



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Mill's Intentions and Motives

**Citation for published version:**

Ridge, M 2002, 'Mill's Intentions and Motives', *Utilitas*, vol. 14, no. 01, pp. 54-70.  
<https://doi.org/10.1017/S0953820800003393>

**Digital Object Identifier (DOI):**

[10.1017/S0953820800003393](https://doi.org/10.1017/S0953820800003393)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Utilitas

**Publisher Rights Statement:**

© Ridge, M. (2002). Mill's Intentions and Motives. *Utilitas*, 14(01), 54-70. doi: 10.1017/S0953820800003393

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Utilitas

<http://journals.cambridge.org/UTI>

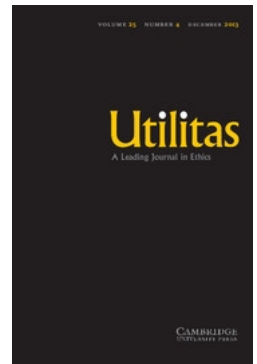
Additional services for **Utilitas**:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



---

## Mill's Intentions and Motives

Michael Ridge

Utilitas / Volume 14 / Issue 01 / March 2002, pp 54 - 70

DOI: 10.1017/S0953820800003393, Published online: 26 January 2009

**Link to this article:** [http://journals.cambridge.org/abstract\\_S0953820800003393](http://journals.cambridge.org/abstract_S0953820800003393)

### How to cite this article:

Michael Ridge (2002). Mill's Intentions and Motives. Utilitas, 14, pp 54-70 doi:10.1017/S0953820800003393

**Request Permissions :** [Click here](#)

# Mill's Intentions and Motives

MICHAEL RIDGE

*Australian National University, University of Edinburgh*

One might have thought that any right-thinking utilitarian would hold that motives and intentions are morally on a par, as either might influence the consequences of one's actions. However, in a neglected passage of *Utilitarianism*, John Stuart Mill claims that the rightness of an action depends 'entirely upon the intention' but does not at all depend upon the motive. In this paper I try to make sense of Mill's initially puzzling remarks about the relative importance of intentions and motives in a way that highlights the importance of other elements of his moral philosophy and action theory.

A full appreciation of John Stuart Mill's moral theory requires a proper understanding of his action theory. For in a puzzling and infrequently discussed passage in *Utilitarianism*, Mill claims that there is 'no point which utilitarian thinkers (and Bentham pre-eminently) have taken more pains to illustrate than' the distinction between intentions and motives.<sup>1</sup> Mill goes on to claim that while the rightness or wrongness of an action does not depend upon the agent's motive, it does depend 'entirely upon the intention.' In spite of the apparent importance of this distinction to his view, Mill does not explicitly return to it in *Utilitarianism*. The footnote in which he responds to Rev. J. Llewellyn Davies is the extent of his discussion of the distinction in that work. This presents sympathetic interpreters of Mill's moral theory with a problem. It would seem that for a utilitarian, motives and intentions are on a par with respect to the rightness of an action. The most straightforward understanding of utilitarianism makes the rightness or wrongness of an action a direct function of its consequences. Either intentions or motives may have some influence on the consequences of an action, but there is no obvious reason to suppose that intentions, but not motives, are especially strongly connected to the consequences of actions. It is therefore difficult to see why a utilitarian would think intentions, as opposed to motives, are uniquely important to the rightness or wrongness of an action. My aim here is to dissolve this puzzle, and show why it makes sense for Mill to put special weight on the intention, but not the motive, with which an action is done.

## I

An opponent, whose intellectual and moral fairness it is a pleasure to acknowledge (the Rev. J. Llewellyn Davies), has objected to this passage, saying,

<sup>1</sup> See John Stuart Mill, *Utilitarianism*, Indianapolis, 1979 [1861], p. 18. The footnote is one of two footnotes of any substantial length in the entire work, but does not appear in all contemporary versions of the book (it was added in Mill's second edition).

'Surely the rightness or wrongness of saving a man from drowning does depend very much upon the motive with which it is done. Suppose that a tyrant, when his enemy jumped into the sea to escape from him, saved him from drowning simply in order that he might inflict upon him more exquisite tortures, would it tend to clearness to speak of that action as a "morally right action"? Or suppose again, according to one of the stock illustrations of ethical inquiries, that a man betrayed a trust received from a friend, because the discharge of it would fatally injure that friend himself or someone belonging to him, would utilitarianism compel one to call the betrayal "a crime" as much as if it had been done from the meanest motive?'

