Mill's Proof

TIM MAWSON

I. Introduction

> It is a most painful position to a conscientious and cultivated mind, to be drawn in contrary directions by the two noblest of all objects of pursuit, truth, and the general good. Such a conflict must inevitably produce a growing indifference to one or other of these objects, most probably both.[1]

On the one hand, in making a whole-hearted conversion to Utilitarianism, one might be supposed to have a two stage journey to complete. Starting from a general willingness to be moral, firstly, one would need to establish to one's satisfaction that Utilitarianism was true; secondly, one would need to establish to one's satisfaction that believing that it was true was more likely to lead to the end that it set for one (we shall take this to be the maximization of the aggregate happiness[2]) than believing that an alternative theory was true, or not having any beliefs about such matters at all. The second leg of the journey would be forced on one by the 'goal-post moving' conclusion of the first leg. By showing Utilitarianism to be true, the mere fact that it was true would have been shown not to be an overriding reason for believing that it was true. Of course, one cannot switch beliefs on and off like a tap and this might lead one to suspect the motives of anyone undertaking the second leg of the journey, including oneself. It might *prima facie* seem that one could have no reason to think that undertaking the second leg would or even ought to be motivated by an impartial desire to determine the truth about the matter. A belief that Utilitarianism was true might be a belief the having of which was sub-optimal in maximizing aggregate happiness. If so, then—on the truth of Utilitarianism—if one has, unfortunately, got to the stage of having the belief that Utilitarianism is true, it might appear likely to be better to be deceived into believing a falsehood about how conducive to

[1] J. S. Mill, 'Utility of Religion', in his *Three Essays on Religion* (London: Longmans, 1885), 71.

[2] It will be obvious how what I shall say could easily be adapted if an alternative formulation were preferable.

doi:10.1017/S0031819102000359 ©2002 The Royal Institute of Philosophy

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

happiness the having of that belief was. However, the argument in favour of such self-deception would entail only that it would be better for oneself: even if what one believes is not under the direct control of the will, if anything is under the control of the will, then how one acts is and one's actions can have good or bad effects on others. Thus, believing that Utilitarianism was true, one might have a genuine—truth-directed—interest in finding out whether acting always with the conscious and explicit aim of maximizing aggregate happiness was the best strategy for doing so. If one were to find out that it was, one would be a convert to Act-Utilitarianism. If one were to find out that it was not, one would be a convert to Rule-of-thumb-Utilitarianism, i.e. the only sort of Rule-Utilitarianism that deserves the name 'Utilitarianism'.

On the other hand, if one were to think that there was a qualitative distinction between sentences of the 'is type' and those of the 'ought type', with only the former directing themselves towards truth, then in converting to Utilitarianism—conceived of as an ought type doctrine—one would tread a quite different path. Proving Utilitarianism could not then involve proving to one's satisfaction that it was true; nor would it seem likely to involve proving to one's satisfaction that believing that it was true was conducive to the happiness of oneself or others. It might, I shall suggest, nevertheless involve proving that *following* the Principle of Utility was reasonable. If one may prove that it is reasonable to follow a principle by (and only by) showing that doing so is the strategy most likely to lead to an end which it is rational to set oneself, then if, as a matter of empirically determined fact, it turned out that there was only one end which people ultimately set themselves and this end was the maximization of their own individual happiness, it would follow that were people to be rational in choosing this end, they would be rational in following the Principle of Utility just if doing so were to be shown to be most probably the best means to the maximization of their happiness. One could argue that it must be most rational to set as one's ultimate end the maximization of one's own happiness if this is what everyone regards as most rational, understanding this claim of impeccability in the choice of end(s) in one of two ways. It could be understood either as a claim made within Realism—as an anti-sceptical assumption of, what might be called, 'Moral Epistemic Optimism'—or it could be understood as a consequence of a Quasi-Realist view of how the intersubjectively valid standard of rationality in the choice of end(s) is determined by peoples' attitudes. Intuitionism would incline one towards the former account; Naturalism would incline one towards

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

the latter. Henceforth, I shall thus be assuming that Mill sees the desirability—in the sense of worthiness—of the ultimate end as generated in a Quasi-Realist way by facts about what people actually desire and I shall not be arguing with this strategy. But on either understanding, if the empirical premise and claim of impeccability were accepted, the only further question would be whether, as another matter of empirically determined fact, it could be shown that following the Principle of Utility was more likely to lead to the maximization of one's happiness than following any other principle—including the principle of acting with the conscious aim of maximizing one's happiness—or following no principle at all. If this too could be shown, then—in the sense one would be operating with—the Principle of Utility would have been proved. One would not have rationally determined one's intellect to believe that Utilitarianism was true, but one would have rationally determined one's will to follow the principle that is Utilitarianism and this would be the only sort of proof to which one would expect a principle of practical reasoning to be susceptible. At this stage, one would need to rejoin the path of the argument sketched in the previous paragraph by considering whether acting always with the conscious and explicit aim of maximizing utility was the best means to do so. One would need to decide whether the best means to the end of promoting one's happiness was indeed to follow the Principle of Utility in reaching every decision (to be straightforwardly an Act-Utilitarian) or to follow secondary principles in most cases (to be a Rule-of-thumb-Utilitarian), and so the argument would continue.

In this paper, I shall argue that Mill's proof of Utilitarianism follows this second line of thought. Although, after this introduction, I shall speak without qualification of the position I shall outline as Mill's position and offer textual evidence in favour of attributing it to him, my priority in this paper will be philosophical and not exegetical: my main aim will be to present a particular valid and interesting argument. It is an argument suggested to me by my reading of Mill and when expounding it, I describe it as his. However, I am happy to admit that the Mill of whom I shall speak hereafter may or may not have much in common with the historical Mill and when I wish to venture a claim which I believe crosses the vague boundary from the history of philosophy into the history of ideas I shall therefore prefix the name 'Mill' with the word 'historical' to indicate this deviation from my main concerns. Having said this, I would say that the number of passages which have to be discounted as not indicative of the historical Mill's true opinions on my interpretation is very small and we shall see in a moment that there is a peculiar reason why the historical Mill, if

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

he were attempting to advance the argument which I attribute to his namesake, would find it difficult to present his case clearly. Thus, even if my interpretation were not to be preferred to others as an account of the historical Mill's views in eliminating a greater number of his apparent confusions, it would have the unique virtue of providing a philosophical explanation of why these confusions should arise in the first place.[3]

I shall argue that, according to Mill, the truth is that the only thing ultimately worth promoting for an individual is his or her own happiness and yet that acting with the aim of promoting the aggregate happiness is the best means for him or her to promote his or her own. It seems plausible to suggest that one would most wholeheartedly act to promote the aggregate happiness, and thus—if Mill as I shall interpret him is right—succeed best in promoting one's own, if one pushed from one's mind the fact that the only thing worth promoting was one's own—as it were, if one kicked away the ladder of his proof after having climbed up it. After all, '[i]f we wish men to practice virtue, it is worth while trying to make them love virtue, and feel it an object in itself, and not a tax paid for leave to pursue other objects.'[4] And what is true of other men will be equally true of ourselves.

[3] Day was the first to make the point that Utilitarianism cannot admit of proof in the sense of proof of its *truth*: as he put it, 'the question of the truth of [Mill's]... principle cannot arise'. Unfortunately, Day went on to add that 'Mill's main aim in *Utilitarianism*, which is to prove the truth of the utility-principle, is misconceived, and any attempt by a critic to establish its truth or falsity would be misconceived also.' According to the argument of this paper, Mill's main aim in *Utilitarianism* is not to prove the truth of the utility-principle, precisely because—according to the version of the is/ought distinction to which he subscribes—being a principle, it does not make sense to speak of it as true or false. Rather, Mill's proof may best be construed as a consideration designed to be capable of rationally determining the will to follow the Principle of Utility by rationally determining the intellect to believe in the ultimate egoistic psychological hedonism of humans and the aggregate-happiness-maximizing end which they must directly pursue if they are to maximize their own happiness indirectly. J. P. Day 'On Proving Utilitarianism', from *A Critical History of Western Philosophy*, edited by D. J. O'Connor (New York: The Free Press, 1964). Reprinted in J. Schneewind (ed.) *Mill: a collection of critical essays* (London: Macmillan, 1969).

