Utilitarianism without Consequentialism: The Case of John Stuart Mill

Daniel Jacobson
Bowling Green State University

In this essay I will argue, flouting paradox, that Mill was a utilitarian but not a consequentialist. According to the textbook definition, of course, utilitarianism just is the combination of a certain sort of theory of the good (as pleasure, happiness, or flourishing) and a consequentialist theory of the right. My conclusion thus seems necessarily false. Nevertheless, the argument will proceed in two stages. First, I argue that there is logical space for a view that deserves to be called utilitarian despite its rejection of consequentialism. Then I argue that this position was in fact occupied by the most renowned utilitarian, John Stuart Mill.

The first step in my argument rests on what might be considered a clever trick, at best; but this is to be expected when one argues for a claim that seems necessarily false. I concede from the beginning, though, that were my conclusion merely a semantic or conceptual point, this argument would be of little interest. But the conclusion is quite interesting because the tricky step in the argument exposes an implicit feature of consequentialism that is both substantive and dubious and has received inadequate attention. Examination of this implicit assumption will reveal an important contrast between consequentialism and its

I would like to thank David Boonin, Julia Driver, Chris Heathwood, Doug Portmore, Henry Richardson, Walter Sinnott-Armstrong, and David Sobel; audiences at the University of Colorado, the University of Illinois at Urbana-Champaign, the International Society for Utilitarian Studies, and Bowling Green State University; two anonymous referees for this journal; and especially Don Hubin for helpful comments on earlier drafts of this essay.

Philosophical Review, Vol. 117, No. 2, 2008
DOI 10.1215/00318108-2007-035

cognates and will help illuminate an insufficiently explored aspect of Mill's moral and political philosophy. I refer here to Mill's sentimentalist metaethics, which proves crucial for understanding his view of morality as comprising just one distinct sphere within what he called "the Art of Life, in its three departments, Morality, Prudence or Policy, and Aesthetics; the Right, the Expedient, and the Beautiful or Noble, in human conduct and works" (*System of Logic*, *CW* 7:949).[1]

The failure of the standard interpretation of Mill to account for his sentimentalism has led to a profound misreading of his moral theory, which makes Mill out to be a predecessor of the most fashionable view in the neighborhood of utilitarianism: namely, multilevel maximizing act-consequentialism.[2] Despite the popularity of this theory and the predominance of this interpretation, it cannot be reconciled with some of Mill's most distinctive and important claims. Its advocates therefore must ignore or traduce crucial aspects of his work, to the point of attributing a dissimulation hypothesis to Mill, on which he (sporadically) conceals his genuine but "esoteric" morality for instrumental purposes.[3] Surely this should be an interpretive strategy of last resort, especially when another reading both comports better with what he actually wrote and fits more neatly into his historical context. Moreover, I find Mill's highly unorthodox brand of utilitarianism especially attractive for several reasons, including its engagement with disparate emotions, both moral and nonmoral, and its resistance to the pervasiveness

1. J. M. Robson, ed., *Collected Works of John Stuart Mill* (Toronto: University of Toronto Press, 1969). References to Mill will be to the *Collected Works* (*CW*), except for *Utilitarianism* (*U*) and *On Liberty* (*OL*), which will be given in the text as (Title: chapter. paragraph).

2. While there have been various other interpretations of Mill, none has won wide favor. Because of the immensity of Mill scholarship, I make no attempt to be comprehensive, but instead aim to reflect the current state of play as illustrated in important recent surveys and anthologies published by major academic presses.

3. The idea of an esoteric morality, which must be hidden from the masses, is due to Henry Sidgwick, *Methods of Ethics*, 7th ed. (Indianapolis, IN: Hackett, 1981), esp. 490. Sidgwick (ibid., 490) notes further that "similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric." Derek Parfit considers the possibility that consequentialism might be "self-effacing," in that it would be better by the theory's own lights if few, or even no one, believed it. See Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1986). I am not claiming that a self-effacing utilitarianism is no utilitarianism at all, or that self-effacing views are thereby self-defeating. Rather, I aim to illustrate the extreme cost of attributing a dissimulation hypothesis to Mill, when he never champions—or even considers—the doctrine of esoteric morality, and elsewhere makes claims that deeply conflict with this approach.

of more orthodox forms of the theory which allow moral considerations to occupy the entire evaluative domain rather than only one among three spheres of value.

Mill thought this error characteristic of moralists of all theoretical bents, though he attributes it expressly to Bentham:

> This error, or rather one-sidedness, belongs to him not as a utilitarian, but as a moralist by profession, and in common with almost all professed moralists, whether religious or philosophical: it is that of treating the moral view of actions and characters, which is unquestionably the first and most important mode of looking at them, as if it were the sole one: whereas it is only one of three, by all of which our sentiments toward the human being may be, ought to be, and without entirely crushing our own nature cannot but be, materially influenced. ("Bentham," *CW* 10:112)

Both Mill's sentimentalism and his resistance to the narrow-mindedness of moralism are illustrated nicely by this quotation. In what follows, I will argue that Mill's sentimentalist ethics renders his account of right and wrong both less strictly impartial and less pervasive than consequentialism presupposes.

More than fifty years ago, J. O. Urmson, lamenting the state of Mill scholarship, wrote that if Mill were interpreted with "half the sympathy accorded to Plato, Leibniz, and Kant, an essentially consistent thesis can be discovered which is very superior to that usually attributed to Mill and immune to the common run of criticisms."[4] Undoubtedly, matters have improved since then, thanks in no small part to Urmson, although his own rule-utilitarian interpretation of Mill has not been widely accepted.[5] The standard view of Mill's moral theory today still attributes to him a maximizing act-consequentialist moral theory despite

4. J. O. Urmson, "The Interpretation of the Moral Philosophy of J. S. Mill," in *Mill's Utilitarianism: Critical Essays*, ed. David Lyons (Oxford: Rowman and Littlefield, 1997), 1.

5. Urmson's brief sketch of an interpretation of Mill's moral theory has been effectively challenged by Fred Berger, *Happiness, Justice, and Freedom: The Moral and Political Philosophy of John Stuart Mill* (Berkeley: University of California Press, 1984). For another influential challenge to Urmson, which I find less convincing, see D. G. Brown, "Mill's Act-Utilitarianism," *Philosophical Quarterly* 24 (1974): 67–68. Although David Lyons defends and develops one crucial aspect of Urmson's reading, which I will utilize also, Lyons puts his view forward as a reconstruction of Mill's view and admits that his is the minority opinion: "Scholars generally prefer an act-utilitarian reading of Mill. This may be a reflection of the fact that consequentialists generally and utilitarians in particular tend to favor 'act' versions of those theory-types." David

its inconsistency with much of Mill's writing and its vulnerability to familiar and powerful objections. As Roger Crisp, perhaps the leading advocate of this interpretation, expresses it: "*The* right action will be that which produces the greatest balance of happiness over unhappiness overall," according to Mill, and "any other action [than this optimific one] will be wrong."[6] Geoffrey Sayre-McCord claims, similarly, that "according to [Mill's] standard of conduct, an agent has performed the right act if and only if that act is among the agent's best available options. To have taken any less than the best available option is, Mill thinks, to have performed the wrong act."[7]

Although this standard interpretation attributes fundamentally the same moral theory to Mill as those readings Urmson excoriated as "so unsympathetic and so incorrect," its current advocates far exceed Urmson's antagonists in their scholarship and philosophical acumen.[8] And since the *multilevel* version of the direct and maximizing theory they advocate—of the sort developed most compellingly by R. M. Hare and Peter Railton—surpasses older and less sophisticated versions of consequentialism, their reading is also more charitable to Mill.[9] Yet the move to a multilevel view does not change the theory's criterion of rightness, which still identifies right action as the best available option: the one that creates the greatest net good. It simply differentiates between

---

Lyons, "Introduction," in *Rights, Welfare, and Mill's Moral Theory* (New York: Oxford University Press, 1994), 21.

6. Roger Crisp, ed., *J. S. Mill: Utilitarianism*, Oxford Philosophical Texts (Oxford: Oxford University Press, 1998), 115; Roger Crisp, *Mill on Utilitarianism*, Routledge Philosophy Guidebooks (London: Routledge, 1997), 96. As these prestigious recent commissions suggest, Crisp's reading of Mill can fairly be called standard.

7. Geoffrey Sayre-McCord, "Mill's Proof of the Principle of Utility: A More Than Half-Hearted Defense," *Social Philosophy and Policy* 18 (2001): 331–32. In fairness to Sayre-McCord, it should be noted that he makes this claim in passing, while in pursuit of a different argument. However, the fact that he advances it without argument provides even more support for my contention that this has become the conventional interpretation of Mill's moral theory.