I submit that he who saves another from drowning in order to kill him by torture afterwards does not differ only in motive from him who does the same thing from duty or benevolence; the act itself is different. The rescue of the man is, in the case supposed, only the necessary first step of an act far more atrocious than leaving him to drown would have been. Had Mr Davies said, 'The rightness or wrongness of saving a man from drowning does depend very much' – not upon the motive, but – 'upon the intention', no utilitarian would have differed from him. Mr Davies, by an oversight too common not to be quite venial, has in this case confounded the very different ideas of Motive and Intention. There is no point which utilitarian thinkers (and Bentham pre-eminently) have taken more pains to illustrate than this. The morality of the action depends entirely upon the intention – that is, upon what the agent wills to do. But the motive, that is, the feeling which makes him will so to do, if it makes no difference in the act, makes none in the morality: though it makes a great difference in our moral estimation of the agent, especially if it indicates a good or bad habitual disposition – a bent of character from which useful, or from which hurtful actions are likely to arise.<sup>2</sup>

This lengthy footnote comes in the context of Mill's having claimed that utilitarian moralists have 'gone beyond all others in affirming that the motive has nothing to do with the morality of the action, though much with the worth of the agent'. Reverend Davies presents Mill's claim with two counter-examples – that of a tyrant who saves a drowning man simply to torture him and the case of a man who betrays a trust received from a friend because keeping the trust would fatally injure that friend or someone dear to him. The idea seems to be that the tyrant's action is wrong in virtue of his motive while the man who betrays a trust deserves at least some exculpation in virtue of his motive. Mill argues that Davies has confused intentions with motives. According to Mill, the rightness or wrongness depends 'entirely upon the intention', while the motive, 'if it makes no difference in the act, makes none in the morality'.

What, though, is the distinction between intentions and motives? Since Mill refers to Bentham as having given the definitive account of the distinction, it makes sense to turn to Bentham to see what Mill has in mind. After all, Bentham devotes an entire chapter of *The Principles of Morals and Legislation* to motives (chapter X) and

<sup>2</sup> Ibid., p. 18.

another chapter to intentions (chapter VIII), and he discusses them in some detail in numerous other chapters as well. Bentham's most general characterization of a motive is 'any thing that can contribute to, give birth to, or even prevent, any kind of action'.<sup>3</sup> Bentham then distinguishes a wide variety of senses of 'motive' and its cognates, but for present purposes only two of those distinctions are important. First, there is the distinction between motives as prospects and motives *in esse*. A motive as prospect is that posterior to the action to which the agent looks forward. A motive *in esse* is an event which takes place upon a person's anticipating a motive as prospect. Second, there is the distinction between external and internal motives. An internal motive is internal in the sense of being a mental event; in particular, it is a perception of pleasure or pain which is 'looked upon as calculated to determine you to act in such or such a manner'. An external motive is an external event, 'the happening whereof is regarded as having a tendency to bring about such pleasure or such pain'.<sup>4</sup> Given these distinctions, it is easy to see which of Bentham's senses of 'motive' fits best with Mill's discussion. Recall Mill's claim that a motive is a feeling which makes one will something. Since a motive in Mill's sense causes one to will something, it must not be posterior to that action and hence must be a motive *in esse*, rather than a motive *qua* prospect. Since a motive in Mill's sense is a feeling, it clearly is an internal rather than an external motive. So Mill's conception of a motive maps nicely onto Bentham's notion of a internal motive *in esse*.

On Bentham's account, our intentions can regard either the action to be performed itself, or its consequences. This is meant to explain the way in which an action can be intentional even when its consequences are not, as when I intend to give my friend some news but do not realize that the news will upset her, so that I do not intend one of the consequences of my intentional action.<sup>5</sup> More unorthodox is Bentham's distinction between consequences which are 'directly' intended and those which are only 'obliquely' intended. A consequence is directly intended when the prospect of that consequence is one of the 'links in the chain of causes' which led the agent to perform the action. So if the consideration of some prospective consequence of an action is part of what motivated me to perform that action, then I directly intended that consequence. By contrast, a consequence is obliquely intended if

<sup>3</sup> Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, New York, 1948 [1789], p. 97.

<sup>4</sup> *Ibid.*, p. 99.