[4] J. S. Mill, 'Inaugural Address Delivered to The University of St. Andrews' (London: Longmans, Green, Reader, and Dyer, 1867), 44. This is of course a point that also occupies much of Mill's attention in chapter four of *Utilitarianism*. Utilitarianism's ends are best served by maintaining 'not only that virtue is to be desired, but that it is to be desired disinterestedly for itself' J. S. Mill, *Utilitarianism* (London: Routledge, 1895) 67.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

If this is right, then accepting Mill's proof would put one in the predicament of which he speaks in the passage with which we commenced this discussion: one would be 'drawn in contrary directions by the two noblest of all objects of pursuit, truth, and the general good'. Accepting Mill's proof would entail a growing 'indifference' to the truth of its premises inversely proportional to one's growing passion for the general good—its conclusion—and this would naturally lead to the mistreatment of facts and arguments pertinent to those premises. All of this would help explain the manifest ambiguities and inadequacies in Mill's own presentation of his argument, faults perhaps most manifest in the passage with which he infamously begins chapter four of *Utilitarianism*.

II. Mill's Metaethical Assumptions

> No reason can be given why the general happiness is desirable, except that each person, so far as he believes it to be attainable, desires his own happiness. This, however, being a fact, we have not only all the proof which the case admits of, but all which it is possible to require, that happiness is a good: that each person's happiness is a good to that person, and the general happiness therefore a good to the aggregate of all persons.[5]

In this passage and those that surround it, Mill offers us what he elsewhere describes as 'equivalent to [a] proof'[6] of the 'grand elementary principle of pleasure' that is Utilitarianism: always act so as to produce the maximum happiness. There have been numerous and imaginative efforts to present Mill's proof in a light more favourable than that shone on it, for example, by Moore, which exposes it as being so 'naïve and artless' that it becomes 'quite wonderful'[7] that Mill could ever seriously advance it. Under the unforgiving glare of an expectation of a formally valid argument for the truth of Utilitarianism, it is all too easy to see Mill as starting from the inherently unstable foundation of an extreme psychological hedonism, and, in the space of the opening paragraph of chapter four of *Utilitarianism*, jerry-building an ostensibly deductive proof of his theory by committing the fallacy of equivocation (in conflating 'desirable' meaning 'capable of being desired' with 'desirable' meaning 'worthy of being desired') and the fallacy of composition

[5] J. S. Mill, *Utilitarianism*, 65–66.
[6] J. S. Mill, *Utilitarianism*, 8.
[7] G. E. Moore, *Principia Ethica* (Cambridge: CUP, 1903), 66.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

(in moving from each desiring their own happiness to each desiring the happiness of all). The paragraphs which follow that infamous paragraph—where the cracks in the slender foundation of Mill's argument appear to be papered over only by his willingness to oscillate between trivially true and obviously false versions of psychological hedonism, emptying the notion of happiness of any content—can then do nothing but increase one's amazement that Mill could have so completely and unwittingly lost control of his material.

Those of a more charitable disposition than Moore are thus quick to point out that convicting Mill of these logical fallacies ignores a crucial mitigating circumstance: it has become a commonplace to draw attention to the fact that Mill explicitly and persistently demurs from unqualifiedly referring to his arguments in favour of Utilitarianism as a 'proof' of it. Rather, Mill insists, they are to be understood merely as 'considerations … capable of determining the intellect'[8]. As such, one might think, they could reasonably be expected to be 'equivalent to proof' merely in their result—rational belief—and not in their formal structure. Whilst I have sympathy with the intentions of the charitable tradition as sketched thus far, this paper will suggest that taking this line does not entail—as it is normally thought to—that a) we must interpret Mill as offering us something *less than a deductive proof*, and b) we must interpret him as offering us this in support of the *truth* of Utilitarianism. Whilst locating myself within the tradition of charitable interpretation, I shall argue that both (a) and (b) are false: Mill is offering us *nothing less than a deductive proof*, but it is not a proof of the truth of Utilitarianism; it is a proof of the truth of the claim that being a Utilitarian is the most reasonable course of action available to any thoughtful reader.

In the rest of this section, we will look at the two metaethical foundations on which Mill rests his proof, one which we shall call his broad consequentialism and one which we shall call his endorsement of the is/ought distinction. We shall see that the first of these is uncontentious and the second is not.

Firstly then, his broad consequentialism:- 'That the morality of actions depends on the consequences which they tend to produce, is the doctrine of rational persons of all schools'[9]. So much, one might think, for Kantian persons who might have considered themselves rational. What ought we to make this claim? To argue that one

[8] J. S. Mill, *Utilitarianism*, 8.
[9] J. S. Mill, 'Bentham', *Collected Works*, vol. X, 111.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

should prefer to refer the rationality of one's actions to their ends as the alternative would leave us with nothing more than 'vague feeling or inexplicable internal conviction'[10] would be manifestly weak. The problem of one's moral judgements resting on a 'vague feeling or inexplicable internal conviction', if it is a problem, would arise, *mutatis mutandis*, in one's ultimate endorsement of ends of a certain type just as much as it would do in any ultimate endorsement of actions of a certain type. As Mill himself observes, the 'ultimate sanction ... of all morality (external motives apart)...[is] a subjective feeling in our own minds'.[11] A more promising line to take in support of Mill here then might be to argue that the movements of one's body could not be interpreted as actions one was performing, rather than merely events one was undergoing, without being interpreted as directed towards ends. Even if we allow things like 'to have performed an action of a certain type' as potentially amongst the ends towards which one might direct oneself, the type of action itself could not be specified save as one directed towards a certain end. For example, one could not set oneself to have performed an action of truth-telling without, more ultimately, having set oneself to convey to someone what one believed to be a truth. Thus, any sort of practical reasoning must be consequentialist in this—very broad—sense: in reasoning what one ought to do, one must be reasoning about what end or ends one ought to be trying to bring about. In this—very broad—sense, consequentialism is not contentious and it is only in this sense that Mill's proof needs us to assume that consequentialism is true.[12] Thus, as long as we remember how broad a sense we are operating with (it does not allow a distinction to be drawn between consequentialism and deontology), we can allow ourselves to assume this form of consequentialism; let us call it 'broad-consequentialism'.

Secondly then, his endorsement of the is/ought distinction:- 'A proposition of which the predicate is expressed by the words *ought* or *should be*, is generically different from one which is expressed by *is*, or *will be*.'[13] Sciences 'assert a matter of fact'; arts, by contrast, 'do not assert that anything is, but enjoin or recommend that something should be'. 'Science is a collection of *truths*; art a body of *rules*

[10] Ibid.

[11] J. S. Mill, *Utilitarianism*, 53.

[12] This is the sense of 'consequentialism' in which Kant would be a consequentialist, as discussed by Williams in J. Smart and B. Williams, *Utilitarianism for and against* (Cambridge: CUP, 1991), 83.

[13] J. S. Mill, *A System of Logic*, vol. II, 526–7.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

or directions for conduct'[14]. Mill holds then that there is an is/ought distinction, a conceptual gap between the thoughts expressed by is type sentences and those expressed by ought type sentences. According to Mill, ought type sentences are more or less disguised forms of imperative and as such do not state any facts but merely enjoin certain actions. Ought type sentences do not state that their speaker is desirous of the ends which the action enjoined would serve; neither do they state that the ends which would be served are worthwhile. Of course, one would not (sincerely and understanding what one was doing) enjoin an action which one did not believe served an end which one desired and believed to be worthwhile and thus the fact that a certain person utters an ought type sentence may be taken as necessarily good evidence that he or she does desire the end that would be served by the action and believes that end to be worthwhile (because such utterances are of linguistic necessity most often correlated with such desires and beliefs), but nevertheless ought type sentences have no fact stating element.[15]

Mill's assumption that there is this division between the thoughts expressed by is type sentences and those expressed by ought type sentences is a contentious assumption, but it is one which he shares with many other great thinkers. Kant thinks that ought type thoughts do not direct the intellect towards truth; Kant thinks that they direct the will towards duty. Mill, as we shall see, thinks they direct the will towards happiness. Rather than examine in more detail whether such a distinction can be defended, let us look ahead slightly and see how Mill uses it in his proof.

We may observe that Mill needs a strong is/ought distinction in order to avoid embracing in his conclusion a version of Moore's paradox—'P and because P, do not believe that P'. He does not want to end up arguing that ultimate egoistic psychological hedonism (a thesis we shall define in due course) is true but, because it is true, we should not believe that it is true but should believe that Utilitarianism is true instead, and he can avoid this destination with this assumption. A form of egoistic hedonism is true: it is an is type

[14] J. S. Mill, 'On the Definition of Political Economy and on the Method of Investigation proper to it', *Essays on Some Unsettled Questions in Political Economy*, *Collected Works*, vol. IV, 312.