8. Urmson, "Moral Philosophy of J. S. Mill," 8.

9. Although the idea of an esoteric utilitarian morality was broached by Sidgwick, the *locus classicus* of its modern development, in the form of a sharp separation between the criterion of rightness of a moral theory and its recommendations for ordinary moral thinking, can be found in R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981); and Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (1984): 134–71. Hare introduced talk of different levels of moral thinking, while Railton calls this approach "sophisticated" act-consequentialism.

this criterion and the decision procedure, or ways of moral thinking, recommended by the theory. The basic insight is that consequentialism need not recommend that a moral agent think like a consequentialist, in the sense of aiming at maximizing the common good, any more than an egoistic hedonist must aim directly at pleasure. Indeed, if overtly consequentialist thinking has the problems often claimed of it, then the theory will *not* so recommend. This is an important insight and a real advance, but the more sophisticated theory accepts the same old account of right and wrong; its novelty consists in its recommendations for how to engage in moral thinking and moralizing. In particular, the multilevel view allows consequentialists to avow—and even to believe—claims that, strictly speaking, are false according to their theory. It recommends alienation from the most exigent demands of morality, and even dissimulation (to oneself and others) about the true but esoteric morality, whenever that would have better consequences than sincerity.

This essay will not attempt to demonstrate that Crisp's reading is largely wrong or that Urmson was fundamentally correct, contrary to the conventional assessment, although that is my view.[10] My complaint against Crisp and the other advocates of the standard interpretation is nearly antithetical to Urmson's complaint against his antagonists. Mill is now read not unsympathetically but anachronistically: too much through the lens of twentieth-century developments in ethical theory, specifically in the evolution of consequentialism to its current level of sophistication. By reminding ourselves of the state of play in the nineteenth-century debates over utilitarianism, we can avoid problems arising from reading Mill as a sophisticated, twentieth-century consequentialist, engaged in principled dissimulation about an esoteric morality. Moreover, by bringing to light an implicit presupposition of consequentialism, we can come to see the theoretical advantages of rejecting the strict form of impartiality it embraces.

10. Although there are problems of both omission and commission in Urmson's arguments, I ally myself with Urmson and Lyons on the whole, while deviating from each in some crucial respects. For another like-minded reading, see Alan Fuchs, "Mill's Theory of Morally Correct Action," in *The Blackwell Guide to Mill's Utilitarianism*, ed. Henry West (Oxford: Blackwell, 2006), 139–58. However, in my opinion no one has adequately connected Mill's sentimentalist metaethics with his classical liberalism, as expressed especially in *On Liberty*. See also John Skorupski, "The Place of Utilitarianism in Mill's Philosophy," in West, *Guide to Mill's Utilitarianism*, 46–60. An excellent treatment of many of these issues can be found in John Skorupski, *John Stuart Mill* (London: Routledge, 1989), but I think even Skorupski does not adequately appreciate Mill's sentimentalism.

Hence, even readers who are not convinced by the argument of the first section of this essay—which claims that there can be nonconsequentialist forms of utilitarianism—can grant the possibility of a moral theory with some distinctive advantages over ordinary consequentialism and which corresponds better to some famous aspects of Mill's thought. Furthermore, readers who are not persuaded by the argument of the second section—that Mill is best read as rejecting an essential element of consequentialism—can grant that several of Mill's explicit claims and deepest commitments cannot be reconciled with that theory. My thesis that Mill was a utilitarian but not a consequentialist is not merely a rhetorical flourish, however. If I am right to think that current Mill scholarship suffers from anachronism, and in particular that modern readers have lost sight of the fact that Mill's conception of utilitarianism was considerably broader than can be accommodated by contemporary consequentialist theory, then this way of putting the point is especially apposite because 'utilitarianism' is a nineteenth-century term, whereas 'consequentialism' was coined in the twentieth century.

I.

Let us begin with the trick: the move that avoids self-contradiction. Surely *consequentialism* is a philosophers' term of art, which means whatever philosophers have meant by it over the past half-century or so, when the term was coined and earned its place in the philosophical lexicon. By contrast, *utilitarianism* was a movement in the history of ideas.[11] Hence, that appellation must be understood broadly enough to include the views of the classical Utilitarians—in particular, Jeremy Bentham and John Stuart Mill, unquestionably the two most important utilitarians prior to Henry Sidgwick. But because 'utilitarianism' has a semantic debt to the history of philosophy that 'consequentialism' does not, it is possible for the two theories to diverge in unexpected ways. This despite the fact that the very concept of consequentialism was derived from utilitarianism—that is, from the orthodox form of the theory that became standard after Sidgwick—by abstracting away from its theory of the good, whence the textbook definition, on which utilitarianism entails consequentialism. Nevertheless, if some of the classical

11. While there is some doubt as to the term's provenance, it is typically attributed to Bentham and its popularity ascribed to J. S. Mill's founding of the Utilitarian Society in the 1820s.

Utilitarians did not accept the presuppositions that philosophers now routinely associate with consequentialism, then they were not consequentialists.

In claiming that consequentialism must be understood however philosophers commonly understand it, I am not supposing that its philosophical usage is uniform or precise. It is not.[12] But we can bracket our target quickly and relatively uncontroversially and then home in on it in a principled way. Some characterizations of consequentialism imply that it requires maximal promotion of the good, impartially considered, thereby ruling out both indirect (for example, rule-based) and satisficing (that is, nonmaximizing) versions of the theory. Shelly Kagan thus defines consequentialism as "the view that an act is right if and only if it leads to the best consequences."[13] Yet Kagan's usage seems artificially stipulative since he grants that "a maximizing approach is not the only one compatible with act consequentialism," though historically, "the fact remains that act consequentialists have almost always been maximizers."[14] By his own admission, his official definition is too narrow to capture either common usage or theoretical possibility. Even so, Kagan's gloss is common and describes the most popular form of the theory; it can aptly be called *orthodox consequentialism*.

12. Indeed, a notion of consequentialism has recently been introduced according to which, by stipulation, "every moral view is consequentialist." See James Dreier, "Structures of Normative Theories," *The Monist* 76 (1993): 24. This approach fixes on the dictum that it must be right to produce the best possible consequences, and then understands the concepts of *value* and *consequence* in whatever manner is necessary in order to preserve this maximizing dictum. In particular, it does so by adducing the notion of agent-relative value. I am not convinced of the merits of this approach since the claim that it is always right to bring about the best consequences seems to me far from trivial, and agent-relative conceptions of value seem quite problematic from the perspective of value theory. In any case, this will surely remain an idiosyncratic usage. Obviously Mill's utilitarianism, like every other moral view, counts as a form of consequentialism in this truistic sense. See also Douglas Portmore, "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation," *Ethics* 113 (2003): 303–32. Portmore develops a conception of consequentialism without the presupposition of agent neutrality, on which every agent is given the same set of aims. This approach does not make every moral view consequentialist, but it does include many theories that are not typically classed as such. This distinction will not be crucial to my argument, and I will largely ignore it hereafter. Compare Portmore, "Dual-Ranking Act-Consequentialism," *Philosophical Studies* (forthcoming), where he acknowledges that this sort of theory falls outside the range normally considered consequentialist.

13. Shelly Kagan, *Normative Ethics* (Boulder, CO: Westview, 1998), 61.

14. Ibid., 219, 223.

On the other hand, one can also find characterizations of consequentialism broad enough to include any moral theory on which the rightness and wrongness of action is somehow determined by consequences alone. Thus Walter Sinnott-Armstrong first defines consequentialism as the view that "whether an act is morally right depends only on consequences," but he later concludes that, although this thesis is necessary, "it is less clear whether that claim by itself is *sufficient* to make a theory consequentialist."[15] Sinnott-Armstrong here deliberately avoids two thorny issues for consequentialism: what count as consequences, and which consequences count for moral judgment. But however we answer those questions, consequentialism requires more than this minimal thesis if we are to avoid drastically revising our taxonomy. As Sinnott-Armstrong notes, this capacious conception would make some unlikely theories consequentialist: not only ethical egoism, but also such theories as Bentham's imaginary principle of asceticism, which instructs us to minimize happiness. I propose to use Mill's name for the class of normative theories with this structure, on which considerations of goodness are the sole normative determinant of rightness: these are *teleological* theories.[16] Although the distinction I propose between teleology and consequentialism is not uncontroversial, it should be noted that Samuel Scheffler defends a teleological theory, in this sense, in a book entitled *The Rejection of Consequentialism*.[17]

Following both Kagan and Sinnott-Armstrong in spirit, though neither in letter, I will eschew both these admitted extremes and seek a more standard gloss of consequentialism.[18] This choice is not just a

15. Walter Sinnott-Armstrong, "Consequentialism," in *The Stanford Encyclopedia of Philosophy*, Summer 2003 Edition, ed. Edward N. Zalta, 2, 13, plato.stanford.edu/archives/sum2003/entries/consequentialism (emphasis in original).

16. In his *System of Logic*, Mill differentiates between the propositions of Science, which assert matters of fact, and normative propositions, which identify the goal of any practical domain or Art (as he calls them). "Every art is thus a joint result of laws of nature disclosed by science, and of the general principles of what has been called Teleology, or the Doctrine of Ends" (*Logic*, *CW* 7:949). In all of practical reasoning, including morality, Mill claims the principle of utility to be the final end: its teleology.

17. See Samuel Scheffler, *The Rejection of Consequentialism*, rev. ed. (Oxford: Oxford University Press, 1994). In Scheffler's view, agents are allowed to favor their own welfare, though it is always permissible to maximize utility impartially considered. My reading of Mill resembles consequentialism considerably less than does Scheffler's "hybrid" (which is to say, avowedly nonconsequentialist) theory, as will be illustrated in the following section.