<sup>5</sup> Bentham also gives an example in which, on his account, the consequences are intentional though the act is not intentional 'throughout'. See Bentham, p. 83.

the agent was aware of the consequence but the prospect of that consequence 'did not constitute a link in the aforesaid chain'.<sup>6</sup> So if I knew that hurting my friend was a likely consequence of giving her the news, but I was not at all motivated by the prospect of hurting her, then in giving her the news I only obliquely, and not directly, intended to hurt her.<sup>7</sup> It must be admitted that Bentham's notion of oblique intending is idiosyncratic. Nonetheless, this is rather clearly his meaning, and it would be unhelpful to quarrel over the word.

Does Mill, in the footnote from *Utilitarianism*, mean by 'intention' to include both direct and oblique intentions? He refers the reader to Bentham, which provides some evidence that he does, especially since he does not bother explicitly to invoke the distinction and say he is only talking about the one rather than the other. However, the examples from Revd Davies that Mill is discussing seem compatible with thinking that he only has in mind what Bentham would refer to as direct intentions, especially since this is arguably the more natural understanding of 'intention'. In the first case, the tyrant rather clearly has a direct intention to torture the man, as that is what moves him to act. Presumably, he also has a direct intention to save him as well, as he intentionally saves the man; of course this direct intention is purely instrumental, but the direct/oblique distinction cuts across the non-instrumental/instrumental distinction. It is less clear that his intention to then torture the man is direct; even though it is an action of his, it is also understood here as a consequence of the first action, as if it necessarily follows from the first. One might argue that the tyrant's intention to torture the man is direct under one description – as an action of his without reference to its etiology, and indirect under another description – as a consequence of his directly intended action of saving the man. The second case Davies provides is considerably less clear, since the nature of the fatal injury which would befall the person's friend if he breaks his trust is left open.

Fortunately, we have independent evidence that Mill understands 'intention' as including direct and indirect intentions. In commenting on Chapter XXV, 'Intention', of his father's (James Mill's), *Analysis of the Human Mind* (John Stuart), Mill remarks that,

<sup>6</sup> Ibid., p. 84.

<sup>7</sup> Oddly, Bentham characterizes the direct/oblique distinction as a distinction falling only within the category of intended consequences, but it seems straightforward and plausible enough to extend that distinction to intended actions themselves. If the realization that my action falls under a particular description – an insult, say – motivates me to perform the action, then the action is directly intentional under that description. If, by contrast, I am aware that my action falls under that description but am not at all motivated by that fact, then the action is only obliquely intentional under that description.

*Intention*, when we are said to intend the consequences of our actions, means the foresight, or expectation of those consequences, which is a totally different thing from desiring them. The particular consequences in question, though foreseen, may be disagreeable to us: the act may be done for the sake of other consequences. Intention, and motive, are two very different things. But it is the intention, that is, the foresight of consequences, which constitutes the moral rightness or wrongness of the act. Which among the many consequences of a crime, are those, foresight of which constitutes guilt, and non-foresight entitles to acquittal, depends upon the particular nature of the case. We may say generally, that it is the hurtful consequences.<sup>8</sup>

In this passage Mill not only explicitly embraces Bentham's very broad notion of intention, he also explicitly connects his embracing that notion to his claim that it is the intention, but not the motive, 'which constitutes the moral rightness or wrongness of the act'. This makes the conclusion that it is this sense of 'intention' that Mill had in mind in the footnote in *Utilitarianism* in which he makes this very same claim almost irresistible.

At this point, we have some idea of how Mill understands the distinction between intentions and motives. An intention is either an agent's aiming to do something, or the agent's foreseeing a consequence of what the agent aim to do.<sup>9</sup> To aim to do something, in this sense, one must not simply desire to do it but must also be committed to doing it, which will at the very least involve not seeing it as beyond one's control.<sup>10</sup> A motive, by contrast, is a feeling, and such feelings may (though they need not, and do not, always do so) cause one to intend to bring about that state of affairs, the contemplation of which gives one the feeling. A more intuitive way of putting this point would be to say that such feelings may cause one to aim to bring about the intentional object of the feeling, but Mill does not clearly think that feelings have intentional objects. He may very well follow Hume in supposing that they are 'original existences,' having no intentional objects.<sup>11</sup>

Mill's language is admittedly a bit strained here, in that he seems to equate 'intending' with 'willing' and it does seem odd to claim that someone wills something when she merely foresees it as a consequence of her action.<sup>12</sup> Still, this pretty clearly seems to be what Mill has in

<sup>8</sup> *Collected Works of John Stuart Mill*, ed. John Robson, Toronto, 1961–91, xxxi. 252 f.

