

PHIL 202: Ethics Core Course
Winter 2005; David O. Brink
Topic: Mill's Moral and Political Philosophy
Handout #4: The Proof and Sanctions of Utility

THE PROOF

We've focused so far on understanding Mill's version of utilitarianism, especially his conceptions of happiness and duty. Now we should consider his justification of utilitarianism, which he offers in his discussion of the "proof" of the principle of utility in chapter iv. Mill claims that the utilitarian must claim that happiness is the one and only thing desirable in itself (iv 2). He claims that the only proof of desirability is desire and proceeds to argue that happiness is the one and only thing desired. He argues that a person does desire his own happiness for its own sake and that, therefore, happiness as such is desired by and desirable for its own sake for humanity as a whole ("The aggregate of all persons") (iv 3). He then turns to defend the claim that happiness is the only thing desirable in itself, by arguing that apparent counterexamples (e.g. desires for virtue for its own sake) are not inconsistent with his claim (iv 5-8). Let's see if we can reconstruct the steps in this argument. A traditional reconstruction might look something like this.

1. Utilitarianism is true iff happiness is the one and only thing desirable for its own sake (and not for the sake of something else).
2. The only proof of desirability is desire.
3. Each person desires his own happiness for its own sake (and not for the sake of something else).
4. Hence, happiness, as such, is desired for its own sake (and not for the sake of something else) from the point of view of humanity (= the aggregate of persons).
5. Hence, happiness, as such, is desirable for its own sake (and not for the sake of something else).
6. Happiness is the only thing desired for its own sake (and not for the sake of something else). Other things – such as virtue, health, music, money, and power – can come to be desired for their own sakes, but then they are desired as parts of happiness.
7. Hence, happiness is the only thing desirable for its own sake (and not for the sake of something else).
8. Hence, utilitarianism is true.

SPOT THE HOWLER

The "proof" has, at least in some quarters, threatened Mill's reputation as careful philosopher. It is commonly thought to be riddled with major mistakes – mistakes of inference and implausible assumptions. Here is a partial list of concerns about Mill's argument, as traditionally conceived.

- (1) is true only if desirable means worthy of being desired, not if it means capable of being desired.
- Even so, (1) is false. Even if happiness were the one and only thing desirable for its own sake, this would establish only a claim about the good or "ends". It is not a claim about duty or right action. Utilitarianism not only claims that the good is human happiness but

goes on to define the right in terms of promoting the good. The second claim does not follow from the first. Mill appears to recognize this insofar as he at least tacitly distinguishes the two claims (iv 9; cf. ii 2). However, he goes on to infer the second claim from the first without any independent argument.

[I]f human nature is so constituted as to desire nothing that 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. If so, happiness is the sole end of human action, and the promotion of it the test by which to judge all human conduct; from whence it necessarily follows that it must be the criterion of morality, since a part is included in the whole [iv 9].

- (2) is false. Desire is not proof of desirability. People can and do have mistaken desires about what is good. Indeed, if Mill is either a hedonist or a perfectionist he must think that people can and do have desires that fail to track the good or even the apparent good.
- Is (3) true? Do masochists or selfless altruists desire their own happiness for its own sake?
- (4) may be incoherent and certainly does not follow from (3). It's not clear that aggregates of persons have desires. Even if they did, it's doubtful that you could infer what the aggregate desires from facts about what its members desire. That would seem to involve a compositional fallacy.
- (5) is presumably equivalent to the claim that happiness is good. But is it good simpliciter or good for the aggregate?
- It's not clear how to understand (6). One would think that the aim is to make claims parallel (4) and (5). But then (6) needs to be understood as making another claim about aggregate psychology. And this raises some of the old questions about aggregate psychology. However, much of the discussion in iv 5-8 seems to be about individual psychology. Mill seems to be saying that insofar as individuals do have intrinsic desires for things other than their own (?) happiness the objects of intrinsic desire are desired as parts of their own (?) happiness. Perhaps this is Mill's initial claim from which he then hopes to infer, as he did from (3)-(4), that the general happiness is the only thing desired by the aggregate for its own sake (and not for the sake of something else). This inference would, of course, give rise to the same sort of worries we raised about the inference from (3)-(4). In particular, we might doubt that aggregates of persons have any aims, much less ultimate aims. And even if we conceded that they did, it's not clear that we could infer facts about the desires of aggregates from facts about the desires of its members. That would seem to involve a compositional fallacy.
- Even if we accepted this defense of (5) and (7), this would only establish that happiness as such was the only thing desirable or good for the aggregate. It looks like we could have parallel claims about the agent's own happiness being the only thing desirable or good for the individual. But this might seem to imply that while the aggregate should pursue or promote the general happiness individuals should pursue or promote their own happiness. That would not be a defense of utilitarianism.