18. The most developed exposition of the varieties of consequentialism comes from Peter Vallentyne who holds, somewhat idiosyncratically, that indirect (for instance,

matter of splitting the difference or adopting the consensus view, though I think it does both. More important, there is a principled reason to adopt the gloss I favor. This reason is given by Rawls, who writes:

> The most natural way, then, of arriving at utilitarianism (although not, of course, the only way of doing so) is to adopt for society as a whole the principle of rational choice for one man. Once this is recognized, the place of the impartial spectator and the emphasis on sympathy in the history of utilitarian thought is readily understood. For it is by the conception of the impartial spectator and the use of sympathetic identification in guiding our imagination that the principle for one man is applied to society.[19]

Although Rawls speaks of utilitarianism, he notes, "the kind of utilitarianism I shall describe here is the strict classical doctrine which receives perhaps its clearest and most accessible formulation in Sidgwick."[20] I will therefore refer to this direct, maximizing, and strictly impartial theory as *orthodox utilitarianism*, on analogy with orthodox consequentialism, whereas I will call the philosophers associated with the historical movement Utilitarians, with a capital 'U'. Though Sidgwick was both an orthodox and a classical Utilitarian, Mill was, in his own words, a utilitarian only "in quite another sense from what perhaps anyone except

---

rule-based) theories are not genuinely consequentialist. Vallentyne characterizes the two fundamental claims of what he calls *core consequentialism* as supervenience and value promotion. See Vallentyne, "Against Maximizing Act Consequentialism," in *Contemporary Debates in Moral Theory*, ed. James Dreier (Oxford: Blackwell, 2006), 21–37. The supervenience thesis (ibid., 22) states: "The permissibility of actions in a given choice situation supervenes on (is fully determined by) the value of their consequences." The value promotion thesis states: "If, in a given choice situation, one action is permissible, and a second is more valuable, then the second action is also permissible." Since neither thesis is compatible with rule-consequentialism, this usage would make my task too easy: I am not merely arguing that Mill was no act consequentialist. But it is worth noting that, on my reading, Mill's moral theory rejects both the supervenience and value promotion theses *for different reasons than do indirect consequentialist theories* (though it rejects them for those reasons as well). Specifically, it rejects both core claims of consequentialism because they fail to distinguish between two equally valuable states of affairs: (the same quantity and quality of) the agent's happiness and the happiness of another.

19. John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 26–27.

20. Rawls, *Theory of Justice*, 22. My only quibble with Rawls here is over the word 'classical', which I would replace with 'orthodox' so as to avoid any misleading historical implications.

myself understands by that word" (letter to Carlyle, no. 95 [1834], *CW* 7:207).[21]

Nevertheless, Rawls is surely correct to note that the metaphor of the impartial spectator has played a crucial role in the history of utilitarianism.[22] Indeed, he highlights exactly the feature of consequentialism on which I want to focus, albeit for a different reason. "The striking feature" of this view, he writes, "is that *it does not matter, except indirectly, how this sum of satisfactions is distributed among individuals* any more than it matters, except indirectly, how one man distributes his satisfactions over time."[23] Another implication of the metaphor, which has lately become more remarkable, is that it entails an agent-neutral conception of value, on which the value of a state of affairs does not differ depending on who evaluates it.[24] Since the spectator's perspective determines the value of a state of affairs, and does so in just the same way for everyone, the metaphor clearly presupposes an agent-neutral conception of value. Indeed, agent neutrality tends to be taken as partly definitive of consequentialism. But I want to focus on another implication of this trope.

Whereas Rawls was concerned primarily with issues of distribution, my interest in the impartial spectator metaphor concerns its implication of a kind of moral symmetry between self and other: everyone's happiness counts in exactly the same way, when it comes to evaluating acts as right and wrong. We might call this commitment *deontic impartiality* because it adopts impartiality as an abstract rule governing the morality of action. We can then differentiate it from a less stringent notion of *axiological impartiality*: the claim that everyone's happiness, if equal in quantity and quality, is equally valuable. Since the metaphorical

21. For discussion of the importance of this letter, and for argument that Mill's conception of himself as an unorthodox utilitarian remained stable throughout his life, see Daniel Jacobson, "J. S. Mill and the Diversity of Utilitarianism," *Philosophers' Imprint* 3 (2003), www.philosophersimprint.org/003002/.

22. See Stephen Darwall, "Hume and the Invention of Utilitarianism," in *Hume and Hume's Connexions*, ed. M. A. Stewart and John Wright (University Park: Pennsylvania State University Press, 1994), 58–82.

23. Rawls, *Theory of Justice*, 26; emphasis added. This point was crucially important for Rawls, of course, because it underwrites his complaint that utilitarianism does not respect the separateness of persons. My purposes are different, though they too focus on the metaphor's implication that it does not matter, from the point of view of morality, whose happiness is promoted or diminished; only the sum (or perhaps the average) of satisfaction matters.

24. Compare Amartya Sen's development of the notion of agent-relative value, especially in his "Evaluator Relativity and Consequential Evaluation," *Philosophy and Public Affairs* 12 (1983): 113–32.

impartial (and perfectly sympathetic) spectator takes on everyone's pleasures and pains as if they were his own, that metaphor entails a deontically impartial conception of morality. By contrast, axiological impartiality only concerns the theory of value. This value theory, and not any particular way of applying it either as decision procedure or criterion of rightness, Mill considered definitive of utilitarianism. He calls the doctrine that "one person's happiness . . . is counted for exactly as much as another's" nothing less than an "explanatory commentary" on the principle of utility (*U*: 5.36).[25]

I propose, then, to adopt the broadest conception of consequentialism compatible with the deontic impartiality and agent neutrality implied by the impartial spectator trope. Specifically, we should understand consequentialism to allow for either direct or indirect theories of the right, and for both subjective versions of the theory, which focus on expected consequences of action, and objective versions, which focus on actual consequences. Finally, we should not assume that consequentialism requires the maximization of value. Instead, we can allow for satisficing accounts on which suboptimal promotion of the good can be morally permissible and greater promotion supererogatory, without undermining the impartial spectator basis for consequentialism. (It is important to note that these distinctions are *not* anachronistic, though I have framed them in modern terms, since in every case there were

25. My implicit claim that Mill rejects deontic impartiality, which would condemn all action that favors the interests of those one cares about over anyone else's, is admittedly contentious. It may be recalled that Mill (*U*: 2.18) writes that "the happiness which forms the utilitarian standard of what is right in conduct, is not the agent's own happiness, but that of all concerned. As between his own happiness and that of others, utilitarianism requires him to be as strictly impartial as a disinterested and benevolent spectator." However, his point here is to differentiate between utilitarianism and egoism. For this purpose one need not embrace a stronger form of impartiality than the axiological version to which Mill certainly subscribes, though of course one does need to count at least some harms and goods of others as contributing to the rightness and wrongness of actions. Obviously Mill accepts this extremely weak claim. Yet he repeatedly denies the stricter demands of deontic impartiality, as in his 1862 letter to Grote (no. 525), where he explicates his argument and intentions in *Utilitarianism*. There (*CW* 15:762) he writes, "people must not be required to sacrifice even their own lesser good to another's greater, where no general rule has given the other a right to the sacrifice." This is not an idiosyncratic remark, as it gets repeated throughout Mill's work. Whereas deontic impartiality requires just such sacrifice, axiological impartiality can allow that one's reasons, even moral reasons, differ depending on whose good is affected by an action—even though everyone's happiness is equally valuable. Note too that Mill (in *U*: 5.9) expressly rejects deontic impartiality "as an obligation of justice."

advocates of both sides of these disputes among Mill's contemporaries.[26]) So long as a satisficing or indirect theory treats all value identically, without regard for whose interests it affects (if anyone's), the theory abides by deontic impartiality, and therefore counts as consequentialist.

Let me clarify the dialectical situation at this point. My adoption of the broadest conception of consequentialism compatible with the impartial spectator metaphor deliberately makes my task more difficult and significant, in two respects. First, the more capacious a notion of consequentialism I adopt, the more interesting becomes my claim that utilitarianism need not be consequentialist. Second, this broad conception circumvents a simpler (and less interesting) argument that Mill was no consequentialist. Mill repeatedly insisted on recognizing a class of supererogatory action, and he held that it is impermissible to violate certain basic rights for the sake of maximizing the good.[27] While it is true that these views are inconsistent with orthodox consequentialism, they are nonetheless compatible with other versions of the theory; therefore, I will not rest my argument that Mill was no consequentialist on these points. Instead, according to the conception I will take as standard, a moral theory is consequentialist just in case it holds that the rightness and wrongness of an action is determined (perhaps indirectly) solely by consequences (actual or expected), evaluated under strict deontic impartiality and in an agent-neutral manner.[28]

26. In particular, the denial of supererogation was the fourteenth of the thirty-nine articles of the Church of England, which Mill rejected throughout his career, defending supererogation against the charge of "Popish laxity" made by the likes of Godwin. Although 'satisficing' is a modern term, any utilitarian view that defends supererogation entails such a view. If I am correct, Mill advocated a satisficing, subjectivist, and indirect (sanction-based) form of utilitarianism: he explicitly endorsed supererogation, claimed that the "natural" or expected consequences of an action determine its morality, and identified wrong action with the blameworthy. Or so I will argue.