<sup>9</sup> In the footnote in question, Mill actually says that the intention is 'what the agent wills,' making an intention an external state of affairs rather than a mental state, but presumably he would simply distinguish intentions *qua* what is intended from intentions *qua* acts of intending.

<sup>10</sup> It may require the stronger condition that you see it as, to some degree, within your control.

<sup>11</sup> David Hume, *A Treatise of Human Nature*, ed. L. A. Selby Bigge, rev. P. H. Nidditch, 2nd edn., Oxford, 1978 [1739], bks. I and II.

<sup>12</sup> Thanks to Jonathan Dancy for useful discussion here.

mind. Nor is Mill's language completely counter-intuitive anyway. For if one is aware of some consequence of one's action and still wills the action then it is intuitive to claim that one is 'willing to allow that consequence to happen,' which is not a million miles from claiming that one wills the consequence. Fortunately, whether Mill's language fits especially well with ordinary language is not really crucial to his position so long as the theory he intends is plausible when properly interpreted. Having now a reasonably clear picture of the distinction Mill is drawing, we are in a better position to see why he goes on to say that intentions, but not motives, 'entirely determine' the rightness of an agent's action.

## II

Mill does give us one clue as to why he draws this connection between an agent's intentions and the rightness or wrongness of her action. He tells us that when the intention differs, 'the act itself is different'. This suggests the following account.<sup>13</sup> Before we determine whether a given action is right or wrong, we need to know just what action is in question. Hence, determining the rightness of a given action requires having some way of individuating actions. Mill's suggestion seems to be that actions should be individuated, at least in part, in terms of their intentions. On this account, if an agent had acted with different intentions, she would have performed a different action. By contrast, Mill is suggesting, we should not individuate actions in terms of the motives which give rise to the intentions. If the intentions and the external bodily movements (or, in the case of a mental action, the train of thought, etc.) remain the same, we have the same action.<sup>14</sup> In the case of the tyrant, Mill's idea seems to be that the relevant action is a complex whole – the action of rescuing one's enemy in order to torture and kill him.<sup>15</sup> This seems to be the relevant action in the case because the agent intends this combination (actually, the case is arguably underdescribed on this score, but we can assume that Mill intended this reading). Whereas if the agent had intended to rescue his enemy and reform him, the action might not be wrong. On this account,

<sup>13</sup> Though there are other ways of reading this sentence (for some discussion, see Jonathan Dancy, 'Mill's Puzzling Footnote,' *Utilitas* xii (2000), 220), I think the account offered in the text makes the most sense, all things considered.

<sup>14</sup> This interpretation receives further support from a passage in Mill's *A System of Logic*: 'Now what is an action? Not one thing, but a series of two things: the state of mind called a volition, followed by an effect. The volition or intention to produce the effect, is one thing; the effect produced in consequence of the intention is another thing; the two together constitute the action.' (Mill, *A System of Logic*, London, 1875, bk. 1, ch. Iii, sect. 5.)

<sup>15</sup> Jonathan Dancy advocates a similar reading of the tyrant case. See Dancy, 221.



motives can at best be indirectly relevant to the individuation of actions by having an influence on the agent's intentions, whereas intentions provide constitutive grounds for the individuation of actions.

This certainly is part of what Mill has in mind, but it cannot be the full story. It is true that we need some way of individuating actions, and that doing it in terms of the agent's intentions is one way of doing so. The problem is that it is obviously not the *only* way of individuating actions. We could instead take a more coarse-grained approach, and individuate actions in terms of bodily movements. On this more coarse-grained approach, someone who goes through exactly the same external bodily movements as me performs the same action as me, even if he or she has wildly different intentions. Or we could take a more fine-grained approach, individuating actions in terms of external bodily movements, intentions, and motives. On this account, a difference in motives which cause the intentions will entail that we have different actions, even if the intentions are identical. Or we could individuate actions in terms of external bodily movements and the feelings that gave rise to them. Why should we favour Mill's approach over these alternatives? Mill could respond by claiming that actions are just defined in terms of intentions but not motives, but such a stipulative answer hardly does justice to the force of the question. For the question can then be reformulated in terms of why that particular definition is the one we should embrace for purposes of moral theory.<sup>16</sup> Moreover, the question is especially pressing for a utilitarian. Intuitively, utilitarianism makes the rightness of an action a direct function of its consequences. An agent's intentions may, in a given case, have some effect on the consequences of her action. However, an agent's motives may also, in a given case, have some effect on those consequences. For example, if I give to charity simply because I am motivated by greed and a contemplation of tax benefits, and if others can perceive this, my acting from those motives may have significant consequences – it might set a bad example. In terms of what is important to a utilitarian – promoting good consequences – motives and intentions seem to be on a par.

