AGAINST

UTILITARIANISM

Champions of utilitarianism regularly insist on its thoroughgoing precision and comprehensiveness; even its critics often grudgingly agree. Here I argue that utilitarianism is intolerably sketchy and, more important, that we are given no good reasons to adopt the utilitarian standard.

I proceed by explaining what I take utilitarianism to be and undertaking some textual exegesis of Bentham, Mill, and Sidgwick. Mill, incidentally, I take to be at most an indecisive, inconsistent utilitarian, whose thought cannot be explicated adequately within the confines of utilitarianism. I include him here in deference to convention.

I go on to reject an all too common line of attack against utilitarianism, that of showing that it does

not match our "moral intuitions." The line I adopt instead is that of showing that utilitarianism seems radically incomplete and that the arguments for it are readily refuted. Throughout, I focus on classical utilitarianism, which seems to me the most compelling version. But classical utilitarianism has spawned countless variants, often deemed improvements; so I also examine the theories of Harsanyi, Hare, and Brandt. I conclude with some critical remarks on the foundational structure of utilitarianism.

This chapter wanders away from politics, into terrain occupied largely by economists, and I should explain why. Utilitarianism, with its unswerving devotion to the greatest happiness, presents a marvelously clear picture of the practical syllogism:

We ought to realize the greatest happiness.
Doing X will realize the greatest happiness.
We ought to do X.

Here correct political arrangements would fill in X. I mean to leave aside arguments about what X in fact is. First we need to figure out whether the major premise is right. If it is, then we can worry about the minor premise.

We can note immediately, though, that adopting utilitarianism would mean drastically revising our understanding of politics. It would mean viewing politics as a completely technical activity, simply finding means to a given end, not as anything like a Weberian clash of ultimate ends. This reshaping is only one of the dubious surprises utilitarianism has to offer.

Utilitarianism Characterized

It might be best to begin by explaining what I do not take utilitarianism to be. I take a teleological theory to be one identifying

---

the right as that which realizes the good, and a consequentialist theory to be one holding that actions, dispositions, rules, and the like are to be evaluated on the basis of their consequences. Both sorts of theories are often called utilitarian, but I wish to reserve the word for a distinctive family of theories that falls within these broader classes.

I take utilitarianism to be any theory holding that the average or total happiness of the group ought to be maximized. Utilitarians differ on their interpretation of happiness: some conceive of it as pleasure or as agreeable mental states of whatever kind; others as preference satisfaction. I insist that happiness be construed descriptively and that it be tightly linked to mental states for a theory to qualify as utilitarian. Utilitarians differ too in their strategies for maximization: act-utilitarians ask agents to calculate the consequences of each and every one of their actions; rule-utilitarians ask agents to follow rules justified in turn by their consequences.¹ Utilitarians also differ on whose happiness is to be maximized: some include the unborn, some those in other societies, some animals. But the crucial element, however elaborated, is maximization of average or total happiness. I do not, for example, consider Moore’s “ideal utilitarianism” to be a form of utilitarianism at all, but rather just another teleological theory, for Moore nominated other goods to take the place of happiness.³

Suppose we are discussing Bentham’s views on dicastic thelematoscopic pathoscopy pneumatoscopic idioscopy eudaemonics, as he named ethics.⁴ We all know, without bothering to read Bentham, that his principle of utility dictates striving for the greatest happiness of the greatest number; and indeed there is


 AGAINST UTILITARIANISM

textual evidence for that traditional view. Yet, as David Lyons has argued, we actually find in Bentham’s writings “a dual standard, with community interest the criterion of right and wrong in public or political affairs and personal interest the proper standard for ‘private ethics.’” Here I argue that the dual standard is best understood as a strategy for realizing the greatest happiness.

At the beginning of An Introduction to the Principles of Morals and Legislation, Bentham offers “an explicit and determinate account of what is meant by” the principle of utility: “that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words, to promote or oppose that happiness.” Bentham, while hesitant to define interest, is clear that “a thing is said to promote the interest, or to be for the interest, of an individual, when it tends to add to the sum total of his pleasures: or, what comes to the same thing, to diminish the sum total of his pains.” So to understand the explicit and determinate account of the principle of utility, we need know only whose interest is in question. Bentham never, as far as I am aware, argues that when private individuals act, only their own interests are in question, but he apparently holds that view: “By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered: if that party be the community in general, then the happiness of the community: if a particular individual, then the happiness of that individual.” These claims are unequivocal. The government is to pursue the happiness of all; I am to pursue my own happiness. That seems a far cry from utilitarianism: it looks

1For example, Works 1:269n.; and Introduction, p. 285.
2Here I follow, up to a point, David Lyons, In the Interest of the Governed: A Study in Bentham’s Philosophy of Utility and Law (Oxford: Clarendon, 1973); the quoted passage is from p. vii.
3Introduction, pp. 11-12.
4Introduction, p. 12, and 12n.c.
5Introduction, p. 12; also p. 293. But compare p. 284.

[113]
more like utilitarianism for governments, egoism for individuals. The split here is puzzling and demands some explanation.

Let me work backward a moment. Suppose that Bentham’s fundamental principle was in fact the greatest-happiness principle as we generally think of it. Why might he have been led to the dual standard? Because, I conjecture, his psychological theories placed obstacles in the way of realizing the greatest happiness in a more direct way. The problem is clear already in the opening flourish of the Introduction: “Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne.”¹⁰ Here Bentham invites cheap criticism. “How convenient!” we might say. “We are determined by Nature to do what we ought to. Immorality is precluded by the laws of causation.” But this criticism is too quick. We need to appreciate the gravity of the riddle Bentham poses himself in adopting these views.

Each individual, it seems, seeks his own pleasure, yet (I am supposing) Bentham holds the greatest happiness of all to be the good. If I am causally determined to seek my own pleasure, how can the greatest-happiness principle have any impact? Bentham’s hedonism is not purely egoistic. Bentham explicitly allows for simple pleasures and pains of benevolence and malevolence, which he calls “extra-regarding.”¹¹ He does hold, however, that “the only interests which a man at all times and upon all occasions is sure to find adequate motives for consulting are his own.”¹² With his discovery of the sinister interests blocking the success of utilitarianism, he seems to have become even more firmly convinced of the prevalence of egoism.¹³

For Bentham, there is a second psychological obstacle to ask-

---

¹⁰Introduction, p. 11.
¹¹Introduction, pp. 42, 49.
¹³Deontology I:13, 18, 28–29, 191; II:15, 35, 72–73, 121, 133, 155; Works IX:5–6, X:80. The Deontology is of doubtful authenticity, and I will try here not to rest any interpretation on its pages. But I do think it is characteristic Bentham.
AGAINST UTILITARIANISM

ing individuals to pursue the group's good. Bentham is skeptical of the possibility of adequately knowing other minds: "By competent attention and observation every man will be best acquainted with the character of his own sensibilities. By countenance, gesture, deportment, contemporaneous or subsequent conduct, he may give indications to others; but no evidence will be so complete, no testimony so direct, as that of his own feelings: thence it follows that, with the benefit of experience, every man is a more competent judge of what is instrumental to his own well-being than any other man." In Bentham's psychological views, then, we have two good reasons for allowing the pursuit of self-interest. It is what people are most prone to do anyway, and each person knows better than others what makes her happy. Bentham's dual standard thus offers a way of maximizing the utility of a group of largely egoistic agents with basically private mental states.

But the dual standard is unsatisfactory. Bentham claims, "If every man, acting correctly for his own interests, obtained the maximum of possible happiness, mankind would reach the millennium of accessible bliss; and the end of morality—the general happiness—be accomplished." Imagine now, though, that transfers of income, or of other happiness-producing goodies, can on the whole (counting in all the indirect effects Bentham and later utilitarians so strenuously insist on) increase total happiness. Imagine, say, that Bentham is right in arguing from the principle of decreasing marginal utility to economic egalitarianism. Then this claim is wrong. Even a doctrine of the harmony of interests between individual and society, one that Lyons attributes to Bentham in the Introduction, may not do the job here, depending on just how we spell out that elusive idea. For

14Deontology I:78-79; see too I:29, 59, 68, 189; II:121, 289-290.
15Deontology I:12, also I:18; Works II:121-122.
16Lyons, In the Interest, pp. 50-81. This is surely the view in Deontology I:18-19, 164; II:36-41, 80, 189-190, 295. Sidgwick, in Methods, pp. 87-88n., sees two principles in Bentham and says Bentham reconciles them by holding "that it is always the individual's true interest ... to act in the manner most conducive to the general happiness"; but as support he cites Works X:560, 561, where nothing of the sort is to be found. Nor, as far as I can tell, is the reference a simple typographical error for some similar location. A sound critical edition of the Methods is badly needed.
The threat is not exhausted by situations in which I can advance my interests by trampling on yours. Some social settings may present prisoners' dilemmas, situations in which the pursuit of private interests produces results irrational from everyone's point of view.\footnote{A delightful sketch of this phenomenon, and related ones, is Thomas C. Schelling's \textit{Micromotives and Macrobehavior} (New York: Norton, 1978). Note too Henry Sidgwick, \textit{Elements of Politics}, 4th ed. (London: Macmillan, 1919), pp. 144–151.}

The dual standard is not the only problematic element of Bentham's utilitarianism. Bentham, as we will see, faces objections for holding that all sorts of mental states are commensurable, and the more he insists on a variety of pleasures and pains—in one manuscript he has fifty-four pleasures and sixty-seven pains\footnote{Works 1:205–206.}—the more plausible the objection seems. He flirts too, especially given his inclination to debunk poetry, with the accusation that utilitarianism is coarse and degrading. Mill, concerned about these issues, introduces a distinction between higher and lower pleasures. He writes that "the creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness," and adds, "By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure." So far we are on familiar Benthamite ground. Mill, however, holds that "it is quite compatible with the principle of utility to recognise the fact, that some \textit{kinds} of pleasures are more desirable and more valuable than others." There follows his notorious procedure for vindicating the higher pleasures: "Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure."\footnote{Utilitarianism, pp. 210–211.}

The coherence of the utilitarian calculus hangs in the balance here. If we cannot plot all pleasures and pains on the same metric, we cannot make an unequivocal judgment that this action, or
rule, or disposition, or whatever else, produces the greatest happiness. One may produce more "higher pleasures," another more "lower pleasures." Perhaps Mill can be successfully defended as a utilitarian here, though the defenses are, I think, a bit too ingenious to rank as compelling interpretations. But Mill presses on, past any recognizably hedonic calculus: "It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied." At this point, efforts to show that Mill is a consistent utilitarian seem more foolhardy than courageous. Surely the right hypothesis is that Mill is torn between a Benthamite utilitarianism and an ethic of self-realization. Commitment to self-realization recurs through his works: witness the chapter on "Individuality" in On Liberty, his account of his breakdown and sympathy for Coleridge in the Autobiography, and the chapter "Of the Probable Futurity of the Labouring Classes" in his Principles of Political Economy. The commitment is not readily accommodated, and indeed may be unintelligible, within the Benthamite psychology so conducive to classical utilitarianism. And the commitment is much to his credit. It ought not to be stamped out or glued over to demonstrate Mill's consistency.