[15] Whilst composing the *Logic*, Mill says in a letter that it is to be 'a logic of the indicative mood alone—the logic of the imperative, in which the major premise says not *is* but *ought*—I do not meddle with' (J. S. Mill, 'Letter to Sterling', *Collected Works*, vol. XIII, 412.) How could the *logic* of ought type sentences differ from that of is type sentences if there was a fact stating, indicative, element to both?

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

empirical theory about what people actually desire and thus—we shall come to the 'thus' in due course—about what is most desirable in the sense of worthwhile. Given that it is true, it is reasonable to follow the Principle of Utility, an ought type principle, as, given the broadly-consequentialist nature of practical reasoning, it is reasonable to adopt that policy most likely to realize the ultimate end which one rationally sets for oneself (allegedly, the maximization of one's happiness) and, as another is type matter of fact, following the Principle of Utility is the best means to the maximization of one's happiness. As mentioned in the introduction, to think that one should follow the Principle of Utility because a version of egoistic hedonism is true is peculiarly psychologically uncomfortable, but—with this form of is/ought distinction—it avoids being paradoxical.

In this section then, we have allowed ourselves to bring out into the open two assumptions that Mill is inclined to make, assumptions which inevitably affect the style of the structure for which they form the foundations. We have argued that his assumption of broad-consequentialism should be uncontroversial. Any extended analysis or underpinning of his contentious and important assumption of a strong form of the is/ought distinction would take us well beyond the scope of this paper. However, having noted that it is both contentious and important, we have no need to excavate this particular foundation further to examine with profit the strengths and weaknesses of Mill's proof as a structure built upon it. This is the task to which we shall turn in the next two sections.

III. Mill's Meta-Psychological Assumption and His First Psychological Premise

> If the opinion which I have now stated is psychologically true—if human nature is so constituted as to desire nothing which is not either a part of happiness or a means of happiness, we can have no other proof, and we require no other, that these are the only things desirable. If so, happiness is the sole end of human action.[16]

Even if we think of happiness in the relatively simple way that Mill does (as the purely mental state of pleasure and the absence of pain[17]), then that human nature is so constituted as to desire

[16] J. S. Mill, *Utilitarianism*, 72–3.

[17] My discussion here and hereafter is premised on taking it that Mill's claim that by 'happiness is intended pleasure and the absence of pain' (J.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

S. Mill, *Utilitarianism*, 13) is indicative of his position as what Crisp calls a 'pure hedonist' with a pure mental state conception of happiness (R. Crisp 'Mill on Virtue as a Part of Happiness', *British Journal of the History of Philosophy*, 4, 1996, 372ff). In taking this 'traditional' reading of *Utilitarianism*, I go against a recent trend in the interpretation of Mill whereby he is seen as positing happiness as an 'inclusive end', akin to Aristotelian eudaimonia, or as holding a 'deliberative conception of happiness'. (F. Berger, *Happiness, Justice, and Freedom: The Moral and Political Philosophy of John Stuart Mill* (University of California Press, 1984); J. Gray, *Mill on Liberty: A Defence* (Routledge and Kegan Paul, 1983); and R. Hoag, 'Happiness and Freedom: Recent Work on John Stuart Mill', *Philosophy and Public Affairs*, 15, 1986, 'Mill's conception of happiness as an inclusive end, *Journal of the History of Philosophy*, 25, 1987, and 'J. S. Mill's language of pleasures', *Utilitas*, 4, 1992; and D. Brink, 'Mill's deliberative utilitarianism', *Philosophy and Public Affairs*, 21, 1992.) Unfortunately, space does not permit me to justify going against this trend in a paper focused on other issues. The issue is taken up again briefly in the last note to section III.

On a traditional reading, it seems that Mill would be best advised—for reasons of consistency—to drop the claim—which he should cleave to for reasons of verisimilitude—that there are differing qualities, not just quantities, of pleasure: thinking that there were differences in quality of pleasure, differences which did not amount simply to differences in quantity, would not be consistent with Mill's central thesis. If we say with Mill that the pleasure attendant upon an hour engaged in an intellectual pursuit (say reading poetry) is of a 'higher quality' than that attendant upon an hour spent in some more lowly occupation (say playing pushpin) and hence that we would be rational in seeking out the former in preference to the latter, then we must be positing that the two 'qualities' of pleasure may in fact be compared on some scale (so that one may be said to be higher than the other) and thus that they are, after all, differing quantities of some one thing, a thing which it would then be natural to call 'happiness' so that we might say that the two 'qualities' of pleasures may both be considered potentially 'ingredients' or 'parts' of happiness. Given Mill's initial definition of happiness—pleasure and the absence of pain—this would then entail that we were committed to two other claims (both of which contradict ones Mill makes explicitly in *Utilitarianism*). We would need to say firstly that the two 'qualities' of pleasure were in fact differing quantities of the one pleasure and thus secondly that it was not true that *any* amount of a 'higher' pleasure was worth more than *any* amount of a 'lower' pleasure. Whilst committing Mill to the second of these claims would be offering him an advantage, in that it is obviously true, committing him to the first would be to take an advantage away, in that it is obviously false. Thus the implausibility of maintaining that all values can be 'boiled down' to the one homogenous mental state—happiness—is one which I consider ineliminable for the traditional Mill. However, I do not regard it as so great an implausibility as to justify abandoning interpreting Mill in the traditional way.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

nothing which is not either a part of happiness or a means of happiness (what we might call 'psychological hedonism') is still an assertion which can be taken in a number of different ways. A division between an egoistic and a disinterested reading is perhaps the first to suggest itself. On the former, Mill would be commencing his argument for Utilitarianism from the unpromising starting point of the claim that humans are, by nature, essentially selfish in always looking to their own happiness without any non-instrumental regard for the happiness of others. On the latter reading, Mill would have apparently assisted himself in reaching Utilitarianism by starting from an assumed tendency to put other people's happiness on a par with one's own when deciding how to act. It would still be the case that people by their nature were so constituted as to desire nothing other than happiness, but the happiness which they were so constituted to desire would be or would include perhaps the aggregate or average of the society in which they found themselves.[18] On either reading, 'we have evidently arrived at a question of fact and experience dependent, like all similar questions, upon evidence'.[19] If one accepts that ought implies can (and I shall be assuming that, at least in this context, one should), the question must turn on whether the evidence needed to establish some version psychological hedonism will present itself. We shall be assuming that Mill starts his proof from a version of egoistic psychological hedonism. In making this assumption, we must acknowledge the fact—observed by many recent commentators—that much of what Mill says in *Utilitarianism* and elsewhere would naturally be taken to suggest that the historical Mill did not subscribe to any version of egoistic psychological hedonism at all. However, we must weigh against this the less popular fact that much of what Mill says in *Utilitarianism* and elsewhere *would* naturally be taken to suggest that the historical Mill did subscribe to some—rather inchoate—version of the theory. Given the peculiar argument which—if the thesis of this paper is correct—Mill thought established the Principle of Utility as that which all thoughtful readers should follow (an argument the conclusion of which enjoins on one the forgetting of its premises), this is exactly what we would expect from the writings of someone who had already 'kicked away the ladder' of his proof after having climbed up it, a ladder the bottom rung of which was a version of egoistic psychological hedonism. Let us assume then that we may say that Mill starts from some version of egoistic psychological hedonism.

[18] Of course, there are other intermediary positions too.

[19] J. S. Mill, *Utilitarianism*, 73.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

It is true that in his *Remarks on Bentham's Philosophy*, we do find the historical Mill somewhat ambivalent—if not ambiguous—on this issue. There he explicitly tells us that it is 'entirely false' to suggest 'that all our acts are determined by pains and pleasures *in prospect*, pains and pleasures to which we look forward as the *consequences* of our acts. This, as a universal truth, can in no way be maintained.' But he goes on to add that the 'pain or pleasure which determines our conduct is as frequently one which *precedes* the moment of action as one which follows it. ... [a person] recoils from the very thought of committing the act; the idea of placing himself in such a situation is so painful, that he cannot dwell upon it long enough to have even the physical power of perpetrating the crime. His conduct is determined by pain; but by a pain which precedes the act, not by one which is expected to follow it.'[20] Just after this, Mill calls acting from a deliberate and conscious aim to reach some end, for example the maximization of one's happiness, acting from 'interest' and acting from a desire the acting on which is in itself and immediately the most pleasurable option, acting from 'impulse', the implication being that the two are logically separable and empirically capable of independent instantiation. How are we to understand Mill's claims here? Let us consider them in order.