<sup>16</sup> An analogy might help make this point more clear – there are different ways we might go about individuating books. Suppose the only thing on my desk is a copy of Hume's *Treatise of Human Nature*. On one way of individuating books, there are three books on my desk, since the *Treatise* is divided into three books. On another way of individuating books, there is only one book on my desk, since the *Treatise* is itself just one book. Either way of individuating books seems reasonable, so long as we are clear about which we are using – nothing much needs hang on which way we go. However, Mill is going to make whether an agent performs an action which is right (or which is wrong) hang upon individuating actions in terms of intentions, but not in terms of motives. Since it is quite significant whether we end up saying that an agent performed a right action (or that she performed a wrong action), we need an account of why we should individuate actions, and hence attribute rightness and wrongness, as Mill suggests.

The key to dissolving this puzzle is first to see what Mill means in calling an action right or wrong. Utilitarians typically are understood as defining right action as an action, of those available to the agent, which maximizes aggregate happiness, while wrong action is defined as any of the other actions available to the agent. Mill comes close to endorsing these definitions in chapter II, where he notes that 'the creed which accepts as the foundation of morals "utility" or "the greatest happiness principle" holds that actions are right in proportion as they tend to promote happiness; wrong as they tend to promote the reverse of happiness'.<sup>17</sup> However, the characterization of right and wrong action in chapter II is a rough and ready account meant to get the general idea of utilitarianism across; it does not represent Mill's most refined and considered view of the concepts of morally right and morally wrong action. For his considered view of those concepts, we must turn to chapter V of *Utilitarianism*. The primary aim of chapter V of *Utilitarianism* is to deal with the apparent problem that utilitarianism conflicts with justice. In the course of dealing with that objection, Mill also presents in a rather unequivocal manner his account of what it is for someone to be under a duty. To be under a duty is intuitively to be obligated in some way to do something. Mill argues that we can capture this notion of duty in terms of one's being rightfully compelled to fulfil it. Importantly, a given action might be recommended by considerations of utility – it might maximize aggregate happiness – but not be something that the agent is under a duty to do, simply because it would not maximize aggregate happiness for his or her community to adopt laws, conventions or other practices which punish people for not performing such actions. In this context, Mill distinguishes the expedient actions – those which maximize aggregate utility and hence are recommended by utilitarianism and ought to be done – from the morally required actions – those which are such that an agent ought to be compelled to perform them.<sup>18</sup> Finally, Mill argues that justice is a proper subset of the morally required – he argues that only those requirements crucial to any kind of flourishing community, and that involve rights, are requirements of justice.<sup>19</sup> In articulating the distinction between what is recommended by expedience and what is morally required, Mill offers the following definitions of 'right' and 'wrong':

<sup>17</sup> Mill, *Utilitarianism*, p. 7.

<sup>18</sup> Mill seems to suppose that the morally required is a proper subset of the expedient, but this would require some argument. For present purposes, I overlook this subtlety. For useful discussion, see John Skorupski, *John Stuart Mill*, London, 1989, p. 324.

<sup>19</sup> For a discussion of some problems with Mill's claim that duties involving rights are all duties of justice, see David Lyons, *Rights, Welfare, and Mill's Moral Theory*, New York, 1994.

For the truth is that the idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong. We do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other for doing it – if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience. This seems the real turning point of the distinction between morality and simple expediency ... there is no doubt that this distinction lies at the bottom of the notions of right and wrong; that we call any conduct wrong, or employ, instead, some other term of dislike or disparagement, according as we think that the person ought, or ought not, to be punished for it; and we say it would be right to do so and so, or merely that it would be desirable or laudable, according as we would wish to see the person whom it concerns compelled, or only persuaded and exhorted, to act in that manner.<sup>20</sup>

Mill's considered view, then, is that to call an action right is to say that it ought to be compelled and to call an action wrong is to say that its performance ought to be punished – either by the state, one's fellows, or one's own conscience. Rightness and wrongness are reduced to 'to-be-compelledness' and 'to-be-punishedness' where these claims bottom out in appeals to the utility of the relevant practices of compulsion and punishment.<sup>21</sup> That these claims bottom out in appeals to utility is not,

<sup>20</sup> Mill, *Utilitarianism*, p. 47.