Sidgwick hesitates at being described a utilitarian, and he has good reason for doing so, given his views on the duality of practical reason and his critique of hedonism. I will not examine those views here, however. Instead I will take the Methods of

---


21Utilitarianism, p. 212.


23The argument for higher pleasures and self-development at the end of Logic, p. 952, shows clearly that Mill is willing altogether to surrender a hedonistic conception of happiness.


25Methods, pp. x–xi.


[117]
Ethics as an argument for utilitarianism, for the exposition and defense of the doctrine there offered is clearer and more elegant than either Bentham's or Mill's.

Sidgwick takes the greatest happiness to mean "the greatest possible surplus of pleasure over pain, the pain being conceived as balanced against an equal amount of pleasure, so that the two contrasted amounts annihilate each other for purposes of ethical calculation," and he takes pleasure and pain "to include respectively all kinds of agreeable and disagreeable feelings." He admits "distinctions of quality . . . only in so far as they can be resolved into distinctions of quantity." Finally, he counts the feelings of all sentient beings. It may be appropriate to formalize the notation a bit, for while Sidgwick is aware that calculations will be difficult, he does seek as much scientific precision as he can. A utilitarian wants to maximize

\[ U = \sum_{i=1}^{n} (x_i - y_i), \]

where \( n \) is the number of beings in the group, \( x_i \) is the happiness or pleasure of the \( i^{th} \) being, and \( y_i \) is the unhappiness or pain of the \( i^{th} \) being.

Sidgwick is aware of what must be the case for this view even to be coherent. First, we must be able to place pleasures and pains on a single scale and be able to add and subtract them. More formally, these mental states must be commensurable and cardinal. Second, interpersonal utility comparisons must be possible. It will do no good to tell us to maximize a sum that cannot be summed. If these conditions are prerequisites for the

\[ ^{27}Methods, \text{ pp. } 413, 120-121. \]
\[ ^{28}Methods, \text{ p. } 121. \]
\[ ^{29}Methods, \text{ p. } 414; \text{ compare } Introduction, \text{ pp. } 282-283n.b; Utilitarianism, \text{ p. } 214. \]
\[ ^{30}Methods, \text{ pp. } 123-137. \text{ Sidgwick plumps for a ratio scale, not merely an interval one.} \]
\[ ^{31}Methods, \text{ pp. } 144-150, \text{ where this point is combined with that of comparing pleasures in the same person at different times.} \]
AGAINST UTILITARIANISM

meaningfulness of utilitarianism, a utilitarian is perforce committed to them. Sidgwick worries about each, but thinks we can make sense of the felicific calculus in at least a rough and ready way.

Maximizing $U$, the utilitarian holds, is what morality and politics are all about. Utility, writes Bentham, "is itself the sole and all-sufficient reason for every point of practice whatsoever"; Mill takes it "as the ultimate appeal in all ethical questions"; and Sidgwick, despite his hesitations in the Methods of Ethics, adopts the principle of utility in his Elements of Politics: "For a State, as for an individual, the ultimate end and standard of right conduct is the happiness of all who are affected by its actions."32

These claims are extraordinarily bold and sweeping. I have noted already that adopting utilitarianism would mean thinking of politics as a technical activity. Arguably, it would also mean radically reshaping our understanding of moral life.33 But it would mean, by the same token, a perhaps welcome relief from the endless haggling of moral and political debate. We ought, then, to subject utilitarianism to careful scrutiny to see if we should adopt it.

Utilitarianism Scrutinized

What sort of scrutiny is appropriate? A common procedure is to check utilitarianism against our moral judgments to see if it tells us to do what we know is right. An especially popular argument is that utilitarians are committed to supporting slavery if slavery maximizes utility. Yet, the argument goes on, surely slavery is wrong. The defense has been that the calculations showing slavery to be felicific are incomplete or skewed, that

they don’t pay enough attention to indirect consequences, that they ignore deep-seated facts of human nature.34 (Utilitarians show marvelous ingenuity in developing such defenses, and are lucky to have such a pliable theory.) I want to avoid such discussions, and I will explain why.

Recall the sketch from the introduction on the variety of projects available in moral and political theory: justification, explication, genealogy, and so on. Were utilitarianism to pose as an explication of our moral beliefs, the appeal to slavery and the like would be quite sensible as an objection. Here our moral judgments are appropriately seen as data the theory must fit, and a theory that cannot account for important data is in bad shape. But our beliefs can be explicated whether they are right or wrong, or neither. Utilitarians want to say not that “we all think this is right and wrong, good and bad,” but rather that “this is right and wrong, good and bad.” They are presenting their view as the right one, not necessarily the one we do use, but the one we should use.

So the appropriate considerations are those of justification, not explication. A clear-headed utilitarian can airily dismiss the appeal to the injustice of slavery by saying, “If injustice serves the general happiness, we ought to be unjust.” At that point, it will not do to remind the utilitarian that we think slavery is unconditionally wrong. He knows that, and he thinks we are wrong to be so rigid. Why, he might ask, insist on a view so singularly unresponsive to shifting circumstances? Although utilitarians have generally been anxious to minimize the apparent discrepancies between their code and our shared views, they do sometimes take the less timid line I am suggesting. Bentham’s Deontology may have been extensively revised by Bowring, but it is hard to detect the latter’s work here: “If it could be proved that evil, in the shape of a balance of suffering on the whole, grew out of a

34For example, Bentham, Works I:343–347; R. M. Hare, “What Is Wrong with Slavery,” Philosophy & Public Affairs 8 (Winter 1979): 103–121. The first clear statement of rule-utilitarianism, R. F. Harrod’s “Utilitarianism Revised,” Mind 45 (April 1936): 137–156, was motivated by a desire to reconcile utilitarianism and our moral views: “I conceive it to be the task of the moral philosopher to determine and explain the subject matter of that body of [common moral] opinion” (p. 137).

given line of conduct, and it were agreed that such line of conduct ought to be called just, the consequences would simply be, that justice and virtue might be opposed to one another, and that to be just would be immoral." Once a utilitarian is willing to take this step, our shared moral views are irrelevant.

What, then, shall we say about the voluminous literature devoted to embarrassing utilitarianism by dreaming up examples in which it has morally counterintuitive implications, and the equally voluminous literature defending it against such attacks? Not only has this line of argument proved inconclusive; it also seems quite beside the point. Once we see the difference between explication and justification, it is tempting to write off this entire literature as simply confused. Still, there are two views that prevent our doing so, views that in different ways threaten the clean logical separation between explication and justification.

The first view is that of Prichard and Ross. I can offer only a bald summary statement here. Suppose there are objective moral truths, and suppose we may discover them by the operation of moral intuition, a sixth sense of sorts. There might be ideal observing conditions: Our intuitions might be most accurate when we are impartial, reflecting in a Butlerian cool moment, and so on. But they would be reports on an autonomous realm of moral truths. We need not litter that realm with scads of judgments. We intuit basic principles, and hammer our way through to individual judgments by applying the principles. So construed, our confident moral judgments are once again data the theory has to fit—but now with a new twist, as they are observational data, automatically true.

This view can perhaps be worked out coherently, but it relies on an extravagant metaphysics and epistemology. Worse, it begs


all the really interesting questions about morality. Here I want
only to note Bentham’s decisive attack on such theories. I refer to
the Introduction’s scathing footnote on the British moralists, the
one Mill quotes as indicative of “the strengths and weaknesses of
his mode of philosophizing.” Reporting a moral intuition is
not the same as giving a reason. Imagine—the case is not coun­
terfactual—two people disagreeing on a basic moral or political
question, each sure of being right. In the intuitionist view, we
are at a dead end. One or the other is not perceiving the realm
of moral truths correctly, but we have no way of checking. Ipse
dixitism, as Bentham would say, with a vengeance. If we are to
have intelligent debate, we will have to find reasons. It would be
embarrassing, even for a full-blooded intuitionist, to concede
that there is no point to morality.

The second view is that implicitly adopted by Aristotle in the
Nicomachean Ethics. Justification need not commence with a slate
wiped clean of all our views. Instead, we can investigate our
view critically, revising them as we go. Nor are our critical stan­
dards magically yanked from some hitherto unknown realm. We
already have such standards, themselves subject to revision as
justification proceeds. Since our views do not form some per­
fectly coherent whole, we have working room to launch forth on
the process of justification. I want to hold this view of justifica­
tion in abeyance here, since Bentham and Mill offer another
view of justification, one promising a much more straightfor­
ward account. Not holding a view like Aristotle’s, they can af­
ford to sneeze at our shared moral judgments.

I do not mean to cast these two views as rivals. In the Ethics,
for example, Aristotle actually draws on both; Sidgwick, as a fol­
lower of Aristotle, does so too. He appeals, as we will see, to a
“fundamental moral intuition” in the argument from egoistic he­
donism to utilitarianism. And he writes, “The present argument

\[122\]
AGAINST UTILITARIANISM

does not aim at proving an exact coincidence between Utilitarian inferences and the intuitions of Common Sense, but rather seeks to represent the latter as inchoately and imperfectly Utilitarian. So one might well challenge Sidgwick by demanding a viable utilitarian rendition of fairness or justice, but that ground is surely trod well and frequently enough, and I will leave it unexplored.

The appropriate scrutiny, then, is not an appeal to our shared moral judgments. Instead, I want to consider the utilitarians' claims to have provided a clear, all-encompassing theory, and to have given us good reasons for adopting it. Neither of these claims, I argue, can be sustained.