A DEFLATIONARY READING OF THE PROOF

Can we do any better by Mill in our interpretation of his proof? At the beginning of ch. iv Mill repeats his admonition that ultimate ends do not admit of proof in the ordinary sense of that term (iv 1). Here, he alludes to his introductory remarks in ch. i that provide caution about

the kind of proof available for first principles.

... I shall ... attempt to contribute something toward the understanding and appreciation of the “utilitarian” or “happiness” theory, and toward such proof as it is susceptible of. It is evident that this cannot be proof in the ordinary and popular meaning of the term. Questions of ultimate ends are not amenable of direct proof. ... We are not, however, to infer that its acceptance or rejection must depend upon blind impulse or arbitrary choice. There is a larger meaning of the word “proof,” in which this question is as amenable to it as any other of the disputed questions of philosophy. The subject is within the cognizance of the rational faculty; and neither does that faculty deal with it solely in the way of intuition [I 5].

So it seems that Mill is looking for arguments that may be rationally persuasive without providing incontrovertible proof. How might this help us better understand the proof?

For one thing, Mill need not confuse desire and desirability. He recognizes that they are distinct, but says that desire is our only proof of desirability (iv 3). In saying this, he need not presuppose that desiring something confers value on (obtaining) it. He can be a perfectionist, as we interpreted him, in the higher pleasures doctrine. As he recognizes there (especially the dignity passage), our desires often reflect value judgments we make, explicitly or implicitly. If so, our desires will be evidence of what we regard as valuable, and our reflectively acceptable desires may provide our best defeasible test of what things are objectively valuable.

Mill first applies this test to what each of us desires for her own sake. His answer is that what each of us desires for his or her own sake is happiness (iv 3). We needn't interpret Mill as endorsing psychological egoism at this point. Mill is not saying that each of us can only focus on her own happiness. Rather, he's saying when each of us does focus on her own ends or sake, we find that each cares about her own happiness. Another way to put Mill's point is that prudential concern focuses on the agent's happiness.

Mill goes on to say that just as each person's own happiness is a good to that person, so too happiness, as such, is a good to the aggregate of persons. But we need not suppose that Mill is attributing a psychology, much less an egoist psychology, to humanity as a group. Instead, we can read Mill as claiming that just as the agent's own happiness is the object of prudential concern, so too happiness as such is the proper object of disinterested or impartial concern.

On this reading, Mill is not trying to derive utilitarianism from egoism; rather, he is assuming that the moral point of view is impartial in a way that prudence is not. If so, just as prudence aims at the agent's own happiness, so too morality, which is impartial, aims at happiness as such. On this reading, the structure of Mill's proof looks something like this.

1. Prudence is partial.
2. Because prudence is partial, it aims at the agent's own happiness.
3. Morality, by contrast, is impartial.
4. Because morality is impartial, it aims at happiness as such.
5. If the moral point of view aims at happiness as such, then it is the moral duty of each to promote happiness.
6. Hence, utilitarianism is true.

If this is the right way to understand Mill's proof, then his justification or defense of utilitarianism consists in assuming that the moral point of view is impartial and claiming that utilitarianism is the right way to understand impartiality. One might wonder if utilitarianism is the only or the best way to understand impartiality. Indeed, this is one way of understanding now familiar worries about the implications of utilitarianism for issues of distributive justice and

rights. But it is, I think, a virtue of this reading of the proof that it identifies the defense of utilitarianism with the feature of it that made it a progressive influence historically. Morality is impartial, and impartiality requires taking everyone's interests into account – and not just those of some select few – and weighing them equally – and not with a thumb in the scales for some select few.

SANCTIONS

In what is probably the most understudied chapter of Utilitarianism (ch. iii), Mill addresses the question of the ultimate sanction of the principle of utility. He understands this alternately as a question about “the motives to obey it” and the “source of its obligation ... [or] binding force” (iii 1). Nor does Mill think that this issue about the sanction of the principle of utility is an idle one. There is a potential worry about the sanctions of utilitarianism that apparently has its source in prudence or self-interest.

He says to himself, I feel that I am bound not to rob or murder, betray or deceive; but why am I bound to promote the general happiness? If my own happiness lies in something else, why may I not give that the preference [iii 1]?

But we can imagine this person unmoved by the moral demands in question and not just by a utilitarian reconstruction of these demands. And Mill is very clear that he thinks this issue about the sanction of utilitarianism arises for any moral theory and so poses no special problem for utilitarianism (iii 1, 2, 3, 6).

