's adults really could be sustained in agreement. An alternative is to go for a cut-down variant of common sense morality, but even in quite simple concrete cases, this would offend against Brandt's sensible requirement that codes are to have determinable implications for action (113), and 'not leave too much to discretion' (9). A further possibility is to leave resilient moral controversies out of the ideal code. But that would mean loss of code comprehensiveness. Would we then say that all actions not mentioned in the code are morally permitted by it, or that none of them are? Both answers are implausible.

Brandt suggests that one could lower the acceptance levels demanded of the ideal code (157), say to 70 per cent or 50 per cent. The troubles then would be twofold; debate on many of the most difficult and interesting moral issues would be pre-empted, and the consequences of any candidate ideal code would be much influenced, but in rather unpredictable ways, by reactions to it among the 30–50 per cent moral minorities. Ideal rule utilitarians might say that like the institutional setting and the moral conventions of the day (117, 121f, 125), resilient moral disagreements would be important features of the context in which the consequences of alternative codes are to be assessed, and that any plausible ideal code would have to say something about these contentious matters. The code might perhaps include a rule that we are to deliberate like act utilitarians when no other rules deliver determinate results. Or it might simply let the greater force of moral motivation win the day, or allow the individual to be a kind of appellate court (146). The main trouble with these suggestions is that the ideal rule utilitarian moral code is no longer doing the work. A further alternative in especially controversial areas is to go for procedural rules, according to which any outcome arising from following some proper process would be morally permissible. In the light of these various possibilities, it may be that I am overly pessimistic about the appetite of my fellows for moral uniformity.

On Brandt's view, the rightness of acts is determined by the moral
code that would be impartially optimific, were it to have ideal currency in one’s society. As he observes (147–8), his view is thus vulnerable to many of the difficulties which bedevil moral relativism. How, for example, is one to individuate societies: on a nation-state basis, an ethnic basis, a professional basis? Brandt first suggests that the issue may not be important, because the central moral principles would be roughly invariant (147). But this is false, for code maintenance costs, and thereby code complexity and content, would be enormously affected by the size and sophistication of the society in question. Brandt then suggests that societies could be carved up on whatever basis would be most beneficial (148). But this is very hard to construe, particularly given that the code is to be ‘one that does best for a given society’ (9). Brandt proposes ‘not to worry about this question’ (148), but his critics will not be satisfied.

There are numerous relativism-related worries for Brandt. What are we to say of the acts of people who, on any plausible societal carve-up, are members of several distinct and morally divergent societies? Surely, and contrary to Brandt’s first formulation of his rule utilitarianism (124–5), the rightness of one’s actions is not determined by which society one happens to act in on any given occasion. Well known and intractable troubles arise also concerning the morality of interactions between societies with different codes. And should one take the good of all societies, or only the good of the society in question, as the measure of its ideal moral code? Brandt gives the second answer (9), but this is implausibly chauvinistic.

In the light of the major relativism-related difficulties for Brandt’s ‘societal code’ form of rule utilitarianism, one might think a ‘universal code’ utilitarianism or consequentialism would fare better. But Brandt has a powerful reply to this suggestion: ‘I would guess that any moral system composed wholly of rules so abstract they could be of world-wide application would be costly, or even impossible, to teach’ (147). This is a form of the ‘no ideal code’ objection urged above against Brandt’s own societal view, and the difficulties are more acute still for any universal form of the theory.

Brandt’s theory determines right action by reference to the rule-set that would be optimific in ideal circumstances. Given than many of us do not currently endorse the ideal rule-set, however, might it not sometimes be disastrous for us to act in accordance with it? Brandt calls this the Partial Compliance problem (154). He points out that ideal codes can, just as actual codes do, prescribe for situations

---

II Thus Brad Hooker, for example, in ‘Rule Consequentialism’, *Mind*, xcix (1990), 67: ‘an act is morally right if and only if it is called for [or allowed] by the set of desires and dispositions the having of which by everyone would result in at least as good consequences judged impartially as any other’.
involving their own breach (127). One of his suggestions is that the ideal would include requirements to avoid serious harm (156–7). To the objection that such disaster prevention rules would surely not be needed in the ideal, Brandt can reasonably respond that even in the ideal, there would be those who would not accept the shared moral code, and those who would deliberately break it (125–6). Even among code believers, furthermore, there would not be full behavioural compliance, for many individuals would retain conflicting extra-code motivations of both a moral and a non-moral sort, and there would no doubt also be some weakness of will. If there would be disaster-prevention rules in the ideal, then disaster prevention is open to us non-idealized folk. Nor need this amount to a collapse into extensional equivalence with act utilitarianism, for not all failures to maximize utility are disastrous.

Brandt perhaps has a fully satisfactory reply to the 'disaster objection', but that difficulty is itself generated by a deeper structural problem with his theory. A basic thought of utilitarianism, as of other forms of consequentialism, is that consequences fundamentally determine rightness. A vigorous and well known debate continues between those who favour expected-outcome consequences, and those who favour actual-outcome consequences, but it is generally agreed that either way, it is this-worldly consequences which fundamentally determine rightness. On Brandt's theory, however, those consequences which would ensue, were the world ideal in a certain way, are the only fundamental determinants of rightness. This-worldly consequences are made fundamentally irrelevant to the determination of right action. The consequentialist ambition of this-worldly improvement is thereby abandoned for an other-worldly ambition to do what would make the world a better place, were it only an ideal world.