Utilitarianism as Incomplete

Consider:

1. You are floating across the Lake of Lucerne in a canoe on a balmy spring day, admiring the wispy clouds and piping birdsongs.
2. You are about to devour, courtesy of a friend, a triple-fudge cake from Rosie's, your favorite bakery, after a day without food.
3. You are one of a swirling mob exulting over the accession of Charles the Fat to the throne.
4. You finish solving a difficult crossword puzzle.
5. You have an orgasm.
6. You are reinstated with your best friend of thirty years, thought lost in an airplane crash.

These, let us imagine, are all pleasurable.

Now consider:

1. You are caught in a thunderstorm on top of Half Dome and fear death at the hands of wind and lightning.
2. Your promised triple-fudge cake fails to materialize; your friend is playing a practical joke and solemnly brings out an empty plate.


*See Methods, pp. 439–448, for as good a rendition as anyone has offered.
3. You are found hiding during a pogrom.
4. Your first concert performance is poorly reviewed.
5. You break your arm; it is a compound fracture, and your friend swoons at the sight.

These, let us imagine, are all painful. I trust that I need not disclaim any attempt to show there are six kinds of pleasures and six kinds of pains.

As Sidgwick notes, if we are to sum pleasures and pains, they must be commensurable, cardinal, and interpersonally comparable. Here I will concede cardinality and interpersonal comparability. I want to contest commensurability. We are inclined to think of these twelve pleasures and pains as very different. For commensurability to obtain, there must be some metric, a single scale, along which they can all be placed.

The candidate for the metric, in a utilitarian view, will be agreeability of consciousness. Each experience has a raw feel to it, and the feel has an agreeability or disagreeability. Anticipating the chocolate cake will be less agreeable than the orgasm, say; the broken arm more disagreeable than the fear of death on Half Dome. But does the agreeability refer to any aspect of the experiences, or is it rather a disposition on our part to choose one over another? A full-blown theory of pleasure and pain—not a scientific-minded revision, but merely an account of ordinary usage—will have to move well beyond the supposed feel of the experiences. It will have to draw in, for example, various sorts of behavior. Whatever mentalistic strands the concepts now have do not seem to exhaust their meaning. Yet for the theory to be about happiness construed as a mental state, the metric must not collapse into something like preference satisfaction.

"Well," one might say, "we call them all pleasures and pains, so they must have something in common." The common element might be just our disposition to choose, not an element of the pleasure or pain itself. Yet there need be no common element

"For a recent suggestive discussion of the latter, see Alfred F. Mackay, "Interpersonal Comparisons," *Journal of Philosophy* 72 (October 2, 1975): 535–549."
AGAINST UTILITARIANISM

at all. We could have a case of family resemblance. As Mill
notes, "Names creep on from subject to subject, until all traces
of a common meaning sometimes disappear, and the word comes
to denote a number of things not only independently of any com-
mon attribute, but which have actually no attribute in com-
mon." Among all the experiences we call pleasures we find no
peculiar tickle, itch, or tingling in common; among the pains, no
twinge, burning, or stinging. It is possible that utilitarians are
built differently from the rest of us and lead linear mental lives,
but that seems doubtful.

That mental states are incommensurable presents more than a
technical obstacle for utilitarianism. It is just one part of a gen-
eral problem: the utilitarian calculus demands a staggering loss of
information. Just as utilitarians ask us to forget the differences
between orgasms and solving crossword puzzles, they ask us to
overlook distributive considerations, risk and uncertainty, and
time preference. By focusing now on these three, I hope to show
that utilitarians' claims to have provided a complete decision pro-
cedure must be rejected.

Suppose an agent has a choice between what Bentham would
call a pure pleasure of 6 utils and an impure pleasure of 9 utils
with −3 utils (3 units of pain) attached. One would expect a
theory of rational choice to describe an agent contemplating two
ordered pairs: (6, 0) and (9, 3). Instead we are told, in Sidgwick's
words, that we are to strive for "the greatest possible surplus of
pleasure over pain, the pain being conceived as balanced against
an equal amount of pleasure, so that the two contrasted amounts
annihilate each other for purposes of ethical calculation." The
agent, then, must be indifferent between (6, 0) and (9, 3)—indif-
derent, for that matter, among all pairs generated by (6 + x, x)
for all x > 0. Such an agent seems willfully blind or crazy.
Sidgwick offers no independent justification for focusing on the
surplus. The single number is a technical requirement of the the-
ory. Without it, we would face problems of balance: Mightn't we
sensibly prefer a pure pleasure of 10 utils to one of 15 utils mixed
with 3 utils of pain?

"Logic, p. 38; also the superb discussion in Logic, pp. 668–697.
Annihilating equal units of pleasure and pain seems worrisome even for an individual pursuing her own happiness. It seems downright perverse for society, and sharpens the nature of the utilitarian's disregard for distributive considerations. Sidgwick notes the possibility that different actions will produce the same greatest happiness score (not merely a formal possibility in his view, since his calculus is rough). In such cases, he tells us, we need "some principle of Just or Right distribution of the happiness," and he suggests "that of pure equality—as given in Bentham's formula, 'everybody to count for one, and nobody for more than one.'" Equality, though, is hardly a complete and clear principle for ranking two or more distributions. Are we to choose that with the lowest standard deviation? with the fewest severely outlying points? or what?

Suppose we choose a social distribution summing to \((200,000, 100,000)\). The surplus is 100,000; if other possible distributions have lower surpluses, we are not supposed to consider them. Yet in this society there are 100,000 units of pain spread around, and we want to know where they are. It is implausible to suggest that for a given individual a given quantity of happiness annihilates an equal quantity of pain. It is extraordinary to suggest that when I am happy and you are unhappy, a given quantity of my happiness annihilates an equal quantity of your unhappiness. Again, we are given no reason to think it should. The simplification is introduced to make sense of the idea of greatest happiness. This technical necessity is why "utilitarianism does not take seriously the distinction between persons." 

Were society a colossal organism, we could make sense of focusing on the surplus, on the single number. Each of us would be something like cells in a sponge. But it is hard to see how society literally could be an organism, or to find an author willing to assert explicitly that it is. Perhaps a sponge cell, if conscious,

---

45 Hard, but alas! not impossible: Herbert Read, "The Philosophy of Anarchism," in his *Anarchy and Order* (Boston: Beacon Press, 1971), writes, "The right kind of society is an organic being—not merely analogous to an organic being, but actually a living structure with appetites and digestions, instincts and passions, intelligence and reason" (p. 50).
would (properly?) care only for the welfare of the entire sponge, but surely that is not the individual's relationship to society. Strikingly, Bentham and Mill both insist, in their very different ways, on criticizing organic conceptions of society. Bentham detects a dread fictitious entity, and pronounces, "The community is a fictitious body, composed of the individual persons who are considered as constituting as it were its members. The interest of the community then is, what?—the sum of the interests of the several members who compose it." Mill, in the midst of an extremely careful discussion of the logic of the moral sciences, bluntly states, "Human beings in society have no properties but those which are derived from, and may be resolved into, the laws of the nature of individual man." This methodological individualism undercuts a descriptive prop that would help make some sense of the greatest-happiness principle.

Only some sense, though, because there are still more puzzles. I have been writing so far as though we know what the outcomes of our actions are, but of course we don't. Because of our ignorance of the workings of society, we have only an idea of the possible outcomes; and because the causal ripples introduced by one action interact with those introduced by other actions, unintended consequences seem a permanent feature of any remotely complex society. Suppose, in keeping with a rational-choice model, we can formalize for each action we are contemplating the utilities and associated probabilities of a set of possible outcomes. (This is no trivial supposition, even if we take a Bayesian approach to estimating probability.) How ought we to choose among sets?

Economists will be quick to suggest that we maximize expected utility. That is, we collapse each set into a utility score $S$:

$$S = \sum_{i=1}^{n} p_i U_i,$$

"Introduction, p. 12."

"Logic, p. 879. Though it is not clear just what Mill means to deny. He may well believe in emergent properties for this case, but think that those properties must be accounted for by psychology and ethology. If so, I am not sure he would be denying here anything anyone might wish to affirm."
where $p_i$ is the probability of the $i^{th}$ outcome and $U_i$ the utility of that outcome. Recall, however, that we are now considering utility as some agreeable mental state. Maximizing the expected value of this kind of utility is anything but rational; indeed it is absurd. It requires a rational agent to be indifferent between a one-eighth chance of 800 utils and a one-half chance of 200; or, more generally, to be indifferent among all the members of the set defined by $(1/n, kn)$ for all $n > 1$, with $k > 0$ or $k < 0$. This procedure eliminates all reference to risk aversion, and so is at once eminently attractive and decidedly wrongheaded. Attractive, because otherwise utilitarianism again lapses into indeterminacy or arbitrariness. Just how risk averse is it rational to be? (Can levels of risk aversion even be judged as more or less rational?) Wrongheaded, because it again deliberately purges information, leaving us with a travesty of rationality in the interests of meeting the demands of the theory.

Next I wish to consider pure time preference, cases where an agent prefers a sooner good to an equal later good just because it is sooner. Utilitarians often (conveniently) ignore the dimension of time and write as though the happiness produced by an action occurs in an instant. Neglect of time, however, is hardly appropriate to those who insist that we calculate on the basis of consequences, for the consequences of an action stretch through time.

Consider the graphs of Figure 1, where the x-axis is time, the y-axis total utils in society, and the origin the neutral point between pain and pleasure. The two curves represent the utilitarian consequences of two different possible actions. Suppose that the total area under each curve is the same from time 0 to time 10. That is, over the whole time period, each action will give rise to the same amount of pleasure. Which one should we choose? Or is it irrelevant?

Bentham is sometimes fonder of pleasures that occur sooner rather than later. As he writes in the memorable (if obscure) poem in the Introduction:

\textit{Intense, long, certain, speedy, fruitful, pure—}
\textit{Such marks in pleasures and in pains endure.}\footnote{Introduction, p. 38n. The poem continues with a clear statement of the dual standard.}
Since speedy occurs in the same list as certain, we have here a ringing endorsement of pure time preference. In this view, we should choose the first curve—even if the total area under it is less than the total area under the second curve—up to a point depending on the extent of our preference for nearer pleasures. Here we encounter still another source of indeterminacy or arbitrariness. If time preference is to be countenanced, how much time preference? Has rationality anything to say on this score?