<sup>21</sup> Filling in the details of this account in a way that makes the resulting view plausible is no trivial task. For the question naturally arises as to whether we should call an action wrong on the grounds that an ideal set of rules would prohibit it, and call for punishment of the agent, or whether some non-ideal set of rules is the relevant one. If we favour ideal rules then Mill's analysis warrants a charge of implausible utopianism. Perhaps in an ideal world there would be a rule requiring nations never to go to war but this would hardly make it plausible to suppose that in the actual world, where no such rule is in play, that it would always be wrong to go to war. Fred Berger has made a similar point against the idealized reading; see Fred Berger, *Happiness, Justice, and Freedom*, Berkeley, 1984, p. 112. Interestingly, structurally similar charges are often brought against Kant; see Thomas E. Hill, Jr., 'Kant's Utopianism', *Akten des 4. Internationalen Kant-Kongresses, Mainz, Dordrecht, Holland, 1972*, pp. 307–15, repr. in his *Dignity and Practical Reason*, Ithaca, 1992. The basic problem in each case seems to be a failure to distinguish what Rawls refers to as ideal theory from non-ideal theory. If, however, we are not to analyse rightness and wrongness in terms of ideal rules then it is unclear which of the many possible non-ideal rules are the appropriate ones and why they are appropriate in the case at hand. The only obviously salient candidate would be whatever rules are actually in play, but that would be unduly conservative, making it impossible to criticize existing practices on moral grounds (as opposed to grounds of expedience). Further, switching away from rules which would be ideal in the circumstances seems to lose the sense in which one really ought to be punished, since the 'ought' is neither the maximizing 'ought' of expedience nor the 'ought' of morality, if we assume those are distinct 'ought's. Mill seems to need a middle-ground position here, according to which we always take some of our present practices to be justified to some significant degree though any one of them could in principle be called into question. For useful discussion of this sort of holistic interpretation of Mill on this point, see Skorupski, pp. 318–20. The problem is a very tangled one, and I lack the space here to try to resolve it. My suspicion, however, is that a greater emphasis on the rules of one's own conscience, over which one has relatively direct control, would make it considerably easier to defend the ideal conception without falling prey to objections of utopianism. In this respect, I am in full agreement with David Lyons, who also argues that greater emphasis should be placed

however, an analytic truth for Mill, as I read him. One gets that conclusion only when one adds as a substantive premise that the principle of utility is the correct principle of expedience, accurately characterizing what is worth promoting for its own sake.<sup>22</sup>

Granted that this is Mill's considered view, one might reasonably worry whether it is what he had in mind in chapter II, and in particular in the footnote in which he responds to Revd Davies. For now I want to bracket that worry, and see how Mill's account would fare if this were what he meant in chapter II. I shall argue that if this were what Mill meant by 'morally right' and 'morally wrong' in that footnote, it would be easy to make sense of Mill's position. Let us suppose that Mill's response to Revd Davies rests on the thesis that, in determining whether an action should be punished or compelled, we should individuate actions in terms of intentions, rather than motives.<sup>23</sup> From a

upon the sanctions of conscience, though I disagree with the details of his proposal (see the following footnote).

<sup>22</sup> For a defence of this way of sorting out what is analytic and what is not in Mill, see David Lyons, *Rights, Welfare, and Mill's Moral Theory*. In elaborating this view, however, Lyons attributes a view to Mill which seems to me to be unwarranted. Lyons distinguishes himself from D. G. Brown, who claims that Mill's view was that it is analytic that any action which is wrong is one which merits some sort of external sanction. See D. G. Brown, 'Mill on Liberty and Morality,' *Philosophical Review*, lxxxi (1972). I fully agree with Lyons on this point. Lyons, however, goes on instead to attribute to Mill the view that 'wrongness is conceptually connected with justified guilt feelings, but only contingently or synthetically connected with external sanctions' (Lyons, p. 130). The text instead seems to suggest that Mill did think there was an analytic connection between wrongness and external sanctions, though certainly not the strong one advocated by Brown, according to which the appropriateness of external sanctions is *necessary* for an action's counting as wrong. Rather, Mill seems to have thought that it was a conceptual truth that the appropriateness of external sanctions was sufficient, though not necessary, for an action to count as wrong. More specifically, he held that the concept of wrongness is disjunctive (though the disjuncts obviously have something in common – we need not suppose it is disjunctive in some non-natural way): an action is wrong if and only if either (a) external sanctions (legal or non-legal) ought to be applied to it, or (b) internal sanctions ought to be applied to it. Mill's chapter V definition is itself explicitly disjunctive in form (to be wrong is to be such that one ought to be punished 'in some way or other for it'). It is also true, as Lyons no doubt would emphasize, that cases in which external but not internal sanctions are very rare indeed, but they do not seem to be impossible, and we can even imagine circumstances in which they might be relatively common. I see no reason to suppose that Mill would hesitate to call such at least possible actions wrong, given his explicit definition of wrongness, anyway.