It is certainly true that a course of action under a certain description may be so painful to contemplate that one can neither dwell upon it, nor do it. However, the important feature of potential actions which are so repulsive to us that it pains us to this extent even to contemplate them is that in not performing them we cannot truly be said to have exercised genuine choice. If we consider the case of Oedipus, his behaviour on learning that he had unwittingly killed his father and married his mother gives us every reason to suspect that had he known that the man he killed was his father prior to killing him or known that the woman he married was his mother prior to marrying her, he would not have done either of these actions not because he would have *chosen* not to, having performed a calculation that doing them would provide him with less happiness in the short-term or long-term than not doing so or having brought them under some other principle. Rather, it would have been because the very thought of doing them would have been so repugnant to him that he would not have been able to consider them further; he would have been psychologically unable to perform the calculation or reasoning in question. Any action which it did not

[20] J. S. Mill, 'Remarks on Bentham's Philosophy', *Collected Works*, vol. X, 12–13.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

occur to you to perform whilst reading the previous sentence was not an action which you chose to read that sentence rather than perform. And any action so painful to contemplate that should it occur to you whilst reading this sentence you would push it immediately from your mind as an alternative to continuing reading—for example, if you will forgive the intrusion, killing your father or marrying your mother—again is one which you cannot truly be said to have chosen not to perform (rather than, perhaps, to have chosen not to consider whether or not to chose to perform).

Turning then to situations of genuine choice, we may say that if any egoistic psychological hedonist theory of action is right, the pleasures/pains *from which we act* must always be 'in prospect'—otherwise they would never drive us to act, rather than merely contemplate action or push actions from our contemplation. If any egoistic psychological hedonist theory of action is right, then—apart from the pleasures/pains of contemplation, which are inert with respect to genuine action—the pleasures/pains which Mill misleadingly describes as 'preceding' the act must simply be the short-term pleasures/pains which we suppose will be attendant upon performing the act; the ones which he talks of as 'consequences' must be those pleasures/pains which flow from the act in the longer-term. Acting from a consideration of the former might then be called acting from an impulse; acting from a consideration of the latter, acting from an interest. With these observations, within a traditional reading of Mill, all hope of maintaining that wanting something as 'a part of happiness' could be something other than wanting it as a means to (short-term or long-term) happiness disappears. But this is to Mill's advantage, assuming that one thinks it to Mill's advantage that he be attributed a theory with all the virtues of simplicity and consistency that belong to Bentham's, with which his theory becomes in substance identical.[21]

There are five desire theories of actions which deserve the name 'egoistic psychological hedonist' and, with the foregoing observations, we might now think that we have the resources to determine to which of them Mill ought to have subscribed. Firstly, we might consider the theory that all actions are motivated by desires which refer to one's interests—the maximization of one's total happiness, where this is understood as the maximization of the area under a graph plotting one's pleasure/pain against time for the duration of one's life; the amount of pleasure/pain felt at any time being simply

[21] A small caveat concerning their (possibly) different understandings of the importance of propinquity/remoteness and the reason for its importance should probably be inserted here.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

a matter of how one feels along the relevant dimension, independent of the time at which it occurs.[22] Secondly, we might consider the theory that one always acts out of impulse—out of those desires the acting on which is itself the source of most immediate happiness for one, that is one always acts so as to maximize the positive gradient on a graph of the sort described above. If we perform a certain action, that is because doing so is the most immediately pleasurable or, what—on Mill's account—amounts to the same thing, the least immediately painful of the actions available to us: we act always from impulse. Thirdly, we might consider the theory that all actions are motivated by desires which satisfy both conditions. We have then, from an appreciation that in a particular situation of choice one might realize that maximizing the gradient of one's curve did

[22] Unfortunately, though it has the attraction of *prima facie* simplicity on its side, when we translate this thought across into disinterested hedonism, it requires a commitment to a claim which is likely to be found unacceptable enough as a consequence so as to vitiate this attraction entirely. The consequence of this view is that 'the curve must end at some finite points on its axes or else asymptotically approach them. Only under these conditions is a solution possible [i.e. does the injunction that is Utilitarianism have specific content], since only under these conditions is there a definite area under the curve at all.' (M. Nelson in 'Utilitarian Eschatology', *American Philosophical Quarterly*, 1991, page 340.) The corollary of disinterested hedonism that lives enjoying/enduring pleasure/pain will either cease to exist or asymptotically approach neutrality (a state which contains exactly as much pleasure as it does pain) is, at best, far less obviously correct than the analogous corollary of self-interested hedonism: it is obviously true that any particular agent is himself or herself going to die at some future time. The problem of this surprising eschatology being enjoined on a Utilitarian could be avoided by that version of Utilitarianism which prioritizes short-term over long-term gains as discussed below, discounting future pleasures/pains, and this problem is therefore a reason to prefer that account in addition to any empirical ones, e.g. that, pains of anticipation aside, most would choose the certainty of eleven minutes of torture in twenty years time over the certainty of ten minutes of equally painful torture now. Further discussion of Nelson's argument may be found in P. Vallentyne, 'Utilitarianism and Infinite Utility', *Australasian Journal of Philosophy*, 71, 1993 and 'Infinite Utility and Temporal Neutrality', *Utilitas*, 6, 1994; and in J. Garcia and M. Nelson, 'The Problem of Endless Joy: Is Infinite Utility Too Much for Utilitarians?', *Utilitas*, 6, 1994. For further discussion of whether it is rational to discount future pleasures/pains simply in virtue of their being future, see J. Rawls, *A Theory of Justice* (Oxford: OUP, 1973), sec. 45, pp. 293–8; sec. 63, pp. 407–11; and sec. 64, pp. 416–24. See also D. Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), sec 62, page 158ff.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

not lead to the maximization of the area under it (one might realize in a particular case that a short-term benefit would lead to a greater long-term loss), reason to avoid this theory (it would have as a consequence that in such a situation, one could not act). Fourthly, we might consider the theory that all actions are motivated by desires which satisfy at least one of these conditions. Fifthly, we might consider the theory that all actions are motivated by desires which satisfy one or other of these conditions but not both. We have then, from an appreciation that in a particular situation of choice one might realize that maximizing the gradient of one's curve did lead to the maximization of the area under it (one might realize that a short-term benefit did not lead to a greater long-term loss), to avoid this theory (it would have as a consequence that in such a situation, one could not act). Still 'in the running' at this stage then we have the first, second and fourth theory.

From Mill's discussion of Bentham, it seems as if the historical Mill would have preferred the fourth theory and—our own reflections might lead us to think—wisely so. It seems perfectly possible that one might, on learning of the painful death that awaited one in the longer-term if one continued smoking, decide to forsake the pleasures which would await one in the shorter-term were one to continue. Equally, it seems possible that one might prioritize a smaller amount of short-term pleasure (e.g. that attendant upon continuing smoking) over a larger amount of long-term pleasure (e.g. that attendant upon good health in later life). It is a plain fact that when made aware of the likely effects of smoking some people give up and some do not.[23] However, from the very fact that a moment's reflection reveals that acting always so as to maximize one's gradient might not—over the course of one's life—lead one to maximize one's area (and, of course, that acting so as to maximize one's area might not lead to one's always, from moment to moment, maximizing one's gradient), we have reason to think that not *both* (each of which would be allowed by the fourth theory) are of necessity what is best for one. Yet if Mill's theory is to deserve the name 'egoistic', then it must have as a consequence that people never knowingly do what is worse for them. We must therefore ask whether there is an obvious answer as to which is better for one, an answer which would allow us to choose on Mill's behalf between the first and second theories. Of course if there were no sufficiently obvious answer to this question, Mill might rest with the fourth theory as it would not have as a consequence that people sometimes

[23] See also J. S. Mill, *Utilitarianism*, 19.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

knowingly did what was worse for them. Do sufficient people have the sorts of preference needed to make one or other of these accounts the correct one? Is it obvious that they do? These are, I take it, empirical questions. We have raised them and, fortunately, we do not need to answer them to proceed with our argument. Either the right sort of people have these sorts of preferences in sufficient numbers to generate the relevant ethical facts or they do not; either it is obvious that sufficient number of the right sort of people have these sorts of preference or it is not obvious that they do. Just as we may leave determining criteria for the right sort of people and sufficient numbers as a piece of 'unfinished business' for the Quasi-Realist in general, so we may leave the gathering of the evidence as to which of these is the case as a piece of 'unfinished business' for the egoistic psychological hedonist Quasi-Realist in particular, and—suspending judgment on whether it is the first, second or fourth version of the theory that he or she is drawn to by this evidence—proceed with our analysis of the strengths and weaknesses of the position. This is because there is another and far more pressing problem which will remain for egoistic psychological hedonism however this one as to its content resolves itself: any version seems to face a Morton's fork. If it is substantive, it is demonstrably false; if it is insubstantial, it can have no implications.