In Sidgwick’s view, pure time preference is irrational. This view is convenient for utilitarianism, since it avoids the dilemma of choosing a defensible rate of time preference. Unlike the business of pleasures annihilating pains, however, this view is argued for on the merits: “Mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment than to that of another.” So put, this is but a challenge: why should earlier pleasures be preferred to later ones of equal magnitude? I am not sure that this challenge is unanswerable, but I will grant the point. Time preference, then, is to be discarded.

Regardless, we can exploit the two curves to underline once

---

“Methods,” p. 381; note too p. 124 n. See also Bentham, Deontology I:130, 156, 160; II:82–83.

“Works I:206; Deontology I:62.”
again the curiosity of describing a possible world with a single number. Suppose first that the time units are days. We face a choice between ten days of unbroken serenity and a single tempestuous roller-coaster ride. We may decide it does not matter which we choose. But surely a preference (one way or the other) would be legitimate, for the two alternatives are different. That they leave us at day 10 with the same total amount of happiness is the scantiest description of them. Why should we think that this description captures their only relevant feature for moral or rational choice?

The curves also illustrate the odd conception of society built into the utilitarian calculus. Suppose now that the time units are centuries. We may squander scarce resources and enjoy a cornucopia of consumer delights, falling suddenly into a crevasse of scarcity; yet this limits-to-growth scenario is not all doom, for we will eventually wiggle our way out. We may instead husband scarce resources and keep the welfare of future generations firmly in mind. There are about four centuries of misery in the middle of the first graph, but “society,” stretching over a thousand years, is equally happy in the two graphs. It looks as though the utilitarian is committed to the sponge-cell view with a vengeance. That utilitarianism does not take seriously the distinction between persons is a modest way of putting the point. For—imagine now a tiny bit more surface area under the first curve—utilitarians will, in the name of morality, loftily dismiss the life histories of complete generations. (Probably too they will congratulate themselves on transcending the squeamishness of ordinary morality in doing so.)

Handling these points one at a time has cast the difficulties that utilitarianism faces in too charitable a light. A utilitarian choice among alternatives, whether they be rules, acts, character traits, or anything else, will ideally shape up like this: Associated with each alternative will be a set of possible outcomes with probabilities. Choosing between such alternatives will not be easy, so utilitarians will drastically simplify. They will first re-describe each outcome simply as a distribution of pains and pleasures across time and persons. Then they will place all pleasures and pains on a single cardinal axis, however doubtful the avail-
ability of such an axis; they will subtract pains from pleasures, and pursue only surpluses of pleasure over pain; and they will pour the mental contents of all persons, each a little receptacle of pain and pleasure, into one great bucket called society. Then, I imagine, they will decide that they are interested in maximizing expected utility, despite the objections to that view, and they will further decide that they have no pure time preference. They will therefore collapse each possible outcome to a single number, multiply each number by the outcome’s associated probability, add the numbers in each set, and choose the alternative whose set yields the highest utility score.

I have tried to take Sidgwick’s procedure of seeing what is entailed by the notion of maximizing happiness a few steps farther. This procedure, or one much like it, seems to be what is entailed. (Utilitarians who refuse any of the simplifying moves will face problems of indeterminacy or arbitrariness, or will have to supplement their utilitarianism. Worse yet, they will face trade-offs: how much sooner happiness for how much less pain?) It is, we are told, a model, or the model, of moral decision making. Much of the literature on utilitarianism is reminiscent of the Emperor’s New Clothes, and cries out for the same treatment the little boy offers. As far as I know, no one has ever attempted even a sketch of such a decision for any actual problem. (Cost-benefit analysis doesn’t even begin to cash out the alternatives into distributions of pleasures and pains.) For good reason, too: we would need to know unfathomably more about consequences and individuals than we do or can know. Attempting even a sketch would discredit the entire project.

Attempting a sketch would also underline a point readily underplayed in abstract discussions of utilitarianism. The sorts of problems I have been discussing are not mere technical obstacles, issues to be handled within the sterile realm of the theory of rational choice. Instead they are fraught with moral and political import. Being forced to consider how to weigh the interests of future generations will put a different face on time preference and uncertainty. The parameters surrounding maximization, telling us just what and how to count, will be morally and politically crucial. Whether we maximize average or total happiness
will make all the difference for decisions involving population shifts. Whether we count the happiness of animals will speak volumes on questions about vegetarianism, hunting, vivisection, and the rest. All these issues are morally and politically interesting. None of them could possibly be decided by a utilitarian calculus, since that calculus cannot get started until these issues are resolved. So despite the assurances of Bentham, Mill, and Sidgwick, utilitarianism is not and cannot be the sole principle of morals and politics. Performing its complex calculations might exhaust us, but could not exhaust our moral and political concerns.

Yet even as utilitarianism demands fabulously complicated and detailed calculations, it is singularly narrow-minded in its attention to consequences. Only (un)desirable states of consciousness count. A utilitarian cares not the least for economic or political structure, for culture or society, for themselves. All that is interesting about them is how much happiness they yield. What reason are we given to adopt such a standard?

Justifications of Utilitarianism

Bentham flatly denies the possibility of proving utilitarianism correct: "Is it susceptible of any direct proof? It should seem not: for that which is used to prove everything else, cannot itself be proved: a chain of proofs must have their commencement somewhere."51 Still, he argues for it, offering a threefold classification of moral principles. There is the principle of utility, those "constantly opposed to it," and those "sometimes opposed to it, and sometimes not, as it may happen." Constantly opposed is the principle of asceticism, which always approves of unhappiness and disapproves of happiness. Sometimes opposed is the principle of sympathy and antipathy, "that principle which approves or disapproves of certain actions . . . merely because a man finds himself disposed to approve or disapprove of them: holding up that approbation or disapprobation as a sufficient rea-

51Introduction, p. 13; also Deontology 1:276–278.
son for itself, and disclaiming the necessity of looking out for any extrinsic ground."'

Bentham is little interested in discussing the principle of asceticism. It "seems originally to have been the reverie of certain hasty speculators" who wished to guard against alluring pleasures fraught with "pains more than equivalent to them" and went on to forget their own devotion to utility. Bentham reserves his concern (as shall I) for principles sometimes opposed to utility and so scrutinizes the principle of sympathy and antipathy. It is, he decides, no principle at all, for, "disclaiming the necessity of looking out for any extrinsic ground," it provides no standard. Since Bentham thinks that "the various systems that have been formed concerning the standard of right and wrong, may all be reduced to the principle of sympathy and antipathy," he makes short work of enshrining the principle of utility.

Bentham is right in holding that an acceptable moral or political principle must direct our attention to reasons, not to our own sentiments, cloaked in "sonorous and unmeaning words" though they be. That we feel something to be wrong is not enough for justification. But it is illegitimate for Bentham to take the principle of sympathy and antipathy as the example of principles sometimes opposed to utility. He makes it seem as though such principles are irremediably arbitrary, as they would be if pleasure and pain were all that matters.

Bentham observes that "the utilitarian scale vibrates only between good and evil—pain and pleasure—other elements count for nothing in the balance, let them be called by names as pompous as they may." Other scales may not be quite so resolutely unresponsive to other considerations. They may vibrate between freedom and unfreedom, equality and inequality, autonomy and authority, without regard for the felicific consequences. From the point of view of utility, they will oscillate wildly. But provided their standards are suitably well defined and identify con-

"Introduction, pp. 17, 25.
"Introduction, p. 21.
"Introduction, p. 25.
"Deontology II:11.
"Deontology II:25.
siderations that vary from utility, they will provide examples of nonarbitrary principles sometimes opposed to utility. Has Bentham anything to say against such principles?

Bentham frequently denies that the moral concepts can be made sense of on nonutilitarian grounds: "Take away pleasures and pains, not only happiness, but justice, and duty, and obligation, and virtue—all of which have been so elaborately held up to view as independent of them—are so many empty sounds." This denial surely requires support. Bentham may have such support in mind in referring to the "cloud of misty obscurity [that] has gathered round the term" obligation. That view we might speculatively reconstruct. Moral talk is endlessly circular. We ask for an account of rights, and are sent to obligations and permissions, for accounts of those in turn to duties, on to right, wrong, good, bad, legitimate, illegitimate, just, unjust, and so on. We are entrapped, with no escape in sight. Not only is this procedure unenlightening, but it fails to provide a satisfactory standard. Only a purely descriptive concept can provide the bottom line we seek, else moral talk must remain meaningless noise—or so, I suspect, Bentham thinks. One way to meet the argument would be to show that circles big enough to include far-flung elements can be virtuous, not vicious. But a less ambitious approach will do the trick here. Suppose we give a satisfactory descriptive account of freedom and equality and offer a moral principle focusing on them. Bentham would demand of one offering such a principle, "admitting (what is not true) that the word ought can have a meaning without reference to utility, let him say whether there is any such thing as a motive that a man

---

57Works 1:206; see too 211, 248; IV:542–543; VIII:290; Deontology I:10, 31–32, 136–137. Bentham's famed fusillade against natural rights is at Works II:501. None of this talk commits Bentham to defining the moral concepts in terms of pleasure and pain, in which case he would be open to strictures about making his first principle a tautology. His point is rather that the appropriate criteria for using these concepts are pleasure and pain. Though he sometimes writes carelessly—"it is but tautology to say, that the more consistently [utility] is pursued, the better it must ever be for human-kind" (Introduction, p. 21)—surely does invite such strictures—Bentham himself distinguishes is and ought, so it would be unlikely for him to slip into the fallacy. Note Works I:189, 229.

58Deontology I:10.
AGAINST UTILITARIANISM

can have to pursue the dictates of it”; for if not, the principle looks pointless. Here Bentham explicitly appeals to his hedonistic psychology. I will assume that that psychology is deeply flawed, that we can and do desire things besides pleasure for themselves. Other standards need not be anarchical or horribly vague. It is all a matter of comparison. Attempts to perform a utilitarian judgment would themselves be anarchical and vague. So potentially explosive Benthamite skepticism about nonutilitarian moral talk can be defused.

