

Sanctions in Bentham, Mill, and Sidgwick¹

ROGER CRISP

ROSS HARRISON DESERVES THE THANKS of Sidgwick scholars not only for organising this centennial conference, but for his stimulating enquiry into Sidgwick's views on an issue which was considered of great importance by the classical utilitarians of the eighteenth and nineteenth centuries, but which has received little direct attention during this century—the sanctions of morality. What I think is particularly instructive in his discussion is its demonstration that enquiry into the nature of sanctions can illuminate other more well-trodden areas, such as the dualism of practical reason and Sidgwick's attitude to common-sense morality.

It seems churlish to criticise, but such is the way of philosophy; so let me try to niggle a little. First, I want to suggest that Mill was more of a precursor of Sidgwick, and less a follower of Bentham, than Ross suggests. Mill is not a psychological hedonist, nor indeed a psychological egoist of any stripe. He does think that all desire is for the perceived greatest balance of the agent's pleasure over pain, but for Mill 'desire' is a technical term reserved for that particular—admittedly very common—motivation. He allows for weakness of will: 'Men often, from infirmity of character, make their election for the nearer

¹ References to Mill's *Utilitarianism* give chapter and paragraph number; those to Sidgwick's *The Methods of Ethics* are to the seventh edition; and the reference to the *Memoir* is A. and E. M. Sidgwick, *Henry Sidgwick, A Memoir* (London: Macmillan, 1906).

good, though they know it to be the less valuable; and this no less when the choice is between two bodily pleasures, than when it is between bodily and mental' (*Utilitarianism*, 2.7); and for self-sacrifice: 'Unquestionably it is possible to do without happiness; ... it often has to be done voluntarily by the hero or the martyr, for the sake of something which he prizes more than his individual happiness' (*Utilitarianism*, 2.15). So the need to explain the sanctions of self-interest, how—in the case of weakness at least—to convert motivation by will (see *Utilitarianism*, 4.11) into motivation by desire, is as salient for Mill as it is for Sidgwick.

The dualism of practical reason is a doctrine about the sources of reasons. As Ross notes in his concluding remarks, Sidgwick recognised that his dualism need not result in any irresolvable practical conflict, though of course he did think it did so result. Ross also makes what I think is a very important point: anyone who attempts to move to an impartial principle on the back of an account of welfare which itself seems to assume that the agent has reason to pursue that welfare for herself may find herself in danger of a dualism. Here, then, is another place in which I think Mill anticipates Sidgwick. Not only does Mill offer us an account of welfare which implicitly appeals to our desire for pleasure, or our recognition that it is worth pursuing for ourselves, but his proof explicitly makes that appeal. Further, the final paragraph of the chapter on sanctions in *Utilitarianism* is a straightforward appeal to the rationality of self-interest: given the strength of each human being's desire to be in unity with others, and the higher pleasure to be taken in that unity, the best prospect of one's own happiness in a well-organised society will be found in pursuing the happiness of all. And, we assume, Mill meant that to be another argument in favour of impartiality, in certain circumstances at least.

The fact that he ends Chapter 3 like that, incidentally, suggests that he is, on this matter at least, less exact and coherent than either Bentham or Sidgwick. Mill does not keep clearly separate the psychological question—what might motivate a utilitarian agent?—from the ethical or normative question—why *should* anyone feel obliged to act in accordance with utilitarianism? (To be fair, neither does Sidgwick: when he says, in the passage quoted by Ross, '[M]any persons will still

feel that, after all, it has not really been shown why a man should be a consistent Utilitarian', it might have been clearer had he written 'would' rather than 'should'.)

My second niggles concerns Bentham (I'll come to Sidgwick at the end). Or rather it concerns what Ross says about Bentham. Ross attempts to draw a distinction between Bentham on the one hand, and Mill and Sidgwick on the other, relating to the scope of sanctions. The moral problem is how to motivate people to do what will promote the greatest happiness overall, when they are, to a large extent, out to promote their own greatest happiness. According to Ross, Bentham provides a 'political solution' to this moral problem. Well, that is of course correct as far as it goes. But it is not as if Bentham's sanctions don't provide other, non-political—that is, non-legislative—sources of motivation. An obvious example is the moral sanction itself—that is, the opinion of others—which Mill himself made so much of in accounting for the peculiar institution of morality itself. Right across the board, I see Bentham, Mill, and Sidgwick engaged in both the moral and the political enterprises in roughly the same way. None of them sees himself as a mere commentator on a machine baking cakes, but as an adviser to cake-makers—politicians, and anyone else who will listen—on how to build the best machines for the purposes of political or social, or indeed personal, morality, or how to tinker with the steam-driven items we find ourselves already working with, in the hope of improving their productivity.