The falsehood of any simple version of egoistic psychological hedonism as a theory of action may be driven home by another example from Greek mythology. Prometheus, let us suppose, was a god who wished to help mankind. He thus stole fire from the sun to give to them and also arranged that they should get the best part of any animal sacrificed. He did the latter by (temporarily) tricking Zeus into choosing for himself the worst parts by disguising them to look as if they were the best. Prometheus, we are free to imagine, knew when he did these things that he would not go unpunished for long: he had forethought. So it was. At Zeus' command, Force and Violence seized him and bound him to a 'high-piercing, headlong rock in adamantine chains that none can break' telling him, 'Forever shall the intolerable present grind you down. And he who will release you is not yet born. ... No rest, No sleep, no moment's respite. Groans shall your speech be, lamentation your only words.' Prometheus sacrificed his own happiness for that of men. He was the first martyr.

An objector to any version of egoistic psychological hedonism could it seems point out that Prometheus was someone whose actions provide a clear refutation of it as a theory of action. Either a desire theory of action is wrong (Prometheus acted on a principle

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

whilst not in any sense desiring to) or the claim that the only thing one desires is one's own happiness is wrong, because Prometheus knowingly acted in such a way as to fail to maximize his happiness in the long-term and from a desire the acting on which we are free to imagine he knew would not be in itself particularly pleasant to him in the short-term. We cannot afford to open up another front, defending desire theories of action in general against Kantian-style attacks. Let us therefore assist Mill by granting him another large assumption, in this case a meta-psychological one: a desire theory of action is right. The threat from the case of Prometheus then will be that his actions seem to show that people can desire things that have nothing to do with their own happiness (even if in this case they might plausibly be claimed to have something to do with the happiness of mankind) and act on those desires however unpleasant for them they know the longer-term consequences of doing so to be and even if they find acting on them less pleasant than acting on an alternative desire in the shorter-term.

If Mill were to attempt to accommodate stories of apparently altruistic martyrdom such as that of Prometheus to simple egoistic psychological hedonism as a substantive theory of action, he would have only three options, none of which should strike him as particularly appetizing. Firstly, he could maintain that Prometheus believed that between the moment of making his decision to help men and the commencement of his punishment, he would have such a degree of pleasure as to compensate for that suffering which he expected to have thereafter. As an account of the psychological processes which it seems reasonable to suppose Prometheus attributing to himself, this is most implausible. If necessary, we could adapt the story by increasing without limit the suffering Prometheus expected to endure and the certainty of this expectation, the point being that Mill must—in order to be able later to maintain the unity of the fundamental principle of practical reasoning—allow that expected future suffering can be weighed against[24] any pleasure between the moment of choice and Zeus' punishment and that insofar as Prometheus failed to perform the relevant comparison between these two, this avenue of explanation will turn into the third, discussed below. Secondly, Mill could

[24] It can be weighed according to *some* formula however the question we left over as 'unfinished business' earlier is answered, although certain formulae might provide a solution to it, for example one which placed a zero-value on any pleasures/pains to be enjoyed or endured more than a moment after a decision. (There are of course more plausible formulae than this one, which equally solve this problem.)

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

maintain that—being a god—Prometheus was in it for the long-term. Prometheus might have reasonably expected that the pleasure that would come after he who had not yet been born eventually delivered him would be so great as to provide adequate compensation for the torture he would need to incur in the short-term. Whilst plausible in the case of Prometheus, such a theory cannot accommodate the more mundane stories of agents who sacrificed themselves whilst believing themselves to be mortal. We know, for example, of atheists who threw themselves on hand-grenades to save their friends. Thirdly, Mill might contend that the thought of not giving fire to mankind or allowing them to have the best pieces of any animal they might sacrifice to Zeus, was so repugnant to Prometheus that he could not consider it further, just as Oedipus would not have been able to consider marrying Jocasta further had he known her true identity. If this does not already sound implausible, it seems easy to adapt the story so as to rule it out: there is nothing repugnant about failing to offer someone what one thinks of as supererogatory goods and we can, without any absurdity, present Prometheus as thinking of fire or, at least, the best parts of sacrificed animals as supererogatory goods.

Egoistic psychological hedonism as stated above *is* demonstrably false. To defend any recognizably similar position on Mill's behalf we need to stage a tactical withdrawal from the positions the historical Mill had an occasional tendency to occupy: but we can stop this retreat becoming a rout by redrawing his battle lines, introducing the crucial distinction between objects of desire and objects of ultimate desire. According to this more defensible version of egoistic psychological hedonism, which we might call 'ultimate egoistic psychological hedonism', although we each ultimately desire our own happiness in the way suggested by egoistic psychological hedonism, this leaves room (indeed generates the need) for us to desire non-ultimately other things. Ultimate egoistic psychological hedonism thus accommodates an evident fact which simple egoistic psychological hedonism may seem to deny: one cannot simply choose to be happy; if one wants to be happy, one has to choose to do things which will make one so. According to the more cautious Mill we can—as a matter of fact—rule out these other things as objects of ultimate desire. According to this Mill, to the extent that we are made aware of the fact that a certain thing will not bring us happiness, we cease to desire that thing and that this is how our desires operate is a contingent truth. That we regard those whose 'desires' do not operate in this way as not having anything which can properly be called a desire is also, according to Mill, a

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

contingent truth, but that it is a truth gives us reason to suppose that the truth of the claim that 'to the extent that we believe something will bring us happiness we desire it and to the extent that we believe it will not bring us happiness we cease to desire it' is a metaphysical necessity, gives us reason to think that *ultimately* desiring a thing and finding it pleasant are two ways of referring to the same mental state.

In order to see how stories such as that of Prometheus may be accommodated to ultimate egoistic psychological hedonism, let us consider another example, this one not taken from classical literature. Barnaby the teddy bear knows that attending the teddy bears' picnic will make him happy. (There are lots of marvellous things to eat and wonderful games to play associated with this event.) Thus, he has good reason to desire non-ultimately to attend the picnic. Barnaby knows that to attend the picnic he would need to go down to the woods today. Thus, he has good reason to desire non-ultimately to go down to the woods today. And so on. Now we come to another crucial feature of human (and divine and teddy bear) psychology: focusing on a more ultimate end can distract one from the task of pursuing a less ultimate means to that more ultimate end and thus diminish one's chances of achieving it. To return to our example of Barnaby's wanting to attend the teddy bears' picnic, it is quite possible that if he allowed himself to dwell on how happy he would be made by all the wonderful games and food which are associated with the teddy bears' picnic, Barnaby, being an overly excitable bear, would become too excited to get down to the woods as quickly as he would do otherwise. Knowing this, as he leaves his house in the morning, Barnaby would be reasonable in distracting himself from his ultimate end, focusing solely on getting down to the woods as quickly as possible. Were every day to be a day the teddy bears had such a picnic, it might be in Barnaby's best interests were he to fashion himself so that he rushed to the woods first thing every morning without any thought as to how happy it was likely he would be made by the picnic he would find there. This might be so even if he knew that this might not always be in his best interests: Barnaby might know that it was just possible that one morning—having 'fixated' on getting to the woods—as he hurried eagerly from his house towards the picnic, he might miss the opportunity to join a group of teddy bears who were off to do something even more conducive to happiness than attending the teddy bears' picnic, attending the teddy bears' Schoenberg festival for example. The important thing for the rationality of Barnaby's nevertheless forming a character which fixates on getting to the woods as

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

quickly as possible each morning is that such an opportunity is so unlikely that he maximizes his expected utility by adopting a policy which means that he would never avail himself of it were it to arise. Egoism dictates that one never, in any particular situation of choice, knowingly does what is worse for oneself; however, it does not dictate that one never adopts a policy which one knows will entail that in some particular situation(s) one will—unknowingly at that time—do what is worse for oneself. As our example of Barnaby the teddy bear shows, empirical contingencies may mean that it dictates just the opposite.

With the above facts in hand, Mill has the tools available to him with which he can fashion an answer to the question of why martyrs such as Prometheus occur and why we can, to an extent, often rationalize what they do. The ultimate desire for one's own happiness could form a reason for developing a state of character which—in the extremely rare circumstances where martyrs are required—allowed that one might do something for a non-ultimate end that, in one's fixated state, one prized 'more than [one's]... individual happiness'[25], one's ultimate end. The reasonableness of forming such a character obviously depends on the comparative rarity of its leading to the non-maximization of one's happiness, which is why Mill is so keen to stress that society can be, and in those civilized societies which his thoughtful reader finds himself or herself, *is* ordered so that martyrs are not required. Martyrs—qua martyrs—are not rational, but they might have been rational in putting themselves into a mental state in which, acting on a 'divine enthusiasm'[26], it was possible that they would behave as martyrs and thus when they act from this mental state we can still interpret the movements of their bodies as actions.