Like Bentham, Mill cautions the reader that “questions of ultimate ends are not amenable to direct proof.” Still, he thinks, “Considerations may be presented capable of determining the intellect either to give or withhold its assent to this doctrine; and this is equivalent to proof.” This infamous proof is allegedly encapsulated in one paragraph:

The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it; and so of the other sources of our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it. If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so. No reason can be given why the general happiness is desirable, except that each person, so far as he believes it to be attainable, desires his own happiness. This, however, being a fact, we have not only all the proof which the case admits of, but all which it is possible to require, that happiness is a good: that each person's happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons.

Here, it seems, is some blatantly fallacious reasoning, to which we can dutifully rehearse the stock remonstrances. Desirable means “ought to be desired,” whereas visible and audible mean “can

59Introduction, p. 16.
61Utilitarianism, pp. 207, 208, 234.
be seen and heard,” so the analogy breaks down. And only sleight of hand makes the greatest happiness a good to all: “the aggregate of all persons” may not be an entity with a good, and even if it is, Mill gives no reason for any individual to care for others’ happiness. No wonder, then, that Bradley professes himself “ashamed to have to examine such reasoning,” that Moore finds a “fallacy . . . so obvious that it is quite wonderful how Mill failed to see it.”

If Mill were attempting a deductive proof, Bradley’s and Moore’s derision would be apt enough. But he is not; nor, given his views in the Logic, could he be. Though the argument is still faulty, we should try to arrive at a satisfactory understanding of it.

Mill distinguishes in the Logic between art and science. The former, he says, concerns what ought to be, the latter what is. Like Bentham, he holds that “a proposition of which the predicate is expressed by the words ought or should be, is generically different from one which is expressed by is, or will be.” In conceding that if we take a syllogism as an argument, we must grant that it begs the question, Mill commits himself to the view that purely deductive arguments add nothing to what is already contained in the premises. So Mill believes that it is impossible to derive an ought from an is. The conventional view of his proof of utilitarianism, taking him to be attempting just that, is presumably misconceived. We may be driven to hold that Mill simply forgot his own strictures from the Logic when he came to write Utilitarianism, or that he had changed his mind but some-

---


64 Logic, p. 949.

65 Logic, p. 184.
AGAINST UTILITARIANISM

how never noted the change in his later revisions of the Logic. But surely we should first try to find a reading that makes the two works cohere.

We need, then, a view of the proof different from that excoriated by Bradley and Moore. Just what does Mill take to be “considerations . . . capable of determining the intellect?” Look again at part of the proof: “If the end which the utilitarian doctrine proposes to itself were not, in theory and practice, acknowledged to be an end, nothing could ever convince any person that it was so.”66 The point is pragmatic. There is no sense, Mill is suggesting, in nominating moral principles that tell us to pursue ends we take no interest in. We do take an interest in happiness, so “happiness has made out its title as one of the ends of conduct, and consequently one of the criteria of morality.”67 Mill goes on to argue that while we do desire other things—money, virtue, music, health—for themselves, we desire them as parts of happiness: “Happiness is not an abstract idea, but a concrete whole, and these are some of its parts”; so “there is in reality nothing desired except happiness.”68

The argument for our desiring only happiness is the key to the proof, granted Mill’s pragmatic strategy. Consider: “We have now, then, an answer to the question, of what sort of proof the principle of utility is susceptible. If the opinion which I have now stated is psychologically true—if human nature is so constituted as to desire nothing which is not either a part of happiness or a means of happiness, we can have no other proof, and we require no other, that these are the only things desirable.”69 Mill’s view is that happiness is all we desire, so there is a point to hav-

66 It is surprising that Moore takes Mill to task for “as naive and artless a use of the naturalistic fallacy as anybody could desire” (Principia Ethica, p. 66), since Mill does not say, as Moore claims, that “‘Good’ . . . means ‘desirable’” (p. 66). Mill is explicit here that the connection is one of evidence. Elsewhere, though, like Bentham, he flirts with the language of tautology: “What is the principle of utility, if it be not that ‘happiness’ and ‘desirable’ are synonymous terms?” (Utilitarianism, p. 258n.).
67 Utilitarianism, p. 234.
68 Utilitarianism, pp. 236, 237.
69 Utilitarianism, p. 237.

[137]
WITHOUT FOUNDATIONS

ing the utilitarian view. Other views may be proposed, but we can return Bentham's challenge: what motive will we have to adopt them?

Mill's psychology is a good deal more complicated than Bentham's and cannot be dismissed as readily. How can we desire things for themselves and as parts of happiness? Moore decides that Mill is talking "contemptible nonsense" and gaily ridicules the idea that jingling coins are a part of happiness. Should we understand happiness as some agreeable mental state, the scorn is deserved. And that is Mill's official view of happiness in his opening account. Moore's reading is hardly fair, however. Mill's conception of happiness shifts emphatically away from a mental state and toward a valuable state of affairs, from a Benthamite to an Aristotelian view. As long as Mill adheres to this latter view, his theory is not utilitarian at all in the sense I have adopted here. Nor is my distinction pointless. It allows us to mark the important differences between a theory telling us to maximize agreeable consciousness and one telling us to maximize valuable states of affairs. (Though talk of maximization will become increasingly mysterious: what could the relevant metric be?) We may concede Mill the psychology he argues for and turn it against utilitarianism. We do take interest in things besides pleasures and pains. So a perfectly reasonable moral theory may direct our attention to those other things.

The argument, then, collapses in broadening happiness. A question remains: How is it supposed to be an argument for utilitarianism and not, say, ethical egoism? How does Mill propose to show that each person ought to attend to the general happiness, and not her own? Mill is moving from (1) my happiness being a good for me, yours for you, and so on, to (2) the general happiness being a good for all, to (3) the general happiness being a good. Since happiness, Mill thinks, is the only good, we may substitute (3') the general happiness being the good. Having reached (3) or (3'), Mill can explain why each person ought to care for the general happiness, even if it makes her miserable: it is simply the good. Here the traditional criticism is more plausi-

70Moore, Principia Ethica, pp. 71–72.
AGAINST UTILITARIANISM

[139]
evident because they are substantially tautological.” The principles he seeks are to be self-evident yet not tautological. Nor is he satisfied with formal constraints such as universalization, “that whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances.” Such constraints, he believes, are “too abstract” to yield any definite account of what we ought to do. They do though offer an insight that may be exploited to other ends. If similar individuals, Sidgwick suggests, make up a “Logical Whole or Genus,” there is no good reason to prefer one to another. Therefore, he argues, in a passage worth quoting at length:

By considering the relations of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realised in the one case than in the other. And it is evident to me that as a rational agent I am bound to aim at good generally,—so far as it is attainable by my efforts,—not merely at a particular part of it.

From these two rational intuitions we may deduce, as a necessary inference, the maxim of Benevolence in an abstract form: viz. that each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.

We already have the formal structure of utilitarianism. There remains only the task of identifying the good with happiness, a task Sidgwick proposes to accomplish “by a more indirect mode of reasoning” than Mill’s. 75

What might the good be? Sidgwick attempts now to eliminate other candidates. It cannot be virtue, since an adequate account of virtue will inevitably mention the good and “involve us in a logical circle” we cannot escape. Nor can it be “the determina-

AGAINST UTILITARIANISM

tion of the will to do whatever is judged to be right and to aim at realising whatever is judged to be best,” lest we embrace “a palpable and violent paradox”: that the good lies in pursuing something that does not itself exist objectively. Like virtue, “talents, gifts, and graces” will not withstand scrutiny: “Reflection shows that they are only valuable on account of the good or desirable conscious life in which they are or will be actualised, or which will be somehow promoted by their exercise,” Physical processes of life cannot be ultimately desirable: “So long as we confine our attention to their corporeal aspect,—regarding them merely as complex movements of certain particles of organised matter—it seems impossible to attribute to these movements, considered in themselves, either goodness or badness.” Nor is it “all life regarded on its psychical side which we can judge to be ultimately desirable,” since there is still pain; instead, it is “Desirable Consciousness.” Now we face the initial problem in new disguise. What is to count as desirable consciousness? A crowd of contenders—“cognition of Truth, contemplation of Beauty, Free or Virtuous action, as in some measure preferable alternatives to Pleasure or Happiness”—come bursting in, but Sidgwick bravely fends them off, writing, “I think, however, that this view ought not to commend itself to the sober judgment of reflective persons.” Our considered judgment must be that all these things are valuable only because of their “conduciveness, one way or the other, to the happiness of sentient beings.” This conclusion coheres well enough, Sidgwick thinks, with the “ordinary judgments of mankind,” carefully considered.76 So goes the argument for utilitarianism.

There is much to criticize here. I will focus on two points: self-evidence and wondering what the good is.

The business of hammering out self-evident axioms to serve as foundations is eminently attractive for quite a few reasons. It exercises just those skills that philosophically inclined writers concerned with morals and politics possess in abundance. It promises to provide a fabulously solid theory, one maybe even immune to attack. It can be done without any attention to con-

76Methods, pp. 394–402.
Without Foundations

tingent facts of society, history, or politics, since it is not at all clear what the seeker of axioms might gain by being steeped in such knowledge of the world. It promises to sail gracefully past more cumbersome approaches which drag in far-flung considerations and wearily force them into less elegant theoretical structures.

The only sticky point is finding the axioms. How do we know when we have stumbled upon a self-evident proposition? Sidgwick offers a four-part test: “The terms of the proposition must be clear and precise”; “the self-evidence of the proposition must be ascertained by careful reflection”; “the propositions accepted as self-evident must be mutually consistent”; and (here he is a bit elusive) we should not find considered disagreement. But the notion of self-evidence, reappearing in the second part of the test, needs a good deal of elucidation. A prosaic attempt would be, “carries its own evidence along with it.” That gloss, however, leads straightaway to tautologies, in which Sidgwick is decidedly uninterested. Nor will a purely psychological attempt—say, “seems undeniable”—do, for as Sidgwick notes, “Any strong sentiment, however purely subjective, is apt to transform itself into the semblance of an intuition.” We are supposed somehow to scrutinize our beliefs critically from some neutral vantage point, “the point of view of the Universe,” Sidgwick would say, and see if they hold up. But just what is that point of view, and how can we take it up? Considering as coolly as I can, I cannot bring myself to believe that Sidgwick’s second axiom or intuition (that I am bound to aim at the good generally, sacrificing my own good if need be) is true, let alone self-evident. Lurking here is, once again, the sponge-cell conception of the individual in society, a problematic descriptive view whose contingency smears the epistemic purity of the axiom. On such a conception the plausibility of the so-called self-evident principle rests. Again, the conception is utterly implausible. Sidgwick himself, in his discussion of the dualism of practical reason, concedes as much.