<sup>23</sup> Roger Crisp makes a very similar suggestion in his commentary on Mill's *Utilitarianism* (Oxford, 1998). Crisp remarks that, 'these claims about intention are best interpreted as advice about when to blame and praise agents, rather than as attempts to offer a "standard of morals"' (Crisp, p. 122). Crisp does not, however, argue for this plausible interpretative suggestion. In effect, I am trying to vindicate Crisp's claim that Mill has in mind the appropriateness of blame and praise in this passage, though I would (on at least some readings) dissent from Crisp's claim that this is to be opposed to an attempt to offer a 'standard of morals'. For on my reading Mill would hold that to give an account of when blame/praise is appropriate is, in one sense, to give a standard of morals. This standard is, admittedly, based on the principle of utility, the concept of a duty, and various empirical hypotheses and so is a derivative standard. So if by

utilitarian perspective, deterrence is the primary rationale for punishment. Consider whether punishing someone for an unintentional consequence would further this goal. We need to separate two kinds of cases, cases of negligence and other cases. Take first the cases where there is no negligence. An agent took due care and was in no way negligent, but was nonetheless totally unaware that a consequence of her action would be that an innocent person would die. In this sort of case it is plausible to suppose that punishing the person will not further the general aim of deterring the criminal and others from engaging in such crimes. For, by stipulation, the agent already took due care and was not negligent, so it is not as if we need to give her an incentive to be more careful in the future. Nor is there any sense to be made of the idea that we are giving her an incentive not unwittingly to cause the deaths of innocent people. There is no way you can have a direct incentive not to do anything unwittingly, since by stipulation you would be unwitting of what you are doing. We can, at most, give a person an incentive to take steps to avoid being unwitting of certain consequences of her actions, but we have already seen that this is not necessary in the case at hand; by stipulation, the agent was not negligent – she already shows due care and does not need to be more careful in the future (at least, for all we know about her in virtue of this particular case, she does not need such an incentive). So in the case of the non-negligent person, punishment would not further the general utilitarian aims which justify punishment. Moreover, punishment obviously would inflict some harm on the person we punish, so if no such justification is forthcoming, the proper conclusion is that punishing people for their unintentional crimes, when they are not negligent is, as a general practice anyway, not justified.<sup>24</sup>

Here, of course, it is very important that in Mill's sense what one intends includes what one foresees but does not act for the sake of. In the more usual sense of 'intend', one might not intend some consequence but full well know that it is a consequence of one's action, and in *that* sense of 'intend' (the sense at work in the 'Doctrine of Double

'standard of morals' Crisp had in mind the ultimate standard of morality (and he probably does) then he is right that Mill is not trying to do that here, and we do not disagree. I labour this point only because I think it is important to make sense of the fact that Mill does claim in the text that the intention entirely determines the morality of the action.

<sup>24</sup> This is, at any rate, Jeremy Bentham's argument in *Principles*, ch. XIII, 10–11, against strict liability, and there is no obvious reason to suppose Mill disagrees with it. I do not mean to imply that the argument is sound. One might reasonably worry that it overlooks (1) the general deterrent effects of such punishments, and (2) the possible general benefit of people choosing not to engage in activities they cannot guarantee will not have consequences for which they might be held liable. Since my primary aim here is the exegetical one of making sense of Mill's view, I need not worry over whether Bentham's argument, which I strongly suspect Mill would endorse, is in fact sound.