To draw these strands together then, Mill can—assuming the right sort of empirical evidence presents itself—defend ultimate egoistic psychological hedonism as a substantive psychological truth. When faced with the 'counter-examples' of martyrs, he can explain away most cases of martyrdom by virtue of the martyrs having reasonably fixated on a non-ultimate end, one which has—unfortunately for them, if not for us—ceased to serve their ultimate end. Insofar as these 'counter-examples' do not admit of being explained away in this fashion, he can describe them as not meeting the ideal of rationality which is constitutive of action—for being explicable by reasons is what distinguishes actions from mere

[25] J. S. Mill, *Utilitarianism*, 29.
[26] J. S. Mill, 'Utility of Religion', 95.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

bodily events—and thus redescribe these 'counter-examples' to his theory of action as not really examples of actions at all. When it is suggested that this is to build his preferred conclusion into the definition of rationality (expected personal utility maximization) from which he starts, Mill may maintain that rational behaviour must be what the majority of people take to be rational behaviour and those who one might *prima facie* be tempted to describe as (knowingly and after sufficient time for reflection) choosing to fail to maximize their utility are a tiny minority considered by the majority to be paradigms of arationality[27] (even if they are sometimes, for self-serving reasons, publicly 'praised' for bodily movements which are not their actions): this is the discovery of the truth of ultimate egoistic psychological hedonism. That it is a discovery is why evidence is needed to support it, but this evidence is provided by the fact that introspection will reveal to every thoughtful reader that whilst they do indeed want things other than their own happiness and sometimes fixate on them—presenting them to themselves as 'ends in themselves'—they also want their own happiness and, further reflection will reveal, they want their happiness as their one and only *ultimate* end (i.e. they want it more than they want anything else and want these other things merely as a means to it). It is not the job of a chapter entitled 'Of what sort of Proof the Principle of Utility is Susceptible' to present this evidence: its job is to show where the evidence fits in an argument and then delegate the collecting of it to the introspective powers of its reader. As one would expect then if ultimate egoistic psychological hedonism is to be defended as a substantive truth, the crucial question we must ask is whether introspection does in fact yield the right sort of empirical evidence. In answering this I can follow no other path than that taken by Mill: I leave it 'to the consideration of the thoughtful reader'.[28]

[27] I use 'arationality' to mean behaviour that is such a 'big blunder' with respect to rationality, that one is inclined to regard it as falling outside the bounds rational/irrational action altogether.

[28] However, I will report that the vast majority of my students are convinced: they think that maximizing one's happiness (conceived of as a pure mental state) over the course of one's life *is* the only thing which it is rational for one to desire ultimately. They think that the wrong of maligning someone is to be explained entirely in terms of the bad feelings it generates if the person concerned finds out; they are willing to accept that maligning someone after their death thus does them no wrong; they would step into Nozick's experience machine without any regret whatsoever and think that a reluctance to do so can only be based on *ex hypothesi* false beliefs about the machine or irrationality. Making a difference to the real

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

IV. Mill's Second Psychological Premise

Returning to the famous passage with which Mill commences chapter four of *Utilitarianism*, we may now acquit him of the charge of

world; true beliefs; free-will; and everything else have worth to an individual only as a means to his or her happiness and thus, if once they lose this worth, they are to be sacrificed without regret on its altar. For contrary arguments, by which they remain unmoved, see T. Nagel, 'Death' in his *Mortal Questions* (Cambridge: CUP, 1979) and R. Nozick, 'The Experience Machine' in his *Anarchy, State, and Utopia* (Oxford: Blackwell, 1995).

Most often, a discussion of this topic in a tutorial thus leaves me feeling 'out-smarted' in Dennett's sense, where one out-smarts one's opponent in an argument by blithely accepting as obviously true the conclusion of their reductio of one's position. However, I report this not with the aim of providing evidence that something is going wrong in my admissions procedure or tutorial technique (although it may do that), but rather to show that ultimate egoistic psychological hedonism (in contrast to egoistic psychological hedonism) is not the cliff over which to push one's opponent, to use another Dennettism, that many interpreters of Mill seem to think it is. This, admittedly anecdotal, empirical evidence leads me to hazard that most people, if they reflected to the extent that most of my students do, would accept (the first—i.e. 'area under one's graph'—version of) ultimate egoistic psychological hedonism in the form attributed to Mill in the main text preceding this note. If I further—I hope not too optimistically—posit that the level of reflection achieved by most of my students is that which might reasonably be expected of the thoughtful reader to whom Mill addresses his arguments in *Utilitarianism*, then I may conclude that the rush of most commentators to disencumber Mill of the theory of ultimate egoistic psychological hedonism is misconceived (roughly speaking, we might say from the point of view of the history of philosophy if not from the point of view of the history of ideas). If we do say that ultimate egoistic psychological hedonism of the sort sketched in the main text is plausible to the majority of people who have thought about it enough to have an opinion on it, yet it is nevertheless wrong, we must reject Quasi-Realism (and Realism combined with Moral Epistemic Optimism), leaving us (after we have rejected extreme subjectivism), it seems, with Realism and Moral Epistemic Pessimism. The facts about what is in a person's best interests are independent of what they think is in their best interests and the majority of people can be, and in this case are, mistaken about them. I think it is a desire not to be left in such a position that prompts many commentators to under-rate the attraction of ultimate egoistic psychological hedonism (and in many cases conflate it with egoistic psychological hedonism) and thus think they have something of a history-of-philosophy-reason to attribute to Mill a different theory. In any event, ultimate egoistic psychological hedonism is not a wholly implausible empirical premise.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

the fallacy of equivocation by reading his discussion of visibility and audibility as follows. Actually being (ultimately) desired is the only evidence we could have that a thing is capable of being (ultimately) desired. This, when considered with the (suppressed but plausible in this context) premise that ought implies can, i.e. capability of being (ultimately) desired is a necessary condition for being (ultimately) desirable in the ethical sense, and the putative fact, which we called Mill's first psychological premise, that as a matter of fact nothing is (ultimately) desired by anyone except his or her happiness, then—on the further plausible assumption that one could not justify a claim that something was capable of being desired if it was never actually desired—yields the conclusion that an individual's happiness is the only thing (ultimately) desirable in the ethical sense for that individual. We have moved from ultimate egoistic psychological hedonism to ethical[29] ultimate egoistic hedonism. Not only does everyone, insofar as they are rational and know what they are doing, act in such a way as to maximize their own happiness (and, of course, insofar as they are not rational or do not know what they are doing, to that extent they cannot be considered to act), but it is morally good that they do so. Whilst not obviously having advanced Mill towards Utilitarianism—a disinterested hedonism—at least there is now no room for a charge of equivocation.

Assuming Mill has thus established ethical ultimate egoistic hedonism, it follows from the broadly-consequentialist nature of practical reason that the principle which it would be most rational for us to follow would be that which it was most reasonable for us to believe was most likely to maximize our own happiness. Is this the principle of acting with the direct aim of maximizing our own happiness? According to Mill, it is not. Mill argues that it is by making our direct aim the maximization of societal happiness that we most increase the chances of indirectly maximizing our own happiness. By non-ultimately desiring society's happiness, we best promote that which we ultimately desire—our own. This is Mill's second psychological premise and with it Mill completes his journey to Utilitarianism. It will form our focus for this section.

At one stage in his life, Mill famously suffered a nervous break-

[29] Those who build into the notion of the ethical a tendency to treat other people's interests on a par with one's own would view 'Ethical egoism' as a contradiction in terms. 'Ethical' in the sense operative here means merely subscribing to principles which one considers to bind universally, in this case the principle that everyone should ultimately prioritise their own happiness over that of everyone else.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

down. As the changes in his philosophy brought about by this crisis are pertinent to this premise as it moves him from ethical ultimate egoistic hedonism to ethical non-ultimate disinterested hedonism (i.e. Utilitarianism), his own account of the period from his autobiography is worth quoting at some length.