In his pursuit of the good, Sidgwick ascends to increasingly

\[Methods, \text{ pp. 338–342.}\]
\[Methods, \text{ p. 339. Compare Logic, pp. 224–261, especially 238–244.}\]
\[Methods, \text{ pp. 497–498.}\]
AGAINST UTILITARIANISM

rarefied considerations and abstract issues. There are two views we might take of such an approach, views I will crudely label Platonic and Wittgensteinian. In the Platonic view, it is only when we reach such questions that the discussion promises to yield genuine knowledge. More strongly put, knowledge of the basic issues must precede anything resembling knowledge in the circumstances of daily life: Lysis cannot really know who his friends are until he knows what a friend is. In the Wittgensteinian view, the relevant concepts can be used meaningfully only within the more or less concrete contexts that contain the criteria governing their use. When we rip the concepts out of context to examine them singly, they have lost meaning, and our inquiry is doomed to failure. A question such as “What is the good?” may then provoke two responses. We may think, “Aha! now we are finally at the heart of the matter; now we can really understand morality.” Or we may think, glumly instead of exultantly, “What follows is bound to be nonsensical.”

I want to urge the merits of the Wittgensteinian view, or of one much like it. Sidgwick’s query, whether it is the physical or mental side of life in which ultimate good is to be found, is baffling. If physical life can be regarded “merely as complex movements of certain particles of organised matter,” mental life, I suppose, can be cast as “merely the fluctuations in timbre of experience.” So cast, it looks rather uninteresting. Here we can borrow two pages from Hume: “The life of a man is of no greater importance to the universe than that of an oyster”; the point of view of the universe, even if it were available, would dwarf and indeed obliterate all human concerns, interests, and significance. Besides, even contemplating this point of view is disorienting. “We are got into fairy land,” as Hume says in a different context;\(^8\) in a realm so foreign to the ones we are familiar with, our responses become erratic, even zany. We can get on perfectly well in discussing good actions, good people, good knives, and all the rest without ever asking what the good is.

We can still learn much of interest by asking what good means.


[143]
But Sidgwick’s question, “What is the good?” is not a request for a definition or explication of the concept. Nor is it a request (more like the actual Platonic ones) for a theory showing what goodness is, how it arises, how we know it, and so on. Instead Sidgwick is asking, as Moore might say, “What things are good?” — but he will accept only an answer showing that one thing is ultimately good, all other goods being instrumental to achieving that good. This requirement seems excessively harsh. Perhaps there is such an ultimate good, but I see no reason whatever to assume that there must be.

Bentham, Mill, and Sidgwick, I conclude, give us no reason to accept utilitarianism. And should the arguments of the last part have any merit, utilitarianism is much too exotic a view to command our allegiance without good reasons.

Contemporary Utilitarianism

Some readers will think that most if not all of my criticisms so far apply only to classical utilitarianism. They will censure me for beating a dead horse, and will urge the merits of contemporary forms of utilitarianism, some of which reject a hedonistic conception of utility. But I think such forms are no better than their classical ancestors. Accordingly, here I examine the theories of three contemporary utilitarians: John Harsanyi, R. M. Hare, and R. B. Brandt.

Harsanyi invites us to consider the common idea that “the moral point of view is essentially the point of view of a sympathetic but impartial observer.” He suggests that we may formulate this point of view in the notation of game theory. Suppose that we denote different (hypothetical) societies, or the same society under different proposed social policies or rules, as $A$, $B$, $C$, and so on. Now, any individual $i$’s objective position in, say, $A$ “can be regarded as a vector listing the economic, social, biological, and other variables” describing his position, and can be denoted $A_i$. Let $U_i$ denote a von Neumann–Morgenstern (from

\[\text{[144]}\]
AGAINST UTILITARIANISM

now on, vNM) utility function for individual $i$.

We may then introduce the social-welfare function. For individual $i$ to be impartial, in the relevant moral sense, in evaluating social situation $A$, he must pretend uncertainty as to what his position $A_i$ in $A$ would be; accordingly, Harsanyi proposes that the agent assign the same probability, $1/n$, to the chance of coming out in any one of $n$ positions. Thus he says, for some individual $i$:

Any given social situation would yield him the expected utility

$$W_i(A) = \frac{1}{n} \sum_{j=1}^{n} U_j(A)$$

because he would have the same $1/n$ chance of being put in the place of each individual $j$ ($j = 1, \ldots, i, \ldots, n$) and therefore of obtaining the utility amount $U_j(A)$, representing individual $j$'s utility level in situation $A$. In other words, in making moral value judgments individual $i$ would evaluate each social situation $A$ in terms of the average utility level that the $n$ individual members of society would enjoy in this situation.

To choose the morally best situation, one simply chooses the one with the highest expected utility.

The formal similarity between this model and classical utilitarianism should be clear enough. But the introduction of vNM utility functions to replace hedonistically conceived utility marks a striking departure, one that needs emphasis. To review such utility functions: If we make some minimal assumptions about the consistency and sensitivity of a rational agent's preferences, we can find a set of real cardinal numbers, namely her vNM utility function, such that the agent acts as though she were trying

---


84Rational Behavior, p. 50.
to maximize the value of that function. Furthermore, the utility of a lottery ticket of uncertain outcomes will equal the sum of the utility of each outcome multiplied by the probability of that outcome. Such a vivid reinterpretation of the concept of utility changes the whole complexion of utilitarianism.

Some of the objections commonly leveled against classical utilitarianism therefore fizzle as objections to forms of utilitarianism based on vNM utility. There is nothing problematic here about utility being cardinal or commensurable. Nor need we worry about how to handle problem cases of risk and uncertainty; the function effectively incorporates the agent's risk aversion and so gives us her own solution to them. Finally, the perpetual quest for that pure psychic stuff named utility that will serve as the end of all human action happily can be suspended; these utility functions report only preferences, with nary a word on why the agent prefers what she does or how she feels once she gets it. Accordingly, substituting them for the hedonistic conception of utility may seem an unambiguous great leap forward. But vNM utility functions usher in new problems even as they solve old ones.

As Harsanyi recognizes, the relevant sense of preference here is behavioristic: "We shall say that A is preferred (or is strictly preferred) to B by the decision maker if he always (i.e., with probability 1) chooses A rather than B whenever he has to choose between them." The utility an agent maximizes, in this view, has nothing whatever to do with mental states. (We could postulate that agents prefer A to B if and only if A yields more of some identifiable mental state. Any such postulate would be unacceptably strong, however; a theory that incorporated it would hardly describe us.) Indeed, an actor could maximize a vNM utility function without even having mental states. Take an automatic vacuum cleaner that propels itself around the room, surveying

---

"Rational Behavior, p. 27. Yet Harsanyi slips sometimes into thinking of utility as a psychological state, as when he refers to "the basic goal of all morally good actions, viz. to create as much happiness as possible in this world" in his "Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" Theory and Decision 6 (August 1975): 323, reprinted in his Essays on Ethics, Social Behavior, and Scientific Explanation (Dordrecht, Holland: D. Reidel, 1976), p. 76.
possible future routes with an electric eye and evaluating them with a minicomputer, gracefully sliding around furniture, rebounding off walls, and adjusting its brushes on different surfaces. We could construct a vNM utility function for such a vacuum cleaner. We might find, of course, that the function was impoverished, that it showed little variation for wide ranges of choices in the machine's environment. (It always goes straight if it sees nothing coming and has not hit anything within the last two seconds.) But that is irrelevant. The vacuum cleaner would act as though it were intent on maximizing its utility. Yet that it darts about the room maximizing utility hardly makes it a purposive agent.

Now, vNM utility functions have the feature that, if $U$ is an individual utility function, so is $V = aU + b$, for all $a > 0$. That is, "we are free to choose a zero point and a utility unit for $U$ in any way that we wish." For examining the individual decision maker it makes no difference, but now the problem of interpersonal utility comparison is posed in a new, stark form. What are we supposed to be comparing? Again, such utility functions measure quite precisely nothing at all. There have been arguments for interpersonal comparability, and indeed the job may be mathematically tractable. What matters, though, is keeping conceptually clear on what is going on.

As far as vNM utilities go, interpersonal utility comparison is a moral question, not a peculiarly slippery psychological one. There being nothing to measure, problems of ensuring accuracy can hardly arise. We are asking, "How much weight ought we to lend different individuals?"—which invites the ready answer, "Each to count for one, none for more than one." The champion of vNM utility can thus turn what looked like a liability into an additional asset of the theory. We can scale the utility functions so that everyone will assign his best possible outcome a rating of 1, the worst 0. This procedure seems egalitarian; it dispels worries about utility monsters, individuals who experience pleasure

---

Rational Behavior, p. 41.

and pain so intensely that in a felicific calculus they swallow up the rest of us. And the shift to preferences marks, arguably, a new concern in utilitarianism for individual autonomy. We need no longer worry about, say, forcibly implanting electrodes in people’s brains and inducing constant ecstasy.

Again, there may well seem to be unequivocal gains in shifting to the economist’s conception of utility. I mean to concentrate once more on the attempts given to justify utilitarianism, but I would like to note one serious difficulty that the shift to vNM utility creates. As long as an individual’s preferences are appropriately sensitive and consistent, we can construct the utility function. It does not matter whether those preferences are egoistic, altruistic, malicious, or psychopathological. By using the functions, we lose all ability to criticize the individual’s preferences. We lose too all ability to discuss the good life, for the question “What ought our preferences be?” must be rendered, in a view like Harsanyi’s, “What do we prefer our preferences to be?” The new preferences introduced to rank our existing preferences are themselves unimpeachable. Nor does a regress—“What do we prefer our preferences for our preferences to be?” and so on—seem at all illuminating. Note the wide range of concepts, from very different theories, that we cannot accommodate in a vNM view of utility: false consciousness, delusions, vicious dispositions, misunderstandings of one’s interests, worthless or meaningless life plans, and so on. If in a classical utilitarian view Brave New World offers a tempting vision of the good society because its people are happy, in a vNM utilitarian view that same society is tempting just because its individuals would not prefer other states. The loss of information here, the draining of all kinds of pregnant categories into the sink of preference, is appalling.