Mill is not entirely clear how he understands the sanctions worry. Is it just a problem about motivation – the extent to which people are motivated to comply with utilitarianism or other moral conceptions and the extent and manner in which they might be brought to comply more fully or more easily? Or is it more a problem of authority – the extent to which people have reason to comply with utilitarianism or other moral conceptions? The sort of self-interested challenge that Mill identifies at the beginning (iii 1) is usually part of an amoralist challenge to the authority of other-regarding morality. But much of Mill's actual discussion seems more addressed to the motivational issue.

Mill thinks that an account of the sanctions of utilitarianism or any other moral conception should distinguish between external and internal sanctions. The external sanctions are those penalties that can be visited on noncompliers by another, whether the others are mortals, individually or collectively, or God (iii 3). Penalties might include official acts of the state (legal punishment) or community (official ostracism or public humiliation) or God (punishment meted out on Judgment Day). They might also include more informal and/or diffuse sanctions, such as those incurred when my bad acts have bad reputational effects that harm my social and/or economic prospects or opportunities. Of course, mortals are neither omniscient nor omnipotent. So there's no guarantee that noncompliers will be visited with external sanctions, at least in this world. If noncompliers can conceal their noncompliance (concealing either the fact of noncompliance or their responsibility for the noncompliance), they will suffer no external sanctions. And existing external sanctions may be insufficient to deter noncompliance even if the noncompliance cannot be concealed. For instance, I may not be deterred from violating our for mutually beneficial exchange of services by the bad reputational effects of noncompliance if you have already performed your part of the contract and the benefits of “taking the money and running” dwarf the costs of being excluded from future cooperation with you and other members of my community. Of course, we can collectively try to ratchet up the costs for detected noncompliance by criminalizing noncompliance, adopting draconian

punishments for noncompliance, and pursuing extradition treaties with other communities. In this sense, the provision of adequate external sanctions is not entirely outside our control and could be treated as a political or social goal and achievement.

Mill spends somewhat more time discussing internal sanctions for noncompliance (iii 4-11). The internal sanction for noncompliance is Conscience (iii 4). The issue that interests Mill most about Conscience seems to be whether it is innate (iii 6-11). Mill represents the intuitionists (transcendentalists) as claiming that it is innate (ii 6). Mill seems to equate Conscience being innate with its not being analyzable into other desires, passions, or emotions. Conscience might involve a desire to please God (iii 4); it might involve the desire to liked by others (iii 4); it might involve the desire to form a union with others (iii 10); or it might involve sympathy and empathy (iii 7). Mill endorses the claim that Conscience is not innate (iii 8-11). He thinks that Conscience involves these other emotions, passions, and desires and is built up out of them. Some of these capacities are themselves innate, but their proper development and exercise is a historical accomplishment that is not yet complete and that is fragile. So it is within our power individually – but more importantly collectively – to change social conditions so as to make Conscience and the internal sanctions of duty more robust.

In these claims Mill seems clearly to be focused on the motivational dimension of the internal sanctions of duty. But we might wonder what he has to say about the authority of these sanctions. The capacities underlying Conscience may not be distributed equally, and these capacities can wither or flower. But why should we care about Conscience? Is there anything wrong with me – in particular, any failure of practical reason – if I do not have a Conscience or it is weak? Is there any good reason I should cultivate a Conscience if I don't have one (assuming that's possible) or that I should invest resources in maintaining my Conscience if I do have one? Mill seems not to address these questions.

Is Mill right to claim that the worry about sanctions is really a general worry about the sanctions for noncompliance with other-regarding morality of any sort and so poses no special problem for utilitarianism? Contemporary writers have noticed that utilitarianism seems to be potentially very demanding, much more so than commonsense morality. For instance, reformist utilitarians, such as Peter Singer, have argued that utilitarianism entails extensive duties of mutual aid that would call for enormous changes in the lifestyles of all those who are even moderately well off. And critics of utilitarianism have treated the demandingness of utilitarianism as one of its principal flaws. Rawls has argued that the sort of interpersonal sacrifice that utilitarianism requires violates the strains of commitment in a well-ordered society. And Bernard Williams has argued that the demandingness of utilitarianism threatens the sort of personal projects and partial relationships that help give our lives meaning. The common complaint here is that utilitarianism's demands threaten to offend against a requirement of psychological realism, according to which the demands of an acceptable moral theory must be ones that can be incorporated into a reasonable and satisfying life plan. It's debatable whether we should accept this constraint of psychological realism and whether utilitarianism really does violate this constraint. But it's not a crazy constraint and utilitarianism might seem to violate it. If so, even if we accept Mill's psychological focus in discussion the sanctions of utilitarianism, we may conclude that he is wrong or at least premature to assert that there is no special sanctions problem for a utilitarian theory of duty.