27. On these points see respectively Jacobson, "The Diversity of Utilitarianism" and Daniel Jacobson, "Mill on Liberty, Speech, and the Free Society," *Philosophy and Public Affairs* 29 (2000): 276–309. Mill's commitments to rights and to supererogation are obscured by the common tendency to place far too much weight on chapter 2 of *Utilitarianism*, especially its statement of the greatest happiness principle, which I contend was crafted to be just what Mill there claims it is: a creed held in common among Utilitarians rather than any specific version of the theory. Even this famous proportionality statement of the principle conflicts with orthodox consequentialism since it allows for degrees of rightness instead of holding that only the best action is right.

28. Note that I am deliberately avoiding the question of exactly which consequences are relevant: those of the specific act, the "natural" consequences of acts of

This gloss captures what most philosophers mean by 'consequentialism', and it includes not just the orthodox form of the theory, widely held to be its most powerful version, but also the other common alternatives, such as satisficing act-consequentialism and indirect rule-consequentialism. This is what I will mean by the term hereafter. In addition to including all the versions of consequentialism most commonly advanced by its proponents or attacked by its antagonists, my gloss also has an independent motivation, given by Rawls, which is both theoretically grounded and historically important. Now suppose I am right that 'utilitarianism' has a semantic debt to the history of philosophy, such that it must be understood capaciously enough to capture the views of the classical Utilitarians. If their conception of utilitarianism proves more diverse and less orthodox than the consequentialist schema can accommodate, then I will have shown that the textbook definition of utilitarianism obscures an implicit presupposition of consequentialism. Moreover, insofar as this assumption turns out to be both substantive and dubious, this point reveals unexplored avenues for moral theory in the utilitarian style. The best way to bring out this assumption is to consider how a utilitarian might reject consequentialism by holding a teleological moral theory, along with a utilitarian theory of value, while rejecting deontic impartiality—in particular, the claim that it makes no moral difference who is harmed or benefited.[29]

Consider actions whose only evaluative consequences concern the agent's own good: what are often called self-regarding actions.[30] This

---

that type, or even those that would issue from the acceptance of a rule prohibiting or requiring such actions. Moreover, these consequences can include the value of the act itself, and they might be compared to the available alternatives or judged by some noncomparative standard (such as whether the act creates or diminishes happiness). But the crucial point is that whichever consequences count, the impartial spectator metaphor requires that they be considered identically regardless of who is benefited or harmed. This implies both deontic impartiality and agent-neutrality.

29. Notice that rule-consequentialism's commitment to deontic impartiality may be obscured by the fact that the best moral rules are likely to allow for some partiality: perhaps we will be permitted to save our loved ones rather than the Archbishop (to borrow Godwin's famous case). Nevertheless, at the foundational level where moral rules are justified, consequences are assessed under deontic impartiality. Hence, objections to Mill's antipaternalism are not answered simply by attributing a rule-based theory to him because it seems likely that the best set of moral rules will include some paternalistic ones.

30. I am following the standard practice of understanding the consequences of an action in a broad and vague way, so as to include effects that are not caused by the action but follow from it in some looser sense—as my dog's chasing the rabbit is a

term is closely associated with Mill but, unfortunately, its conventional meaning is not quite what Mill meant by it.[31] Therefore I will call such actions *purely self-regarding*; in the following section, where it will be important to differentiate Mill's usage from the conventional one, I will explicate this distinction. As an example of a purely self-regarding but evaluatively significant act, harmful only to the agent, imagine that I decide to hit my thumb with a hammer, causing myself intense and utterly avoidable pain but affecting no one else. The failure to promote my own good does not seem to be a moral failing as such, even when there are no countermanding positive consequences for others. Such bad decisions are *foolish*—when they are truly bad decisions and not just unfortunate outcomes—but not *wrong*. Consequentialism seems to mistake prudence for morality, when considering the purely self-regarding. Moreover, this mistake follows directly from the impartial spectator metaphor, which implies that pains and pleasures count toward the morality of action identically, *regardless of who suffers them*, the agent or another.

According to commonsense morality, by contrast, there is a fundamental asymmetry between self and other. As Michael Slote writes, "over a large range of cases our ordinary thinking about morality assigns no positive value to the well-being or happiness of the moral agent of the sort it clearly assigns to the well-being or happiness of *everyone other than the agent*."[32] Arguably, however, imprudence is not even a pro tanto wrong-making feature of action.[33] This is not merely a freestanding intuition. It can be buttressed by a sentimentalist argument connecting specific evaluative judgments to distinct emotions—as comic judgments about what is funny are connected to amusement rather than

---

consequence of my not commanding her to stay. The value of actions themselves, not just their outcomes, is also standardly included in the consequentialist tally.

31. See Jacobson, "Mill on Liberty."

32. Michael Slote, "Some Advantages of Virtue Ethics," in *Identity, Character, and Morality: Essays in Moral Psychology*, ed. Owen Flanagan and Amelie Rorty (Cambridge, MA: MIT Press, 1990), 441. This asymmetry claim is consistent with acknowledging duties to oneself, such as not to waste one's talents, so long as these duties do not include promoting our own happiness whenever there is no countermanding duty to others.

33. I am here using 'wrong' narrowly, as a distinctively moral term, as (we shall see) does Mill. Sometimes philosophers use it more broadly, as a normative term that can be applied from various evaluative perspectives. Obviously imprudence counts decisively as "wrong-making from a prudential point of view," but that is irrelevant to the point at hand.

disgust. Moral judgments seem tied to different emotions than are prudential judgments: specifically, to guilt rather than regret (in the first-person case). If guilt essentially involves the motive to make reparations, whereas regret involves the motive to change policy, then regret but not guilt will be fittingly felt toward one's purely self-regarding blunders. Although this sentimentalist conception of morality is controversial, it has some influential modern proponents and a significant historical pedigree. Most notably, this was Mill's conception of morality.

It should come as no surprise that moral theories hostile to consequentialism, and commonsense intuitions in tension with it, reject deontic impartiality. More remarkable, though, is the fact that even by the lights of a broadly teleological theory, deontic impartiality seems both optional and controversial. Recall that I mean what Mill meant by a teleological theory, which is only half of what Rawls means by the term. Rawls called views teleological when they hold that "the good is defined independently of the right, and the right is defined as that which maximizes the good."[34] But Rawls's distinction flouts the historical context of this debate by deviating from a different bifurcating distinction between moral theories, drawn by both the classical Utilitarians and their opponents. The importance of this point, beyond its exegetical significance, is that it helps demonstrate what Mill and his contemporaries meant by utilitarianism.

Both Mill and his antagonists routinely differentiated between two schools of ethics. As he explains: "According to one opinion, the principles of morals are evident a priori, requiring nothing to command assent, except that the meaning of their terms be understood. According to the other doctrine, right and wrong . . . are questions of observation and experience" (*U*: 1.3). The first sort of theory was called intuitive, independent, or a priori; the second, inductive, dependent, teleological, or indeed utilitarian. Mill accepted this admittedly coarse-grained dichotomy throughout his career, and he repeatedly defended

34. Rawls, *Theory of Justice*, 24. He then constructs a dichotomy by calling *deontological* all those views that are not in this sense teleological—a class that includes indirect and other nonmaximizing versions of consequentialism, as well as Kantianism, virtue ethics, and more. This seems unlikely to be a perspicuous taxonomy, when one side is so uniform and the other so diverse. Similarly, Peter Vallentyne writes: "Almost all (if not all) authors require that a theory *maximize* the good in order to be teleological." See Vallentyne, "The Teleological/Deontological Distinction," *Journal of Value Inquiry* 21 (1987): 27; emphasis in original. But not all authors so require: John Stuart Mill does not.

utilitarianism against critics from the other school who drew the same distinction. The crucial point here is that even hedonic ethical egoism Mill declared to be "upon the whole on the utilitarian side of the controversy" (letter to Grote, no. 525 [1862], *CW* 15:762). Despite its rejection of axiological impartiality, the cornerstone of utilitarianism, hedonic ethical egoism is both teleological and empirical. It offers what Mill called an "external standard" by which to assess moral judgment: the happiness of the agent.

Hence, a teleological theory can reject both axiological and deontic impartiality since nothing requires such a theory to treat all good and bad consequences equally, much less identically. Some bad consequences might make an action foolish or even shameful rather than—not just in addition to—wrong; indeed, I will argue that this was Mill's view. According to this taxonomy, which was accepted on both sides of the nineteenth-century debate, a teleological theory counts as utilitarian simply by adopting all happiness as the sole intrinsic good: the thesis Mill refers to as the principle of utility.[35] Those teleological theories that deny axiological impartiality and, hence, the principle of utility (such as egoism and asceticism) are not utilitarian, even though Mill took them to be closer to utilitarianism than to its contemporary intuitionist antagonists. Those theories that accept axiological impartiality thereby hold that the agent's happiness is no more or less valuable than anyone else's; however, this does not imply that it gives rise to exactly the same reasons, or the same kind of reasons, as does other people's happiness. Contrary to the impartial spectator trope and the thesis of deontic impartiality, my own good may give me only prudential, not moral, reasons to act.