But perhaps I move too quickly. Harsanyi, after all, is willing to censor individual utility functions: “In our opinion individual i will be perfectly justified in disregarding j’s actual preferences

---


AGAINST UTILITARIANISM

in cases where the latter are based on clearly antisocial attitudes, e.g., on sheer hostility, malice, envy, and sadism. After all, the entire basis for i’s interest in satisfying j’s preferences is human sympathy. But human sympathy can hardly impose on i the obligation to respect j’s preferences in cases where the latter are in clear conflict with human sympathy.” Here Harsanyi conflates two very different senses of sympathy. The first is that of empathy, identifying completely with another. The second is that of approval, going along with another. Harsanyi uses the first sense in constructing the social-welfare function. Each of us is to sum $U_j(A)$, to compute j’s utility level in state $A$ (not, as we might expect, to compute how he feels about being j in $A$). Only in the second sense, though, does sympathy censor antisocial preferences; and the second sense will not begin to yield a utilitarian calculus.

Harsanyi, then, offers no satisfactory reasons for censoring utility functions. Nor does he offer any satisfactory reasons for employing them in the first place. He claims that his social-welfare function “can be obtained by a conceptual analysis of the nature of moral preferences (moral value judgments).” Being moral, the idea is, just means computing and observing these utility functions; if we want to be moral, we must do so. Presumably Harsanyi has this sort of necessity in mind when he claims, somewhat obliquely, “quasi-hypothetical objective validity” for the moral rules yielded by his theory.” Surely, though, the purported conceptual analysis is flawed. Utilitarianism, however conceived, is hardly the only putatively moral doctrine available; it is only one such doctrine.


I conclude that Harsanyi’s variant of utilitarianism is unsatisfactory. While the shift from pleasure to preference may seem promising, again we are given no good reason to adopt the utilitarian standard. Yet the appeal of Harsanyi’s strategy of justification—that of anchoring controversial claims on some formal analysis—is clear. Harsanyi’s analysis of morality will not do the job, but Hare has pursued a more promising line of attack. Hare wants to ground a moral theory in the logic of the moral concepts. He thinks that “once the form of morality is accepted in our thinking, it quite narrowly circumscribes the substance of the moral principles that we shall adopt.”

Indeed, Hare now argues that we will find ourselves constrained to be utilitarians. He has arrived at this position only gradually. In his first book, where the program of studying the logic of the moral concepts is already laid out, Hare notes in passing the possibility of a conflict between justice and utility. That book, however, is occupied chiefly with the moral concepts. In his second book, Hare tries to draw a moral code out of his theory of the moral concepts, and he arrives at a position near utilitarianism. I will take the liberty of paraphrasing the argument.

Harsanyi also offers two mathematical proofs of his theory, and he has complained that “some critics of my concept of an additive social welfare function have apparently failed to notice the fact that in order to refute my conclusions they would have to refute all three arguments” (Rational Behavior, p. 293n.3). The proofs are flawed not by the mathematics, but by the axioms labeling certain preferences moral preferences.

Harsanyi also appeals to conceptual analysis to resolve moral issues. First, on who is to be included in the social welfare function, he opines, “We would like to have an operationally meaningful analytical criterion that would help us to decide whether to include, e.g., higher animals, human idiots, unborn babies in their mothers’ wombs, more distant future generations” (Rational Behavior, p. 60). But it is a moral dilemma, again not soluble by conceptual analysis of “society.” And on the choice between maximizing mean or total utility, Harsanyi declares flatly that “in my view, the mean utility criterion gives incomparably superior results” (“Morality and the Theory of Rational Behavior,” p. 633n.)—pretty heady stuff from one who in the same article discards intuitionist doctrines as “crude forms of obscurantism in ethics” (p. 625).


AGAINST UTILITARIANISM

Moral judgments properly understood, in Hare's account, must be universalizable. Should I hold that some state of affairs is good, or some action wrong, I am logically committed to holding that all relevantly similar states of affairs are good, all relevantly similar actions wrong. So much is guaranteed by the moral concepts. (Other concepts work similarly. Should I hold that $X$ is striking, or large, or infuriating, or sticky, I must grant the same of all relevantly similar $Y$s.) Hare presses on further, demanding of us "a certain power of imagination and readiness to use it" in making our judgments. $B$ "must be prepared to give weight to $A$’s inclinations and interests as if they were his own."$^{96}$ That $B$’s desires are his own counts for nothing, Hare holds, from a moral point of view. They are simply another set of desires, to be taken no more (or less) seriously than anyone else’s. A large part of morality has to do with safeguarding people’s interests, and Hare flirts with the idea that prescribing universally for the satisfaction of everyone’s interests will mean maximizing satisfactions.$^{97}$

But a most important obstacle stands in the way of an argument from universalization to utilitarianism. That obstacle is the existence of ideals, conceptions of human excellence, aesthetic preferences, preferred states of the world of whatever kind. Someone might be willing, in the pursuit of an ideal, to flout others’ interests systematically. Hare has unkind things to say about such a person. He calls him a fanatic and says that his ideal may be a perverted one. But a universalization argument, he recognizes, cannot show why one genuinely devoted to an ideal ought not to attempt to realize it. Hare conjures up a Nazi willing to affirm that, were he a Jew, he should be exterminated. The world, holds the Nazi, would be better off that way. Such a Nazi is universalizing his ideal, and Hare concedes that he can catch him in no violation of the logic of the moral concepts. He may be, as Hare urges, a rare specimen; but he prevents the argument for utilitarianism from going through successfully.$^{98}$

$^{97}$Hare, *Freedom and Reason*, p. 123.
In his recent third book, Hare polemically champions utilitarianism. The book is concerned chiefly with rebutting those criticisms of utilitarianism that appeal to our moral intuitions. Hare argues that those intuitions serve utilitarian ends in most cases, but they are after all only a guide. The "critical level" of moral thinking enables us to move beyond appeals to intuition, to extract the nugget of utilitarianism hidden in the logic of the moral concepts. Perhaps because of the polemical animus of the book, Hare never quite explains how that logic yields utilitarianism.

In an earlier paper, however, he does offer a clear explanation, so I will focus on that paper. Hare first explicitly endorses the view that universalistic concern for interests leads to a maximizing view: "If I am trying to give equal weight to the equal interests of all the parties in a situation, I must, it seems, regard a benefit or harm done to one party as of equal value or disvalue to an equal benefit or harm done to any other party. This seems to mean that I shall promote the interests of the parties most, while giving equal weight to them all, if I maximise the total benefits over the population; and this is the classical principle of utility." Hare makes a deceptively simple attempt to meet "the problem of the fanatic, who has given me so much trouble in the past": "In so far as, in order to prescribe universally, I have to strip away (qua author of the moral decision) all my present desires, etc., I shall have to strip away, among them, all the ideals that I have. . . . This means that for the purposes of the moral decision it makes no difference who has the ideal. It means that we have to give impartial consideration to the ideals of ourselves and others." Morality thus requires that child abusers sus-

"Consider, for example, Hare, Moral Thinking, pp. 94–95, where Hare flatly asserts that we must forget our own preferences in considering others’ positions.


Hare, "Ethical Theory and Utilitarianism," in Lewis, p. 121. Hare dismisses the possibility of a Nazi so devoted to his cause that his desire outweighs
AGAINST UTILITARIANISM

pend their own ideals and pay heed to those of friends of the disabled—and vice versa. Once Hare’s Nazi gives impartial consideration to other ideals, he will be unable to recommend realizing his own; and again, that they are his own is uninteresting.

Hare’s theories have been criticized widely, and I do not wish to rehearse those criticisms here. I mean instead to press one point. Even on a generous construction of universalization, utilitarianism could only with great difficulty be universalized.

Universalization may be construed in quite a few senses, and Hare shifts a bit erratically among them. The crucial phrase is “give equal weight to the equal interests of all the parties in a situation,” a phrase which allows two competing interpretations. In the first, we may keep the interests glued firmly to the persons, and construe equal interests as something like “interests equally important in realizing life plans.” In the second, we may allow the persons as bearers of the interests to fade away, leaving the interests free-floating, and construe equal interests as something like “wants experienced with equal psychological urgency.”

The second interpretation yields Hare a recognizably utilitarian view. Once the persons disappear, all that is left is to maximize the realization of the interests. Questions of distribution cannot arise, for there is no one to distribute to. Why, though, should we take up this gloss of universalization? Surely it does not flow from a formal study of the logic of the moral concepts. The first interpretation is a more plausible candidate for the uni-

all competing ones as extremely unlikely and so irrelevant for our common reactions. This statement suggests that were such a Nazi to exist, he ought to be catered to; and Hare implicitly affirms that he should in Moral Thinking, pp. 171–172.


universalization embedded in morality, and it simply will not yield anything like utilitarianism. I am unable to universalize a view that might condemn a minority (or a majority with flaccid mental lives) to be used as fodder for others, for I take seriously the possibility of being one so used. Nor could I affirm a system in which I profited by others’ being so used. What is called for by the first view of universalization is a view structured like a rights view, or a welfare view with minimal levels guaranteed to each individual—some view on which, generally at least, each individual’s interests are safeguarded.

The assumption that causes the trouble here is that morality must be a matter of maximizing something, and all we need do is figure out what.\(^{105}\) This view prod s Hare toward utilitarianism and forces him to such an odd construal of universalization. It is not that the formal study of the moral concepts yields utilitarianism; it is rather that, with some arm-twisting, it can be made to. The source of the twisting, however, lies outside the formal theory. Regardless, the maximization assumption requires defense. With his meager conception of ethics as “the logical study of the language of morals,”\(^{106}\) Hare will be unable to defend the assumption. And once we wrest free of the hold that conflating rationality, maximization, and morality has on us, we may well wonder whether the assumption is defensible at all.