One reply to the other-worldliness objection is to say that the implications of the ideal moral code for action would in all significant cases be world-improving. This misses the point, however, for the objection is directed at the fundamentally other-worldly character of Brandt's ideal rule utilitarian criterion of right action, and not at its substantive implications. No rule that might be hypothesized into the ideal code will silence the other-worldliness objection either, for the objection is directed against the fundamental standing of that code in the criterion of right action, whatever the code content might be. Another reply would deny that the theory has a fundamentally other-worldly character, on the grounds that Brandt does not idealize the conventional and institutional setting (132, 149). But this too is beside the point. Even if actual conventions and institutions have a bearing on a Brandtian assessment of moral codes, the fact remains that on his theory, consequences which would ensue in ideal conditions, and not
consequences which do or are expected actually to ensue, are the only fundamental determinants of right action.

A further reply to the other-worldliness objection is to say that it fails to establish the implausibility of Brandt's theory, and at most establishes that he is a non-utilitarian. There are, after all, many ideal observer and ideal contract moral theories, and the 'other-worldliness' of their various criteria of right action generates no obvious absurdities. My criticism, however, centres on Brandt's combination of welfare consequentialism with fundamental other-worldliness. Rawls has said, plausibly, that: 'All ethical doctrines worth our attention take consequences into account in judging rightness. Any theory which did not would be irrational, crazy.' My claim is that if one does take consequences fundamentally into account in judging rightness, as Brandt does, it is crazy to give that role to other-worldly rather than to this-worldly consequences.

Maximization of utility presupposes some account of what utility or welfare is. Brandt's papers on rule utilitarianism set the issue aside, but it is the subject of his 'Two concepts of utility'. He begins by setting out what has become the standard three-way distinction in this area, between hedonistic or happiness theories, desire theories, and ideal theories. His interest is in theories of the first two sorts. After some insightful preliminary remarks on the nature and varieties of desire theories, Brandt moves to the core of the paper, a powerful objection to any theory of welfare made out in terms of desire fulfilment. Considering only very simple one-person cases, he points out that what we desire, and how intensely we desire it, changes greatly over time. Some desires come and go in ways more or less predictable on developmental or 'stage of life' grounds. Others are brief but very intense, others again are quiet and long-lived, and others still leave and return to us cyclically (169-70). Brandt then very nicely brings out the cross-currents which arise when one takes into account, as one should, certain past and future desires. And he emphasizes the enormous complexities generated even among various present desires, to continue present states, to secure future states of affairs, and to have done things differently in the past. Desires of all these sorts must somehow be combined in a single function of maximal desire fulfilment, and the matter is complicated further still by that fact that our many different

13 Brandt uses 'utility' and 'welfare' interchangeably, and I shall stick with the latter concept. On the multiple ambiguity of 'utility', see John Broome, 'Utility', Economics and Philosophy, vii (1991), 1-12. To avoid confusion, I shall also prefer 'desire fulfilment' to Brandt's 'desire satisfaction'. Brandt is crystal clear that he means only the obtaining of that which is desired, but very widespread recent usage understands 'desire satisfaction' also to include the pleasure or enjoyment arising from this.
possible future desire-sets depend in part on which of our past and present desires have been or will be fulfilled.

In my view, desire theorists have still to work out a fully coherent notion of welfare maximization in response to these difficulties raised by Brandt. James Griffin’s recent theory does impressively take up the challenge, but even if successful, it appeals to an irreducibly normative account of informed desire, and as Griffin recognizes, this leaves it by no means clear that the notion ‘desire fulfilment’ retains a working theoretical role.\textsuperscript{14} Brandt’s objection to desire fulfilment theories of welfare does, however, raise two important questions about his own theories. The first is that Brandt himself defends theories of rational desire and action which employ the notion of desire fulfilment.\textsuperscript{15} If the notion of maximization is problematical in desire theories of welfare, it seems it will be no less troublesome in desire theories of rationality. The second question concerns Brandt’s view that a person is happy or enjoying a certain experience ‘if and only if the experience is making him want to continue it (or repeat) for itself’ (164). This motivational theory of happiness seems to make essential appeal to the concept of desire fulfilment, and consequently to inherit some of the problems facing desire theories of welfare.\textsuperscript{16}

Finally, Brandt acknowledges in passing one of the basic problems for any enjoyment or happiness theory of welfare (165). Lives of infantile bliss or ambitionless pleasure in whatever life serves up to us appear to be top qualifiers on his theory, yet it is clear that we would be doing our children no favours were we successfully to raise them for such a future. Brave New World still haunts the happiness theory, and any utilitarianism grounded in it.

This study has concentrated on certain difficulties in Brandt’s utilitarianism, but that would be a quite inappropriate note on which to conclude. Every essay in this wide-ranging collection is of exemplary clarity, and Brandt’s positive proposals are always suggestive and illuminating. So scrupulous is he in pointing out himself the problems his views face that it is very difficult to make any wholly original point against him. In short, those who read this book can expect to enjoy and be challenged by ethical writing of a very high calibre.


\textsuperscript{15} Brandt’s ‘Rational Desires’ is reprinted in the volume under review. See also \textit{A Theory of the Good and the Right}, Oxford, 1979, and ‘The Concept of Rational Action’, \textit{Social Theory and Practice}, ix (1983), 143–64.

\textsuperscript{16} Brandt acknowledges and briefly discusses this problem (174–5). For more on his happiness theory, see \textit{A Theory of the Good and the Right}, especially ch. 13, and ‘Fairness to Happiness’, \textit{Social Theory and Practice}, xv (1989), 39–58.