If I am right to insist that we must understand utilitarianism broadly enough to include the views of the classical Utilitarians, then some forms of utilitarianism will not be consequentialist. Thus the initial "trick" that allowed me to deny that utilitarianism entails consequentialism was actually ground clearing necessary to expose a tacit and substantive assumption of consequentialism, which need not be accepted even by a teleological theory that subscribes to the principle of utility. That assumption is deontic impartiality, the claim that evaluative

35. See D. G. Brown, "What Is Mill's Principle of Utility?" *Canadian Journal of Philosophy* 3 (1973): 33–39. Unfortunately, Mill is not very consistent in his terminology, and he can be careless about differentiating the principle of utility (an axiological claim) from the greatest happiness principle (a moral claim).

consequences count identically toward the moral assessment of action, regardless of who suffers or benefits. The first part of my argument is now secured: I have shown that it is possible for a utilitarian theory to reject consequentialism. In what follows, I contend that this argument has not merely mapped some uninhabited logical space. In fact, John Stuart Mill, the transitional figure between Jeremy Bentham (the first "philosophical utilitarian") and Henry Sidgwick (the great systematic utilitarian) held a nonconsequentialist form of utilitarianism. Or so I will now argue.

II.

In this section of the essay, I turn to the question of whether Mill accepted deontic impartiality and, hence, whether his moral theory was consequentialist. Consider the central problem of Mill interpretation: how to make the principle of liberty consistent with his utilitarian commitments. Mill introduces his principle of liberty in the most uncompromising terms, with a forceful rejection of paternalism that seems in obvious tension with the injunction of orthodox utilitarianism to maximize net happiness, impartially considered. "Over himself, over his own body and mind, the individual is sovereign" (*OL*: 1.9), Mill insists; yet many paternalistic rules seem justified according to orthodox utilitarianism. Laws requiring motorists to wear seat belts, for instance, would compel people for their own good, and such laws are probably optimific. But Mill does not pause to consider specifically the most defensible paternalistic laws; he seems to think he has a general argument with strong antipaternalist implications. Otherwise how could he (ibid.) claim, so peremptorily, that a person's "own good, either physical or moral, is not a sufficient warrant" for compelling him to act or to forbear from acting?

Although the currently predominant view attributes an orthodox consequentialist moral theory to Mill, its advocates are hard-pressed to reconcile their interpretation with what Mill actually says. Roger Crisp is exemplary for facing up to these difficulties most forthrightly. Crisp admits that his sophisticated, multilevel act-utilitarian interpretation ultimately entails a dissimulation hypothesis, on which many of Mill's overt claims are held to be misrepresented, insincere, or at least greatly exaggerated. These include such central doctrines as Mill's antipaternalism and uncompromising defense of individual liberty, his embrace of supererogation, and his claim that moral rules issue in genuine obligations. For example, Crisp simply denies that the principle of liberty

gives anything like a fundamental limit to justifiable social interference with the liberty of the individual, as Mill asserts. He claims to the contrary that: "To put it bluntly, if social interference will maximize welfare overall, then that legitimizes the interference, even if it might appear to be an encroachment on the self-regarding sphere."[36] Notwithstanding Mill's explicit claims about rights, obligations, and the limits of morality, Crisp holds that "when he was engaged in doing serious moral philosophy, that is, in making claims about what *really* makes actions right or wrong," Mill embraces the orthodox consequentialist form of utilitarianism according to which: "Actions are right or wrong solely in so far as they promote happiness or unhappiness."[37]

The burden on this interpretation becomes most telling when Mill confronts the still influential objection that utilitarianism makes morality wildly overdemanding, requiring us to treat our commitments and relationships as merely opportunities to do good that must be forsaken whenever there is more good to be done elsewhere.[38] Mill responds by denying that utilitarianism makes anything like such severe demands on ordinary people who have no special duties or exceptional powers to affect public utility. Crisp's commentary is revealing. He writes:

> Utilitarianism is almost certainly much more demanding than Mill allows. It is tempting to think, in fact, that Mill is deliberately being disingenuous here. . . . Better to persuade a reader to become a feeble utilitarian than put them off entirely by stressing the demandingness of utilitarian morality.[39]

No doubt Crisp is correct that the orthodox consequentialism he attributes to Mill has these implications. Moreover, the proponent of an esoteric morality might think such insincerity justified. Yet this

36. Crisp, *Mill on Utilitarianism*, 185.

37. Ibid., 111–12. Like other advocates of the standard view, Crisp focuses on the proportionality statement of the greatest happiness principle (given in *U*: 2.2) and shoehorns the rest of Mill to fit its allegedly consequentialist shape. He thus concludes that the principle of liberty "cannot ground any kind of liberalism in Mill's thought which is inconsistent with his act utilitarianism" (ibid., 175).

38. This complaint gets its most influential modern development in Bernard Williams, "Against Utilitarianism," in *Utilitarianism: For and Against*, ed. J. J. C. Smart and Bernard Williams (Cambridge: Cambridge University Press, 1973); and Michael Stocker, "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy* 73 (1976): 453–66.

39. Crisp, *Mill on Utilitarianism*, 115.

conventional reading does so much violence to what Mill actually wrote that it is almost fair to wonder if its advocates are actually *reading* Mill, as opposed to projecting their own anachronistic consequentialist theory upon him. Not only did Mill never broach the idea of an esoteric morality, he held the virtue of sincerity in such high regard that he chose to jeopardize his political prospects rather than dissemble about his unpopular opinions.[40]

I say "almost fair" because it must be granted that chapter 2 of *Utilitarianism*, in particular, contains passages that lend themselves to this conventional reading (along with others that are inconsistent with it). Yet there are powerful reasons to doubt the centrality of that work and to eschew placing too much emphasis on the proportionality statement of the greatest happiness principle (GHP), which is the linchpin of the standard interpretation. In fact, there is an abundance of evidence that Mill did not place on this work anything like the importance that canonization has placed upon it. Nor did he intend the proportionality formulation of GHP—on which "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness" (*U*: 2.2)—to be the official statement of his own moral theory. Rather, he puts this forward as a vague and equivocally stated "creed," acceptable to all Utilitarians because it merely "denotes the recognition of utility as a standard, not any particular way of applying it" (*U*: 2.1n).[41] In a sense, this can be seen as another respect in which current interpretations of Mill suffer from anachronism. Although his ambitions for what he called this "little work" (*Autobiography*, *CW* 1:265)

40. Mill set as a precondition of his running for parliament that he not be required to answer questions concerning his religious beliefs (or lack thereof). Indeed, the "rule of veracity" was perhaps foremost of the many duties and virtues that, in Mill's view, have such significant indirect consequences as to almost always preclude breaching them in specific cases. As he wrote to Henry Brandreth: "The duty of truth as a positive duty is also to be considered on the ground of whether more good or harm would follow to mankind in general if it were generally disregarded and not merely whether good or harm would follow in a particular case" (letter to Brandreth, no. 1028 [1867], *CW* 16:1234).

41. But compare Crisp (*Mill on Utilitarianism*, 7–8), who writes: "Insofar as Mill was an evangelist, *Utilitarianism* . . . can be seen as his bible. Though it was not written in the high and polished style of *On Liberty* or *The Subjection of Women*, it was clearly intended to be the summation, and defense, of his thoughts on the doctrine which provided the foundation for his views in other areas." For argument that this is exactly the wrong way to approach Mill and *Utilitarianism*, see Jacobson, "Diversity of Utilitarianism."

were decidedly modest, its historical impact has proven immense. This issue is of the utmost importance for Mill interpretation.

I have argued against the standard interpretation elsewhere and cannot attempt a full account of Mill's moral and political philosophy here. Rather than defend my reading against the consequentialist alternative, or canvass problematic passages for each, I propose to take a simpler approach. I will start with the central tenets of *On Liberty*—in particular, Mill's defense of a sphere of self-regarding action and the principle of liberty itself—and consider what sort of utilitarian moral theory can accommodate them. Then I will marshal independent evidence that Mill held such an unorthodox form of utilitarianism, which rejects the demand of deontic impartiality that everyone's happiness must be treated identically in determining the morality of an action. I will not be defending the principle of liberty here but arguing that, despite its tension with consequentialism, Mill's strict antipaternalism and his defense of a substantial sphere of individual liberty cohere neatly with his conception of morality and the limits of obligation.

Recall our preliminary discussion of purely self-regarding action, which focused on acts that have no effect, or no harmful effect, on anyone but the agent. I claimed there, without argument, that Mill had a substantially broader conception of self-regarding action than this conventional understanding allows. Mill clearly states that by self-regarding action he means those acts that primarily *concern* only the agent; and he allows that some self-regarding acts will *affect* others negatively, even in ways to which they would not consent. It is worth quoting the most important passage on this point, from *On Liberty* (4.10); Mill writes:

> I fully admit that the mischief which a person does to himself may seriously affect, both through their sympathies and their interests, those nearly connected with him, and in a minor degree, society at large. When, by conduct of this sort, a person is led to violate a distinct and assignable obligation to any other person or persons, the case is taken out of the self-regarding class, and becomes amenable to moral disapprobation in the proper sense of the term.