> The experiences of this period had two very marked effects on my opinions and character. In the first place, they led me to adopt a theory of life, very unlike that on which I had before acted, and having much in common with what at that time I certainly had never heard of, the anti-self-consciousness theory of Carlyle. I never, indeed, wavered in the conviction that happiness is the test of all rules of conduct, and the end of life. But I now thought that this end was only to be attained by not making it the direct end. Those only are happy (I thought) who have their minds fixed on some object other than their own happiness; on the happiness of others, on the improvement of mankind, even on some art or pursuit, followed not as a means, but as itself an ideal end. Aiming thus at something else, they find happiness by the way. ...The only chance is to treat, not happiness, but some other end external to it, as the purpose of life. Let your self-consciousness, your scrutiny, your self-interrogation, exhaust themselves on that; and if otherwise fortunately circumstanced, you will inhale happiness with the air you breath, without dwelling on it or thinking about it, without either forestalling it in imagination, or putting it to flight by fatal questioning. This theory now became the basis of my philosophy of life. And I still hold to it as the best theory for all those who have but a moderate degree of sensibility and of capacity for enjoyment, that is, for the great majority of mankind.'[30]

Our own individual happiness is the only thing which we desire ultimately and which it is rational for us to desire ultimately; and which we ought ultimately to be trying to maximize. However, the general thrust of this passage is clear: if we act with this knowledge in our minds, we will be less happy overall than if we fix our minds on some other end.[31] This is the 'paradox of hedonism'. Those who try simply to make themselves happy are less happy than those who forget about trying to make themselves happy and try to do something else instead. Of course, this will not be true of every 'something else' and the something else which Mill suggests as the best—i.e. most happiness enhancing for the individual—solution to this

[30] J. S. Mill, *Autobiography* (London: OUP, 1935), 113 and 120–1.
[31] See also: J. S. Mill, *Utilitarianism*, 30–1.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

'paradox' is fixation on making one's society happy.[32] Why is the pursuit of aggregate utility rather than some other pursuit (for example, those Mill himself mentions in his autobiography) the best means to the end of our happiness?

The picture Mill paints is this.[33] In our state of nature, we start with a natural feeling of sympathy—Seth calls this 'Mill's doctrine of sympathy'.[34] This 'feeling of unity with our fellow-creatures'[35], which is 'as much an ultimate fact of our nature, as care for ourselves'[36], forms 'a natural basis of sentiment for utilitarian morality'[37] in driving our attention outwards to the concerns of others. In *Civilization*, Mill sketches a naturalistic explanation for why nurture might reasonably be expected to improve on what nature provides: a tendency for co-operation, allowing individuals to subordinate their immediate advantage to that of the group, is a necessary condition for the development of civilisation and with civilization persons become more and more dependent upon one another; the interests of individuals become more and more intermingled.[38] (A similar biological explanation of our natural sympathy suggests itself. A general tendency to put other peoples' interests—particularly one's children—on a par with one's own will have great survival value for one's genes and—given Mill's first psychological premise—evolution will have ensured that an individual's happiness is correlated with whatever behaviour has great survival value for their genes.) As societies emerge, it increasingly becomes the case that the 'deeply rooted conception which every individual even now has of himself as a social being tends to make him feel it one of his natural wants that there should be harmony between his feelings and aims and those of his fellow creatures.'[39] Thus in the 'good state

[32] Railton distinguishes between Objective Hedonism—the view that 'one should follow the course of action which would in fact most contribute to one's happiness'—and Subjective Hedonism—the view that one should 'attempt to determine which act seems most likely to contribute optimally to one's happiness' (P. Railton, 'Alienation, Consequentialism, and the Demands of Morality', *Philosophy and Public Affairs*, 13, 1984). In Railton's terminology, according to Mill, 'Objective hedonism' does not lead to 'Subjective hedonism', it leads to Subjective Utilitarianism.

[33] See J. S. Mill, *Utilitarianism*, 58–60.

[34] J. Seth, 'The Alleged Fallacies in Mill's Utilitarianism', *The Philosophical Review*, 17, no. 5, Sept. 1908.

[35] J. S. Mill, *Utilitarianism*, 50.

[36] See J. S. Mill, 'Sedgwick's Discourse', *Collected Works*, vol. X, 60.

[37] J. S. Mill, *Utilitarianism*, 58.

[38] J. S. Mill, 'Civilization', *Collected Works*, vol. XVIII, 122ff.

[39] J. S. Mill, *Utilitarianism*, 62.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

of society and education' that characterizes civilization as Mill refers to it in the troublesome letter of 1868, 'every human being's happiness is a good to every other human being'.[40] Those who live in the good state of society and education that we may reasonably suppose characterized subscribers to *Fraser's Magazine* in 1861 (where *Utilitarianism* first appeared) will find that their natural sympathy, fostered by the social life and education provided to them by a civilization, has led them to desire that which their fellows ultimately desire—namely the happiness of society. Mill is writing for thoughtful readers and one's being a thoughtful reader sets conditions on the sort of world one lives in and the sort of person one is. To be reading *Utilitarianism*, one must be in a world where periodicals or books are produced and circulated; to think about its contents, one must be in a world where one has time for thoughts beyond those concerned with one's immediate physical survival. One's world and oneself must be, in other words, civilized in the minimal sense Mill thinks sufficient for the truth of his second psychological premise.

According to Mill, it follows from one's being civilized that when and only when one has developed the habits of thought such that one acts from a desire to maximize aggregate happiness will one have maximized one's expected individual happiness. This must be because following the Principle of Utility gives one a unique 'quality' of pleasure (i.e. one which is actually greater in quantity than the pleasures one would otherwise have enjoyed); or because one has a greater capacity for pleasure and following this principle does not decrease one's efficiency (measured as a percentage of capacity for happiness satisfied) to an extent that outweighs the gain in total capacity; or because following this principle increases one's efficiency; or for a combination of these reasons. We need not decide. Mill calls those with such characters 'virtuous' or 'noble' and thus talks with compassion of those who have not yet attained that level of character development which gives them a 'confirmed noble character', that is one which of habit pursues virtue for its own sake.[41]

The enlightened ultimate egoistic hedonist ought to develop a confirmed noble character, i.e. one that follows the Principle of Utility from habit, having realized that doing so is actually the best means to the end which he or she ultimately desires—the maximization of his or her own happiness. We thoughtful readers

[40] J. S. Mill, *Collected Works*, vol. XVI, 1414.

[41] See J. S. Mill, *A System of Logic*, vol. II, 530 and *Utilitarianism*, 21–2.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

are ultimate egoistic hedonists whom *Utilitarianism* has enlightened, so we ought to develop the confirmed noble character as developing such a character is the best means to the maximization of that which we ultimately desire, our own happiness. 'In other words, this state of the will is a means to good, not intrinsically a good; and does not contradict the doctrine that nothing is a good to human beings but insofar as it is either itself pleasurable, or a means of attaining pleasure or averting pain.'[42] Again, the only question which we can properly raise over this stage in Mill's argument is over his assumption that the empirical evidence needed to establish this contingent matter of fact is indeed available. Does Mill himself ever present any? He does not. Are those who follow the Principle of Utility from a confirmed noble character happier people than others? Again, with Mill, I leave it to the thoughtful reader to decide.

V. Conclusion

Let us conclude, first by summarizing our reconstruction of Mill's proof, stripping it down so that the nuts and bolts may be seen most clearly.

(1) *Ought implies can*. This premise in this context, although arguable as is everything in this field, is as innocuous a starting point as one could ever hope to encounter for an argument in moral philosophy and (2), *if one ought to desire a thing ultimately, that implies that one can desire that thing ultimately*, follows from it without difficulty. (3) *One must desire something ultimately and the only thing that one can desire ultimately is one's own happiness*. The second half of this is ultimate egoistic psychological hedonism, the first substantial and controversial psychological premise of Mill's argument. The first half is a metaphysical necessity: given a desire theory of action, then were it not true, there could not be agency. The second half is an empirical or 'is-type' premise. One might most persuasively argue for the second half by encouraging candid self-reflection. A consideration of what strikes one as the only natural stopping point for subsumptive explanations of one's desires makes the premise not wholly implausible. According to Mill, in the vast majority of cases, honest introspection, if given enough time, would reveal that the ultimate explanation of why one desired any-

[42] J. S. Mill, *Utilitarianism*, 76–7.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

thing which was not one's own happiness would be that one supposed that that thing would make one happy and to the extent that one became aware of the fact that that a particular thing would not make one happy, one would cease to desire it. One would come to the opinion that to desire a thing ultimately and to consider it pleasant were one and the same thing. Therefore, (4) *one's own happiness is the only thing that one ought to desire ultimately,* follows from 2 and 3.