I pause to notice a strikingly counterintuitive implication of the program of deriving moral content from moral form, undertaken in different ways by Harsanyi and Hare. Were the program successfully executed, it would show that all other moral theories are, in fact, not moral theories at all. Note that it could


\(^{106}\)Hare, *Language of Morals*, p. iii.
not show that they were wrong, or unacceptable, just that they were not properly described as moral in the first place.\textsuperscript{107} Harsanyi and Hare are committed, strictly speaking, to saying that the theories of Plato, Aristotle, Kant, Rawls, and Nozick cannot properly be described as moral theories at all. That commitment invites an open-question argument of the kind both Harsanyi and Hare insist on.\textsuperscript{108} Once we focus on the descriptive component of the concept \textit{morality} or on the workings of what we call the moral concepts, it is perfectly sensible to ask, “Why should I be moral?” This question would mean, addressed to Harsanyi, “Why should I take up the position of your impersonal spectator?” To Hare it would mean, “Why should I universalize in this way?” Harsanyi would respond, “Because that is what it means to be moral.” But someone debating opting out of a social practice such as morality—or, more to the point here, pursuing a competing conception of morality—would not be swayed by linguistic reminders.\textsuperscript{109} Hare, officially leery of giving conceptual answers to substantive questions, has sketched an argument reminiscent of Plato and Aristotle that prudence is best served by the cultivation of dispositions to be moral.\textsuperscript{110} The limits of these sorts of arguments aside, the alternatives Hare should be considering are not morality and egoism, but his brand of morality and others. That he implicitly assumes his is the only brand available shows that he still assumes that \textit{morality} just means his theory. In that sense, his position is no better than Harsanyi’s. Given these dilemmas, would it be premature to recommend abandoning the attempt to derive moral content from moral form?

Brandt takes a different tack in arguing for utilitarianism. His

\textsuperscript{107}Note the ambiguity between two senses of \textit{moral} (contrasting with non-moral and immoral) carefully laid out by Onora Nell, \textit{Acting on Principle} (New York: Columbia University Press, 1975), pp. 2–4.


\textsuperscript{110}Hare, \textit{Moral Thinking}, pp. 188–205.
*Theory of the Good and the Right* is in part an attempt to develop “an account of concepts suited for a scientific psychological explanatory conceptual framework.” Accordingly the book bulges with elaborate views of rationality, desire, and so on. I mean to bypass discussion of those views. Brandt proposes that we understand questions about the good and the right as questions about choices that a fully rational person would make with ideally vivid representations of the alternatives. A fully rational person, in his view, is one whose desires and aversions have been flooded with all available information, and so restructured. The effects of this cognitive psychotherapy, Brandt realizes, will depend on the starting point of the individual undergoing it. So the theory will not converge on a unique solution: “It would be nice if we could demonstrate that all fully rational persons would support one and the same moral system. We shall in fact have to settle for something short of that.”

Nonetheless, Brandt thinks, “rational persons would probably opt for one within a narrow range of what we might loosely call ‘utilitarian’ moral systems; that is, ones the currency of which would maximize the expectable happiness or welfare of some large group, the size of the group depending on the benevolence of the chooser.” His argument is brief and a bit elusive. He tries to show that both perfectly benevolent and perfectly selfish rational individuals would choose a utilitarian view, and he infers that those with middling benevolence would do so as well.

Take first the benevolent chooser: “I define a ‘perfectly’ benevolent person as one who, between two options, always prefers the one associated with the greater long-term sum of expectable net happiness, irrespective of who is to receive it.” Unsurprisingly, Brandt decides that “the main inference is quite obvious”; such a fellow will choose utilitarianism. The definition, though, begs the question. Why should benevolence dictate disregard to all distributive considerations?

The argument from the rational selfish chooser warrants quotation at length:

---

112Brandt, *Good and Right*, p. 200.
113Brandt, *Good and Right*, p. 208.
AGAINST UTILITARIANISM

The serious options open to a perfectly selfish man, however, are restricted by the requirement of viability. Let us suppose, for the moment, that the persons in the group with whom he will interact and with whom he must form a moral community are equally as selfish as himself. Obviously a moral system which serves his interests at their expense would not enlist their loyalty, and there is no point in his supporting it. . . . If the selfish chooser wants, as he will, protection against crimes against the person, such as assault, negligent injury, and libel, he must choose a moral system which provides the same protection for others, thereby restricting his activities and giving them what they surely want. A selfish person who supports a rule which provides a desired circumstance for all because it, among feasible options, maximizes expectable welfare for him is inadvertently also supporting a rule which will maximize expectable welfare for the group (put each one on a higher ‘indifference curve’).¹¹⁴

But there is no reason to think that maximizing the utility of each individual means maximizing the utility of the group. The group’s utility may be increased by lowering the utility of some to get a greater increase in the utility of others. That possibility creates the conflict between rights theorists and utilitarians, so prominent in recent literature, yet Brandt seems unaware of it. The confusion here is the confusion animating Bentham’s dual standard, and was noted by Sidgwick. It is disheartening, to say the least, to see it reappearing at this late date.

Since we have no reason to think that the extremes of benevolence (as generally understood) and selfishness lead to utilitarianism, we need not explore the possibility that the spectrum is discontinuous, that those in the middle would deliberate differently.

Foundational Structure of Utilitarianism

Utilitarianism, for all its vaunted precision, cannot tell us what to do. It frames a choice procedure only by purging information, reducing descriptions of outcomes to distributions of

pleasures and pains, and reducing those distributions to one summed surplus of pleasure over pain. Even then, it sets out an incomplete choice procedure which does not guide us in cases of risk and uncertainty. Indeed, that choice procedure can never get off the ground, since pleasures and pains are a wildly assorted host of incommensurable experiences. Neither classical nor modern utilitarians offer any satisfactory justification of utilitarianism, any account that would give us some reason to redouble our efforts to solve these problems. I therefore propose that we write off utilitarianism as incomprehensible.

However incomprehensible it actually is, though, utilitarianism seems at least to permit, indeed to enjoin, the redistribution of felicific goodies. This redistribution is mandated not to render the lives of the poor dignified, not in the name of fairness, justice, or equality, but to maximize the utility score of some mystic whole named society. Suppose that throwing eggs at Jack will make us very happy, happy enough to offset Jack’s unhappiness and any indirect disutilities. (Rule-utilitarians may suppose that a practice of throwing eggs at people like Jack would be felicific. Hare may suppose that critical moral thinking would approve the cultivation of a moral intuition that it is right to throw eggs at people like Jack.) Then utilitarianism is a theory about the moral rightness, indeed obligatoriness, of splattering people with raw egg, on the ground that “society” is happier if we do so. That, surely, is a view we need not take seriously as a leading moral and political theory.

A utilitarian may protest the mention of such an example. “Didn’t you forswear the appeal to our moral intuitions? And aren’t you doing just that here, inviting us to realize that utilitarianism enjoins behavior we think of as pernicious or just plain silly?” Actually, my point is different; it’s that society becomes a mystical whole in utilitarianism. But it is worth noting one last puzzle arising here. Again, utilitarians do often insist that, properly understood, utilitarianism coheres beautifully with our moral judgments. A sophisticated utilitarianism, it has been argued, can accommodate our commitments to liberty, equality, even autonomy. It would not dictate egg-splattering, even for the uniformly unpopular. This move, however, gives up the
fight. If utilitarianism yields the same judgments we do, why insist so strenuously on it? What would be at stake in adopting or rejecting it? Utilitarians too eager to show the fit between utilitarianism and our settled views invite a pragmatic response: if two theories make the same prescriptions, they are for all intents and purposes the same theory. Nor can utilitarians suggest that they at least have a theory, while we have nothing but muddled intuitions. If our intuitions are so perfectly utilitarian, we might as well stick with them; and given its difficulties, it's not clear that utilitarianism delivers the benefits a real theory is supposed to.

Still, there are attractive features of utilitarianism, features attractive enough to explain the support it has commanded during its long and illustrious history. It demands no exotic metaphysical or theological commitments. It makes human welfare the point of morality. It offers (or seems to offer) reasons for its conclusions, reasons that have a nicely hard-boiled air about them. (If for example we challenge Bentham's economic egalitarianism, we are not directed to the misty realm of rights, human dignity, and the rest. Instead we receive a brisk argument about diminishing marginal utility; if we challenge the force of that argument, we are told triumphantly that it follows deductively from the greatest-happiness principle.) It allows facts to affect and even determine our moral principles. Finally, it is in some ways a perfectly egalitarian theory: "each to count for one, none for more than one," as the injunction goes. Yet these features are caught up in a foundational structure that causes lots of problems and solves none. I will conclude by mentioning two problems posed by the structure of the theory.

Utilitarianism poses the is/ought problem, that of justifying a move from the realm of description to that of evaluation and prescription, in a peculiarly sharp way. The move occurs all at once, in the claim that pleasure (or preference satisfaction) is the good. Once that premise is employed as a foundation, the weight it bears—no less, in the utilitarian view, than that of supporting a complete moral and political theory, of serving as "the sole and all-sufficient reason for every point of practice whatsoever"—makes the demand for a justification all the more pressing. The
Without Foundations

utilitarian widens the is/ought gap and makes it look unbridgeable. Nor are hopes of constructing a sound bridge encouraged by the flimsy attempts utilitarians make to justify their theory.

Utilitarianism is further embarrassed by the reappearance of moral notions in the foundation allegedly supporting them. Much of our pleasure and pain is parasitic on moral notions.\(^{115}\)

Agents take pleasure in acting morally and feel remorse at acting immorally. Even if a friend would live his happiest possible life staring at soap operas all day, we are glad when he discards the television and takes up some more challenging life plans. Will utilitarians count these pleasures and pains? Will they count those we feel because we hold nonutilitarian moral views? (It will be much easier to show that slavery is unfelicific if we count the rage and resentment arising from the nonutilitarian understanding that slavery is wrong.) Similarly, Hare’s demand that we universalize our moral judgments is in fact a demand for fairness, for refusing to exploit the contingencies of superior position. We cannot cleanly separate pleasure or universalization from morality. So neither will properly serve as a foundation for morality.

We want, if we can, to retain utilitarianism’s attractive features and discard both its foundationalist structure and its tendency to purge information. I turn now to David Hume and Adam Smith, who, I think, show that we can.

\(^{115}\)That the Deontology contains a lengthy discussion of the morality of “emission of gas from the alimentary canal” (II:237–240) is perhaps neither a sign of Bentham’s senility nor a sign of editorial betrayal, but rather an attempt to evade this issue. Compare David Baumgardt, Bentham and the Ethics of Today (Princeton, N.J.: Princeton University Press, 1952), pp. 491–492.