This passage, which gives Mill's most developed exposition of self-regarding action, should be taken as his official view of the subject. It allows him to grant the obvious point that even his paradigms of self-regarding action, such as the expression of opinion and sentiment, can harm people both directly (by hurting their feelings) and indirectly (by setting back their interests). But although the opinion that corn dealers

are starvers of the poor, for instance, might prove detrimental to their interests, its expression cannot legitimately be punished except in contexts where it constitutes incitement to riot.[42]

Two crucial points follow, the first of which does not depend on any contentious interpretive issue. Mill clearly implies that when a person "does mischief" to himself, his action is *not yet amenable to moral disapprobation*, though both agent and act can be criticized in other terms—as selfish, intemperate, or foolish. This claim directly contradicts the thesis of deontic impartiality by drawing a fundamental self/other asymmetry with regard to the moral relevance of happiness, thereby confuting the impartial spectator metaphor.[43] Hence Mill's advocacy of a sphere of liberty, within which the individual is immune from moral disapprobation and other forms of social coercion, conflicts intractably with consequentialism by treating the agent's interests differently than the interests of others. Moreover, as the passage quoted above strongly suggests, Mill had a considerably broader conception of self-regarding action than just those acts that affect no one but the agent. He thus places a much larger class of action beyond the pale of morality, in the sense that these actions are not apt for moral disapprobation (when things go badly for the agent) or specifically moral approval (when things go well, and the agent maximizes net happiness by improving his or her own lot).[44]

42. This example is developed at *OL*: 3.1. Compare *OL*: 2.1n, where Mill argues that "there ought to exist the fullest liberty of professing and discussing, as a matter of ethical conviction, any doctrine, however immoral it may be considered." See also *OL*: 2.11, where Mill claims that the "pernicious consequences" of an opinion do not justify its repression. See Jacobson, "Mill on Liberty," for more discussion of this point.

43. That is not to say that it violates every standard of impartiality. This view is compatible with Mill's axiological impartiality (on which everyone's happiness counts equally) and with his insistence on the reciprocal nature of rights and obligations (which grants no one any special rights or more exigent obligations). However, since the metaphor of the perfectly sympathetic and impartial spectator is designed to prescind from any consideration of who is benefited or harmed by an action—as Rawls puts it, the trope applies the principle of rationality for one person to society as a whole—it cannot rule out the harms an agent brings upon himself as morally irrelevant. I am grateful to Julia Driver for pressing this issue in discussion.

44. I should emphasize that it is the action (of hitting my thumb with a hammer, say), not the pain caused by it, that seems morally irrelevant according to common sense and that is claimed by Mill to be beyond the pale of moral assessment. This point is best illustrated by contrasting the agent's perspective with that of another person. As my friend, you might have a duty to try talking me out of this foolish behavior; if so, then the pain I propose to inflict upon myself for no reason is relevant to the morality of *your* action or inaction. The crucial claim here is not that my pain somehow doesn't count, as an axiological matter—it does. It makes the world worse than it would be

This latter point will be explicated when we consider Mill's account of supererogatory action.

Rather than attempting to shoehorn what Mill says into a preconception of his moral theory, I propose to take seriously this passage and others equally inconsistent with consequentialism. Mill claims that only when an action that would otherwise be self-regarding violates an obligation, because of the specific circumstances in which it is performed, does it become amenable to moral disapprobation. Note first that if Mill thought we had a general obligation to maximize net happiness, this claim would make little sense. According to the orthodox position, we do wrong whenever we fail to act for the best. Here Mill is considering actions, such as intemperate drinking or gambling, which are not generally optimific; yet he claims that these are, ordinarily, self-regarding actions despite their tendency to be self-destructive and even harmful to nonconsenting others. Crucially, though, these actions harm nonconsenting others only in those respects for which "society admits no right . . . to immunity from this kind of suffering" (*OL*: 5.3). The interpretive challenge this passage poses should be clear. What makes self-regarding action not amenable to moral disapprobation even when it fails to maximize happiness? The answer lies in Mill's account of the meaning of moral terms, which ties them to specific sentiments.

Consider what Mill says about right and wrong—not the GHP given in chapter 2 of *Utilitarianism*, but the metaethical account given in chapter 5. There (5.14) he writes:

> I think there is no doubt that this distinction lies at the bottom of the notions of right and wrong; that we call any conduct wrong, or employ, instead, some other term of dislike or disparagement, according as we think that the person ought, or ought not, to be punished for it; and we say that it would be right to do so and so, or merely that it would be desirable or laudable, according as we would wish to see the person whom it concerns, compelled, or only persuaded and exhorted, to act in that manner.

According to Mill, then, wrong acts are by definition *punishable* or blameworthy. He uses those two terms synonymously because he understands punishment to include not just legal sanction of the agent, or even the

---

otherwise, and this may be relevant to moral judgment of *other* actions. But this is entirely compatible with the denial that it is wrong for me to hit my own thumb. I am grateful to David Boonin for pressing me to clarify my view on this point.

external sanction of others (expressed in such blaming sentiments as outrage), but also "the reproaches of his own conscience" (*U*: 5.14): that is, guilt. Analogously, when Mill calls an action right, obligatory, or a duty, he means that it is *compulsory*.

Mill's metaethical position is commonly misunderstood on just this point, due in part to his loose talk of what "ought to be punished" and when "we would wish to see [someone] compelled," both of which sound like verdictive (all-in) judgments. The common misreading commits two symmetrical errors: it makes light of his claim that some actions—the supererogatory ones—deserve praise and admiration but are not compulsory, and it misconstrues when actions deserve punishment. It thereby mistakes Mill's view of both right *and* wrong. Let's start by considering wrongness. The conventional reading attributes to Mill the view that an act is punishable, and therefore wrong, whenever punishing it has good (or perhaps best) consequences.[45] This position resembles the so-called Utilitarian Theory of Punishment, which only looks forward at the consequences of punishment, not backward at whether punishment is deserved. Since Mill considers the blame of others and even the self-reproach of the agent's own conscience to be forms of punishment, this implies (roughly) that an act is punishable whenever it would be optimific for an agent to feel guilty over doing it.[46] But that cannot be Mill's view.

In the first place, when Mill says, too casually, that wrong acts ought to be punished, he cannot be expressing an all-things-considered

45. This interpretation is put most clearly by David Brink, who holds Mill to claim "that an action is wrong just in case some kind of external or internal sanction attached to it (punishment, blame, or self-reproach) would have good—perhaps optimific—consequences." David Brink, "Mill's Deliberative Utilitarianism," *Philosophy and Public Affairs* 21 (1992): 69. See also John Gray, *Mill on Liberty: A Defense* (London: Routledge, 1983), 31; and Crisp, *Mill on Utilitarianism*, 129. All these authors understand Mill to hold that wrong acts are those it would be optimific to punish.

46. Guilt is just one of the three kinds of punishment Mill discusses—the other two being law and the blame of others—but it is the one least likely to have unintended negative consequences. Hence, Crisp (*Mill on Utilitarianism*, 129) writes: "there is no imaginable case of an agent's failing to maximize happiness to which Mill would be forced to retract any attribution of wrongness. For he can always claim that the nonmaximizing agent should be punished by the reproaches of their conscience." But this is not what Mill actually claims, since he never suggests that an agent who fails to perform a supererogatory act should feel guilty about the omission; on the contrary, *this is just what he denies* by denying that the agent acts wrongly. The conventional reading thus combines two errors by conjoining a maximizing act-consequentialist theory of rightness with the forward-looking theory of punishment.

judgment. Mill explicitly allows that *some punishable acts should not actually be punished* because in the specific circumstances legal sanction, blame, or even guilt would be inexpedient. In such cases, he (*U*: 5.14) writes: "Reasons of prudence, or the interest of other people, may militate against actually exacting [someone's duty]; but the person himself, it is clearly understood, would not be entitled to complain." One who shirks one's duty *deserves* punishment, Mill (ibid.) claims—which is why that person would have no complaint about being compelled before the fact or punished afterward—but it is a distinct question whether, all things considered, such coercion should be employed. People are punishable not whenever it's expedient to punish them, but only when they are "the proper objects of punishment." Doing wrong makes one eligible for punishment, but considerations of expediency also count toward determining when punishable agents should in fact be punished. Then they should be punished by whatever means is optimific, whether legal sanction, the blame of others, self-reproach, or all of the above.[47]

This is as far as Mill's discussion goes, which is far enough to undermine the conventional interpretation. Since Mill does not explain just what are the proper objects of guilt and blame, we must move somewhat beyond the text at this point; but I have some good company in doing so—most notably David Lyons.[48] As I see it, Mill holds the quintessentially sentimentalist thesis that an act is wrong whenever guilt over it would be *fitting* from the agent, and resentment fitting from others.[49] This proposal raises some inevitable questions: What makes guilt fitting, when it is; and how is such talk of the fittingness of emotions

47. On this point see Alan Ryan, "John Stuart Mill's Art of Living," in *J. S. Mill: On Liberty in Focus*, ed. John Gray and G. W. Smith (London: Routledge, 1991), 162–68.

48. Lyons writes, "Mill seems to be saying that wrong acts are those for which guilt feelings are appropriate" and "these [feelings] are appropriate only when corresponding informal social rules could be justified." See Lyons, "Mill's Theory of Morality," in *Rights, Welfare, and Mill's Moral Theory*, 53, 57. See also Walter Sinnott-Armstrong, "You Ought to Be Ashamed of Yourself (When You Violate an Imperfect Moral Obligation)," *Philosophical Issues* 15 (2005): 194: "Mill's considered view seems to be that people who violate obligations are *liable* to punishment in the sense that they themselves are not wronged or entitled to complain if they are punished (to an appropriate degree)." The crucial point is that some notion of emotional fittingness or appropriateness must be adduced, which is necessary but not sufficient for justifying punishment.