Finally, some remarks on Sidgwick on sanctions. Ross's sympathetic and dispassionate account of Sidgwick's intuitionism is, I think, absolutely spot on. But I don't think that this intuitionism, when bolted on to Sidgwick's motivational internalism, is sufficient, by Sidgwick's own lights, to explain how and why we are able to be moral.

'Motivation' is a slippery term in contemporary ethics, and many writers appear ready to take views on it without spelling out exactly what they take it to mean. There is a clear distinction to be made, for example, between motivation which is sufficient for action, and that which is not, as well as between motivation as some kind of introspectively discernible felt impulse, and motivation understood counterfactually (to be motivated to ϕ is to be in a state such that, were

countervailing motivations absent, one would phi). I suspect that most contemporary motivational internalists accept a counterfactual version of a story about insufficient motivations. That is, if I believe that it is wrong to phi, I may indeed phi, but my state is such that, were countervailing motivations absent, I would not—indeed, would deliberately abstain.

So even to state a half-decent version of motivational internalism seems to require a good account of counterfactuals in this area which, as far as I know, we do not have. But Sidgwick does not run into this problem at all, since he understands motivation as an introspectible 'impulse'. Having stated his view on p. 34 of the seventh edition, he begins the next section as follows: 'I am aware that some persons will be disposed to answer all the preceding argument by a simple denial that they can find in their consciousness any such unconditional or categorical imperative as I have been trying to exhibit.' And, rather than offer us a counterfactual story, Sidgwick goes on to suggest—in a way that cannot help but remind one of Hume on the calm passions—that if these persons will only examine themselves carefully enough many of them will find that their denial of the moral impulse is really nothing more than an expression of hostility to deontological ethics.

Interestingly, Sidgwick does not insist that this *must* be the case, which raises another distinction between versions of motivational internalism: the empirical thesis that a person who judges that it is wrong to phi will be motivated not to phi, which can be turned quite happily into a generalisation not to be falsified by the odd exception, and a conceptual version of the same view, which would be in trouble with a single counter-example. I suspect that many modern internalists take the conceptual line; could it be that Sidgwick is here taking the empirical?

To return to sanctions, and the sufficient/insufficient distinction. Sidgwick does indeed say, as Ross quotes him: 'when I speak of the cognition or judgement that "X ought to be done"—in the stricter ethical sense of the term ought—as a "dictate" or "precept" of reason ... I imply that in rational beings as such this cognition gives a motive or impulse to action' (7, 34). That this is not a watering-down of internalism to the thin Korsgaardian gruel that even the toughest externalist

can imbibe is made clear by the implication which immediately follows, that every human being capable of making a moral judgement is *eo ipso* a rational being. But note the form in which this implication is embedded: 'in human beings, of course, this is only one motive among others which are liable to conflict with it, and is not always—perhaps not usually—a predominant motive'. In other words, the moral motivation that follows upon acceptance of a moral judgement is very likely to be sadly insufficient for the action required, and that leaves plenty of room to explain how and why sufficient motivation can be provided—that is, plenty of room for discussion of sanctions.

Unlike John Mackie, Ross is disappointed by Sidgwick's pessimism, and suggests that he might have kept to the plan of the first edition, left out the pessimism, and—Aristotle-like—referred the reader to his works on politics to explain how sanctions might provide a resource for bringing about an overlap between morality and self-interest.

The pessimism was certainly in line with Sidgwick's mood when he completed the book. In February 1874 he wrote of it: 'It bores me very much, and I want to get it off my hands before it makes me quite ill' (*Memoir*, 287), and the authors of the memoir say that the depression that affected Sidgwick in the final stages of writing a book was 'painful'. But even had he been more cheerful I think Sidgwick would have resisted Ross's suggestion. Sidgwick almost certainly thought that any project to make human beings to any great degree more impartial than they are would be self-defeating. The maxim of prudence could not be made consistent with that of rational benevolence:

There are very few persons, however strongly and widely sympathetic, who are so constituted as to feel for the pleasures and pains of mankind generally a degree of sympathy at all commensurate with their concern for wife or children, or lover, or intimate friend: and if any training of the affections is at present possible which would materially alter this proportion in the general distribution of our sympathy, it scarcely seems that such a training is to be recommended as on the whole felificific. (7, 502)

Further, as long as any gap remained between self-interest and impartial morality, and self-interest was of equal rational standing with

impartial morality, Sidgwick would not have found himself able, in a work on philosophy, to hide that fact. Sidgwick allows that it would be 'a most valuable contribution to the actual happiness of mankind' to come up with a machine to close the gap; but he is not, he says, 'now considering what a consistent Utilitarian will try to effect for the future' (7, 499). Sidgwick really thought the utilitarian philosophical project had run into the ground, and his integrity as a philosopher would not have permitted him to hide that all-important fact. And, because he was no utilitarian, he could not have accepted that there was any overriding moral reason to do so either.