(5) *One is rational insofar as one pursues the policy which it is rational to believe is most likely to realize that which is the only thing which one ought to desire ultimately*. This follows from Mill's uncontroversial metaethical assumption that practical rationality is—in the broadest of senses—consequentialist. If one end is the only end which it is rational to desire ultimately, then the only means it can be rational to employ will be those means which it is rational to believe are most likely to be the most successful means to that end. Thus, (6) *One is rational insofar as one pursues the policy which it is rational to believe is most likely to maximize one's own happiness*. This then follows from 4 and 5. (7) *A policy of acting with the explicit aim of maximizing society's happiness* (*the policy followed most effectively by those with a 'confirmed noble character*') *is more likely to have the consequence of maximizing one's own happiness than is any other policy, including a policy of acting with the explicit aim of maximizing one's own happiness*. This is the second of Mill's psychological premises. It is made plausible, at least to some extent, from introspection akin to Mill's reflections on the crisis in his mental history. It is made more plausible by arguments to the effect that although the fact that the Principle of Utility is that principle the adoption of which most successfully maximizes an individual's happiness depends on the contingent psychology of the individual, this contingent psychology may reasonably be supposed to be materially necessitated in the relevant respects by the fact that the individual has grown up in a society. As it happens then—certainly in those countries which may be called civilized and of people who are not in mental hospitals—people are such that once it has become apparent to them in a particular situation of choice that a certain action is that which will maximize the aggregate happiness of their society, they will not be as happy if they do any other action. Further, they will only be as happy as they can be if they think of themselves as doing that which is most likely to promote the aggregate happiness *because* it is that which is most likely to promote the aggregate happiness—if they are of confirmed noble character. Therefore, at least within this domain, (8) *one is rational insofar as one directly aims at*

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

the maximisation of society's happiness. This follows from 6 and 7, as it is the behaviour which it is reasonable to believe is most likely to be the most successful means to the end which it is most rational to have. Acting always with the intention of producing the maximum aggregate happiness is following the Principle of Utility, which is being an Act-Utilitarian. So, Mill hopes, we may finish by saying that *Utilitarianism*'s thoughtful reader ought to conclude that (9) *one is rational insofar as one follows the Principle of Utility, i.e. insofar as one is an Act-Utilitarian* (from 8).

Unfortunately, the argument does not end there.

At the end of section IV, I said that the only question which we could properly raise over the last stage of Mill's argument was over whether, as an empirical matter of fact, those who thus follow the Principle of Utility are happier than those who do not. Another question which one might properly raise—and indeed which would be inescapable had one been convinced by Mill's argument of the rationality of desiring (non-ultimately) the end which the Principle of Utility set—would be 'Are those who thus follow the Principle of Utility most conducive to society's happiness?' Finally then, I shall sketch an argument for thinking that Mill's argument drives one to a Rule-of-thumb-Utilitarianism, which in turn drives one to adopt a confirmed non-Utilitarian character.[43]

A simple Act-Utilitarianism inevitably collapses into Rule-of-thumb-Utilitarianism as to decide on the worth of a particular act one must use general principles about what categories of acts have

[43] In the terminology of contemporary discussion, we might say that One-level Act-Utilitarianism collapses into a Two-level version. The question which we are left with is whether this in turn 'collapses'. In the main text, I sketch an argument for thinking that it at least collapses in the sense of necessitating that one adopt a non-Utilitarian character and possibly in the sense of necessitating that one adopt a non-narrowly-Consequentialist character. For a description of Two-level Utilitarianism see R. M. Hare's *Moral Thinking* (Oxford: OUP, 1981) and, for an earlier version, his 'Ethical Theory and Utilitarianism' in A. Sen and B. Williams (eds), *Utilitarianism and Beyond* (Cambridge: CUP, 1982). Williams' objection that Two-level Utilitarianism is psychologically unstable—leading one either to abandon Act-Utilitarianism or the second level intuitions—may be found in his *Ethics and the Limits of Philosophy* (London: Fontana, 1987) and in his 'The Structure of Hare's Theory' in D. Seanor and N. Fotion (eds) *Hare and Critics* (Oxford: OUP, 1988), pages 188–92. Whether Williams' objection works is of course an empirical matter; whether, given that it works, it is better in Utilitarian terms for one to abandon Act-Utilitarianism rather than the second level intuitions is another empirical matter.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

had what sorts of consequences in the past and subsume the particular act being considered under one of these categories, i.e. one must follow rules. Supposing then that one follows Mill's argument but modifies his conclusion by accepting Rule-of-thumb-Utilitarianism as the best policy for achieving one's penultimate end—societal happiness—and thus one's ultimate end—one's own happiness. So far then, Mill might be willing to accept one's modifications.[44] One problem for then thinking of Mill as having proved to one the worth of being a Utilitarian is that the rules one would be justified in following would contain no mention of the maximisation of societal happiness. They would simply prohibit things like harvesting people's organs without their consent; hanging those whom one knows to be innocent; et cetera. Furthermore, it could be argued that the ends which one should set oneself with these rules, whilst consequentialist in the very broad sense which we introduced, would not be consequentialist in the more usual and narrower sense which would allow a contrast to be drawn between consequentialism and deontology. Just as one would always be unjustified in deciding to perform the calculation required to collapse one's Rule-of-thumb-Utilitarianism back into an Act-Utilitarianism (because the expected utility of such calculations would always be less than the expected utility of simply following the rule), so—it could be argued for reasons related to 'integrity'—one would always, for narrowly-consequentialist reasons, be unjustified in deciding to perform narrowly-consequentialist calculations. One would always be justified in setting oneself ends which were 'agent-relative' to oneself—in other words, in thinking deontologically. One would be obliged to think things like '*I* must never harvest people's organs without their consent'; '*I* must never hang those I know to be innocent'; et cetera: in other words, one would be under an obligation to develop a confirmed *deontological* character.

Of course the argument of the previous paragraph is barely even

[44] After all, Mill himself says to those who would say that Utilitarianism is too exacting a standard—requiring agents always to consider the general happiness—that 'this is to mistake the very meaning of a standard of morals, and confound the rule of action with the motive of it … it is a misapprehension of the utilitarian mode of thought to conceive it as implying that people should fix their minds upon so wide a generality as the world, or society at large.' (See J. S. Mill, *Utilitarianism*, 33–34). Compare D. Hume, *A Treatise of Human Nature* (Oxford: Clarendon, 1990), Book III, Part I, sec ii, pages 483–4. And Mill himself is sometimes interpreted as suggesting a version of Two-level Utilitarianism in his discussion of secondary principles in *Utilitarianism*, 44–5.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press

a sketch. But if even the first half of it correct, then we may consider Mill as in the predicament of which he himself speaks in the passage with which we commenced our discussion: 'drawn in contrary directions by the two noblest of all objects of pursuit, truth, and the general good.' If the truth is, as I have argued Mill believes it is, that the only thing ultimately worth maximizing is one's own happiness and that the best way to do this is to aim at maximizing the happiness of one's society, then initially it does indeed seem as if one ought to follow the Principle of Utility and thus be straightforwardly an Act-Utilitarian. However, when reflection on how one might best maximize the happiness of one's society reveals the further truth that the best policy to adopt is that of following rules which make no mention of utility, then, in following these rules, one cannot be justified in developing a character which generally thinks in a distinctively Utilitarian way—because one will know that the expected utility of doing so is less than that of not doing so. That it is these rules that one ought to follow will lie in the fact that following them maximizes relative to a policy of calculating in each individual case or following others: however, the rules themselves will not make mention of happiness (or indeed if my 'integrity' point is sound any other 'state of affairs' in the non-agent-relative sense of 'state of affairs' that one is trying to realize in any narrowly-consequentialist system) and following them will thus entail that one ends up trying to adopt a character which acts from distinctively non-Utilitarian motives (and perhaps even non-narrowly-Consequentialist ones). Of course, Mill could block this step of the argument by appropriately nuancing his second psychological premise: he might suggest that—contra 'integrity' objections—the happiest are those who whilst normally acting on common-sense moral dispositions—ones which may perhaps be characterized deontologically—know on reflection that they are doing this for the narrowly-consequentialist (specifically, Utilitarian) benefits that doing so brings. Again then, in attempting to assess his proof we 'have evidently arrived at a question of fact and experience, dependent, like all similar questions, upon evidence.'[45] Were the evidence against such a putative fact be greater than that in its favour, Mill should accept that the pursuit of truth will ultimately lead one away from the self-conscious pursuit of utility and away from Utilitarianism however one construes it.[46]

St Peter's College, Oxford

[45] J. S. Mill, *Utilitarianism*, 73.

[46] My thoughts on this topic have benefited from discussion with Roger Crisp and from the advice of the editor.

https://doi.org/10.1017/S0031819102000359 Published online by Cambridge University Press