49. In Mill's view, anger is a proto-moral emotion, which must be refined into resentment in order to be fitting only at moral transgressions rather than at any disagreeable action, indiscriminately. Thus, "a person whose resentment is really a moral feeling . . . considers whether an act is blameable before he allows himself to resent it" (*U*: 5.22).

compatible with Mill's utilitarian and liberal commitments? Since perhaps the three most obvious answers do not work, I will begin by discarding them. First, one might say that guilt is fitting whenever you would in fact feel it. This proposal makes guilt self-ratifying, which conflicts with Mill's repeated insistence on an "external standard" for justifying moral emotions and intuitions. Second, one might say that guilt is fitting whenever you have done wrong; but this would be viciously circular since we are trying to explicate wrongness. Finally, one might say that guilt is fitting whenever it's best to feel it. But this traduces the distinction between fitting and optimific emotions, and it collapses the distinction between the conventional position and my own. So much for the bad answers. Can we do any better?

Let's start with an obviously limited claim, in hopes of developing a more general schema. Surely guilt is fitting when someone commits a murder (for instance). But what justifies even this weak claim? The first thing to note is that Mill differentiates between the *moral* aspect of an action and its other evaluative aspects: both its *prudence* and its *aesthetic* qualities, understood as the act's beauty, nobility, or lovableness.[50] "The morality of an action depends on its foreseeable consequences," as does its prudence, according to Mill, whereas "its beauty, and its lovableness, or the reverse, depend on the qualities which it is evidence of" ("Bentham," *CW*10:112). Although the moral and prudential evaluation of acts depends upon their foreseeable consequences, these other forms of evaluation, which Mill loosely calls aesthetic, are directed instead at the character of an agent as it is revealed in action. Indeed, this difference lies at the heart of the distinction between guilt and resentment, on one hand, and shame and contempt, on the other.[51] I will come back to this point about the difference between moral and nonmoral sentiments presently.

Opponents of utilitarianism frequently complain that so many of the consequences of action are unforeseeable that the theory provides little guidance. Mill responds by saying that this complaint proves too

50. Mill drew a tripartite distinction between the spheres of value throughout his mature work, from "Bentham" (1838) to *A System of Logic* (1843) to *Utilitarianism* (1861). Unfortunately he draws this distinction in slightly different ways each time. I have finessed these differences here; the important point is that in every case morality is understood narrowly, as concerning whether an action is right or wrong, which is just one of its three evaluative aspects.

51. See June Tangney and Ronda Dearing, *Shame and Guilt* (New York: Guilford Press, 2002).

much since it would also tell against ordinary prudence, which everyone accepts. Thus, he writes:

> Whether morality is or is not a question of consequences, [one] cannot deny that prudence is; and if there is such a thing as prudence, it is because the consequences of actions can be calculated. Prudence, indeed, depends on a calculation of individual actions, while for the establishment of moral rules it is only necessary to calculate the consequences of classes of action—a much easier matter. ("Whewell on Moral Philosophy," *CW* 10:180)

It is obviously a good idea to have a general rule prohibiting murder because that class of action has foreseeably bad consequences, and often enough murder can be deterred by sanctions without undue costs. Such a rule both obligates people not to commit murder and gives them the right not to be murdered. Therefore a moral rule can be justified establishing the fittingness of guilt (from the agent) and resentment (from others) over murder—though surely acts of murder will be best prevented by legal sanction.[52]

This schema justifies certain moral rules, both informal social norms and expressly posited laws, by the utility of their acceptance. These moral rules, in turn, determine when guilt and resentment are fitting. Whenever a moral rule has been broken, the agent has done wrong and is punishable—though it may or may not be expedient actually to punish him or her, depending on contingencies of the circumstance. Although the justified moral rules will vary across times and cultures, this is not a form of relativism, in that it does not ratify the status quo morality. The fact that a moral rule has been adopted by some society does not suffice to justify it. On the contrary, Mill was keen to criticize many actual social rules for lacking utilitarian justification. As he puts it, "The contest between the morality which appeals to an external standard, and that which grounds itself on internal conviction, is the contest of progressive morality against stationary—of reason and argument against mere opinion and habit" (ibid.,179). Whence Mill's account of the fittingness

52. This view bears an obvious resemblance to the moral theory developed by Allan Gibbard, who expressly cites Mill as an influence. See Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), 41. It is the sentimentalist aspect of Gibbard's view, not the expressivism, which is crucial here. For discussion of how sentimentalism can cohere with various metaethical theories, see Justin D'Arms and Daniel Jacobson, "Sensibility Theory and Projectivism," in *The Oxford Handbook of Ethical Theory*, ed. David Copp (Oxford: Oxford University Press, 2006), 186–218.

of the moral sentiments: guilt and resentment are fitting over the violations of moral rules proscribing classes of action, justified on the basis of the utility of their acceptance. A crucial requirement of rule acceptance is that it must engage one's moral emotions; to accept the rule against theft is to be disposed to feel guilty about committing theft and to resent others who steal, as well as to deem such responses fitting.

Although Mill's sanction-based moral theory is a form of indirect utilitarianism, his sentimentalist metaethics differentiates the view from ordinary rule-utilitarianism because the moral sentiments distinguish the moral realm from the prudential and the aesthetic. In short (and too crudely): things we cannot feel guilty about doing, or resent other people for doing, cannot be wrong—though they may be amenable to other forms of criticism.[53] In order for a norm for the fittingness of guilt to effectively regulate people's guilt responses, it must answer to that emotion's characteristic concern, to what it is about. Thus one feature of the acceptance-utility of a norm for the fittingness of guilt, absent in moral rules not grounded in the sentiments, is that norms for guilt must answer to the inherent constraints of the emotion. Guilt serves as a discrete motivational system that issues in the motivation to make reparations to the wronged party.[54] This point exposes the psychological underpinnings of the commonsense verdict, considered earlier, that harms the agent does to himself or herself merit not guilt (from the agent) and resentment (from others), but other emotions—depending on the case, perhaps either pity or contempt. When I harm myself for no good reason, I should feel no temptation to make reparations to myself.

The point is not just that we won't in fact feel guilty over harming ourselves, though that is true and important; moreover, we do not endorse feeling guilty about it. When we act foolishly, regret, which motivates policy change, is fitting; not guilt, which motivates making amends. Something similar holds for resentment, which, as a species of anger, motivates retaliation. We feel no temptation to retaliate against

53. While actual dispositions to guilt and resentment do not settle what is wrong, the essential tie to the sentiments constrains moral judgment to those actions we can feel guilty about. A similar treatment of nonmoral value is discussed in Justin D'Arms and Daniel Jacobson, "Anthropocentric Constraints on Human Value," *Oxford Studies in Metaethics* 1 (2006): 99–126.

54. For an account of the sentiments as discrete motivational syndromes, which do not already involve the evaluative concepts sentimentalists seek to explicate, see Justin D'Arms and Daniel Jacobson, "The Significance of Recalcitrant Emotions (or, Anti-Quasijudgmentalism)," *Philosophy* 52 (2003): 127–45. On guilt in particular, see Gibbard, *Wise Choices, Apt Feelings*; Tangney and Dearing, *Shame and Guilt*.

someone who foolishly hurts himself or herself, such as the drunk with no dependents or creditors—the person whose actions Mill says are not amenable to moral disapprobation. Punishment is not then in order, whether or not negative sanctions would be beneficial to the foolish agent. When you harm only yourself, we are moved neither to retaliate against you nor to make reparations to you, even if by harming yourself you bring about less than the best available (and expected) consequences, impartially considered. The crucial point is that plausible norms for the fittingness of guilt and resentment are constrained by the nature of those sentiments, which are not about self-inflicted harms or other failures to respect yourself. Thus a sentimentalist account of wrongness, such as Mill's, lends itself to the self/other asymmetry embraced by commonsense morality and contradicted by the deontic impartiality of consequentialism—even indirect consequentialism. As we shall see, this point helps explain the peremptoriness of Mill's antipaternalism.

First, though, consider Mill's analogous view of rightness and compulsion. The concepts of the obligatory and the punishable are connected, of course, since the principal way we compel actions is by threatening to punish their omission. Hence, right actions are compulsory *in principle*, whether or not the specific circumstances of the case justify such compulsion. Just as not all harmful actions are punishable, on Mill's view, not all beneficial actions are obligatory. He explicitly states the implication of his claim that some "desirable and laudable" acts cannot properly be compelled, as follows (*U*: 5.14):

> There are other things, on the contrary, which we wish that people should do, which we like or admire them for doing, perhaps dislike or despise them for not doing, but yet admit that they are not bound to do; it is not a case of moral obligation; we do not blame them, that is, we do not think that they are proper objects of punishment.

Since such actions are not moral obligations, Mill holds that they are not proper objects of compulsion and punishment. Someone who was compelled to perform such an act, or punished for not doing so, *would* have a complaint against society, unlike the wrongdoer. Hence, these actions are genuinely supererogatory, not merely impractical to compel.[55]

55. Although it is often overlooked, there is abundant evidence of Mill's commitment to supererogation, especially in *Sedgwick's Discourse* (1833), *Auguste Compte and Positivism* (1865), and his letter to Henry Brandreth (no. 1029 [1867]).

At this point someone might want to reject Mill's narrow conception of morality, which limits the moral evaluation of action to questions of right and wrong, obligation and compulsion. Thus Sinnott-Armstrong objects that "morality [also] includes what is ideal and good but not a duty or obligation."[56] Since my primary purpose here is to argue that Mill's theory of right and wrong conflicts intractably with consequentialism, I need not dispute the issue of how broadly to construe morality. Nevertheless, it is worth noting that Mill anticipated the distinction between a narrow and a broad notion of morality, which would later be drawn in similar terms by philosophers such as Allan Gibbard and Bernard Williams.[57] In Mill's view, there are two "co-equal" parts of morality: along with the narrow part concerning the regulation of external action, which has been our focus here, there is a broader part concerning "self-education, the training, by the human being himself, of his affections and will" ("Bentham," *CW* 10:98). There is a decidedly perfectionist strain to Mill's thought about the broader ethical questions of how to live, when narrowly moral issues of obligation are not at play; though, as with all practical concerns, his perfectionism is *ultimately* grounded in the principle of utility. Nevertheless, in considering both supererogatory and ignoble actions, Mill looks not to their consequences so much as to what they show about the character of the agent.

We cannot adequately understand Mill's ethics without keeping in mind what he called the three departments of the Art of Life: morality, prudence, and aesthetics. These three evaluative spheres concern respectively the right, the expedient, and the beautiful or noble in human conduct and character. This idiosyncratic distinction matters because Mill's treatment of the supererogatory parallels what he says about self-regarding character flaws. Recall Mill's claim that the beauty, lovableness, and nobility of an action—as well as their negative counterparts—"depend on the qualities which [the act] is evidence of"

56. Sinnott-Armstrong, "You Ought to Be Ashamed of Yourself," 194. My difference with Sinnott-Armstrong is superficial and terminological; on the fundamental sentimentalist point we agree.

57. Thus Gibbard (*Wise Choices, Apt Feelings*, 6) writes: "We can understand the term ['morality'] broadly or narrowly. Broadly the moral question is how to live. Narrowly, we might try saying, morality concerns moral sentiments: the sentiments of guilt and resentment and their variants." See also Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press, 1985), who contrasts narrow morality with broad ethics—also construed as concerning the fundamental question of how to live.

("Bentham," *CW* 10:112). Thus action that is neither right nor wrong, but which manifests virtue or vice, belongs to the aesthetic sphere, whose characteristic sentiments (such as shame and pride, contempt and admiration) are directed not directly at action but at the character manifested in it. Hence, Mill's treatment of supererogation differs from even the (unorthodox, satisficing) consequentialist account precisely because it matters, for Mill, who gets benefited or harmed. When an action X has better consequences than Y because of its benefits to the agent, someone who chooses to sacrifice his own interests (by doing Y) for the good of others—contrary to the common good, impartially calculated, *which includes his own good*—may be more admirable than one who maximizes net happiness by doing X. Indeed, the person who does Y *will* be more admirable whenever the motivation for that action indicates a virtuous disposition.[58]

The drunk who does not act wrongly because he violates no obligation, and the philanthropist who does more than duty requires, typically are fitting objects of distinctive negative and positive sentiments: specifically, those emotions that do *not* focus directly on action (as do guilt and resentment) but on the qualities of the agent manifest in action (as do shame and contempt). We feel guilty over what we've done; we feel ashamed of who we are. Compare what Mill (*OL*: 4.5) says about self-regarding vices, which are unpleasant but not punishable:

> There is a degree of folly, and a degree of what may be called . . . lowness or depravation of taste, which, though it cannot justify doing harm to the person who manifests it, renders him necessarily and properly a subject of distaste, or, in extreme cases, even of contempt. . . . [B]ut he suffers these penalties only in so far as they are the natural, and, as it were, the spontaneous consequences of the faults themselves, not because they are purposely inflicted on him for the sake of punishment.

Analogously, although supererogatory action by definition creates a better state of affairs than can be compelled, I contend that for Mill such action not only fails to be obligatory, but is praiseworthy only insofar as

58. It might be recalled that Mill (*U*: 2.17) writes, "A sacrifice which does not increase, or tend to increase, the sum total of happiness, [utilitarian morality] considers as wasted." But the case I consider above, where the sacrifice increases the happiness of others but not the sum total of all concerned (including the agent), is just the sort of peculiar example Mill simply did not bother with in this brief treatise. Moreover, he (ibid.) immediately continues by making a statement strictly compatible with my claim: "The only self-renunciation which it applauds, is devotion to the happiness, or to some of the means of happiness, of others."

it manifests a virtuous character.[59] Thus Mill's account of supererogation is also sensitive to who is benefited, in a manner incompatible with deontic impartiality but in line with commonsense morality.

The main point of my argument is not that Mill rejects the maximizing demands of orthodox consequentialism—though this is true, important, and insufficiently acknowledged. The point is rather that, according to Mill's metaethics, an act isn't wrong unless it is punishable; and it isn't right unless it is compulsory. This is crucial because Mill states in the clearest possible terms both that self-regarding action cannot properly be punished and that a sound-minded adult cannot properly be compelled for his or her own good. This, of course, is what Mill (*OL*: 1.9) famously refers to as the principle of liberty:

> The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him or visiting him with any evil in case he do otherwise.

In short, the good of promoting the agent's own happiness does not justify compulsion, and the bad of causing the agent's unhappiness does not give others grounds to inflict punishment. Prudent action is not right action, even if it maximizes utility; and self-destructive action is not wrong action, even if there is an available alternative with better consequences.

Mill's contemporaries, most notably Sidgwick, appreciated the problem that consequentialism poses for antipaternalism and the principle of liberty more generally.[60] From a strictly impartial perspective, it is unclear how Mill can be so peremptory about his antipaternalism,

59. Note that the fittingness of admiration for the philanthropist depends crucially on his motives for acting, whereas the moral quality of the act depends merely on what he (intentionally) does (see *U*: 2.15n). If the philanthropist's true aim is self-glorification, that does not taint the value of the act, but it does affect how it's fitting to feel about him. Similarly, the drunk might be more pitiable than contemptible, depending on the circumstances.

60. See Sidgwick, *Methods of Ethics*, esp. 478. Sidgwick calls the view utilitarianism rather than consequentialism, of course, because the twentieth-century term had not yet been coined.

when it seems that any kind of consequentialist would have to be sensitive to the effects of one's self-destructive actions on others. Indeed, it seems very likely that some paternalistically motivated laws, such as those requiring the use of seat belts, will be optimific. However, Mill's principles follow much more clearly if moral rules—those concerning the fittingness of guilt and anger—are determined by a utility calculus that does not include the effects of self-regarding action on the agent himself (nor those remote effects that might be brought about by the agent's example, which could lead others, by their own agency, to harm themselves). Whereas rule-consequentialism determines the best moral rules by considering all the consequences of their acceptance impartially, Mill's sentimentalist utilitarianism is sensitive to who is harmed or benefited. This sensitivity follows from the nature of the moral sentiments themselves, namely, guilt and resentment, which are inherently partial and asymmetric.

Mill's discussion of self-regarding action and the principle of liberty, in *On Liberty*, coheres with the metaethical account of right and wrong given in *Utilitarianism*. In both discussions, Mill maintains a fundamental asymmetry between self and other, corresponding to the distinction between the spheres of morality and prudence. In so doing, Mill follows the view of commonsense morality that it is foolish but not wrong to act self-destructively, and prudent but not obligatory to maximize one's own happiness *even when that would be optimific*. Thus Mill flouts the strict form of impartiality entailed by the metaphor of the impartial spectator, on which it makes no moral difference whose happiness or unhappiness is affected by an act. In Mill's view, morality does not treat everyone's happiness in exactly the same way (as deontic impartiality demands) even though everyone's happiness is of equal value (as axiological impartiality requires). Mill expressly rejects deontic impartiality by claiming that self-regarding but harmful acts are not amenable to moral disapprobation and that we cannot be compelled for our own good.

Any moral view that treats the agent's interests differently from others, in determining the rightness or wrongness of an action, does not adopt strict, deontic impartiality. And any view that rejects deontic impartiality is not consequentialist in the standard sense, widely adopted by philosophers and motivated in the first section of this essay. Yet Mill was a classical Utilitarian, not just because he was called a utilitarian by his contemporaries and identified himself as one, but because his moral theory is teleological and accepts the principle of utility as its

axiology. Although my interpretation of Mill's view makes him a highly unorthodox utilitarian, this is in keeping with Mill's self-description, quoted earlier. Moreover, it agrees with the assessment of Mill's contemporaries such as John Grote (who termed Mill a neo-utilitarian) and Henry Sidgwick (who called Mill a "conservative utilitarian" for holding that moral rules issue in genuine obligations).[61] Hence, there is no paradox involved in claiming that there is logical space for a utilitarian theory that rejects consequentialism, and there is considerable evidence for ascribing such a view to that most renowned, though not most orthodox, utilitarian, John Stuart Mill.

61. See Henry Sidgwick, *Essays on Ethics and Method*, ed. Marcus Singer (Oxford: Clarendon, 2000), 174.