Effect' literature) it is true that punishing people for unintended (but foreseen) consequences could serve to deter actions with those consequences. So the unorthodox conception of 'intention' at work in Mill, as including Bentham's 'oblique' intentions, is important to keep in mind at this point. It would, however, miss the point to quarrel over whether this unorthodox notion of 'intend' fits well with ordinary language. For the point is that 'intention' in this context might even be a quasi-technical term, which is meant to serve a role in a theory – the role of picking out the mental state(s) which are relevant to appropriate blame and punishment. If the term has a different meaning in ordinary language, that would not undermine a theory which used the term in this somewhat unorthodox way, so long as its meaning was clearly specified. This leaves us with the case in which a person unwittingly causes some bad consequence in which the person is unwitting of the bad consequence due to negligence of some sort. In this case, the agent's intentions may turn out to be irrelevant to whether they are punished, but we cannot know this in advance of individuating the action. Hence we should individuate actions in terms of intentions, rather than motives, and deal with cases of negligence as they arise.

So far we have seen how a utilitarian might plausibly argue that individuating actions in terms of intentions makes sense when trying to decide which actions are right and which are wrong, where 'right' and 'wrong' are understood as meaning 'worthy of compulsion' and 'worthy of punishment'. This leaves open the possibility that we should individuate actions in terms of intentions and motives, but Mill rejects this approach. What utilitarian grounds could he have for rejecting it? Reflection upon the likely consequences of a set of conventions which do individuate actions in this way, and those that do not, illustrates how such an argument might go. Let us call the rules that a society would be justified on utilitarian grounds in legally enforcing regardless of the agent's motives that society's L-Rules. The question, then, is whether the L-Rules should, in individuating actions, be extended to take into account the agent's motives as well. There are at least two ways in which we might do this. First, we might change the rules so that they call for punishing someone even when they follow the L-Rules but when they follow them from a bad motive or fail to follow them from a good motive (where what counts as a good or bad motive is given a utilitarian gloss). Second, we might extend the rules so that they do not require the punishment of those who knowingly break the L-Rules from a good motive. Under the first scheme in which motives are made relevant to whether someone should be punished, then Kant's famous shopkeeper, who gives correct change simply to maximize profits but not from duty, might be punished *even though* he follows the letter of the law. Our natural reaction to such a legal

scheme is that it is unjust, but for a utilitarian this would be beside the point if such a scheme would be optimific.

Consider what we might gain in the way of utility by adopting a scheme in which people are punished for their motives as well as their actions, and then turn to the costs such a scheme would impose. The only apparent benefit of such a system when compared with the more usual system of punishing someone just in case she breaks the community's L-Rules and regardless of her motives, is that there would be a somewhat greater incentive for people to acquire good motives and thereby avoid such punishment. However, there is already a strong incentive for people to acquire such motives, in so far as doing so is quite often necessary simply to be reliable at obeying the community's L-Rules understood independently of motives. In so far as people are rational and self-aware, the main people who would be given a further incentive to acquire such motives would be those who are reliable followers of the L-Rules in spite of having bad motives, but it is not at all clear what advantage there is in getting such people to adopt good motives – by hypothesis they obey the relevant rules anyway. There may be some marginal utility in giving such people a further incentive. For example, they might become more likely to be nice to people in cases in which the L-Rules do not require anything, though they might just as easily become more nasty in such cases having focused all their positive energy into following the L-Rules from good motives. So it would seem that the benefits of such a scheme are, at best, speculative and modest.

By contrast, there are apparently some not so speculative and potentially enormous costs associated with any such scheme. First, it would divert considerable resources from other, potentially more productive, outlets. Determining someone's motives in a given case will be very difficult and require a non-trivial investment of resources, as will the need to punish more people, at least at first. Indeed, given that almost anyone sometimes does the right thing for the wrong reason, it might seem that such a system would, if fully and ideally implemented, incarcerate or perhaps in some more informal way punish almost the entire population! Second, such a system might encourage fraud, threats, corruption, and witch-hunts. Given the subjectivity and uncertainty involved in the determination of a person's motives, people may be more likely to lodge frivolous complaints against people who most likely did not act from bad motives or fail to act from good ones. Third, the increase in punishment is, for a utilitarian, itself bad in so far as it causes pain or prevents pleasure for those punished. Fourth, such a scheme would in almost any imaginable context be vastly unpopular, reducing respect for the law and conventions of morality in general. Fifth, there is a danger of tyranny involved in giving too much

power to the state, as such a scheme seems to do. Mill, of course, is quite sensitive to this kind of worry, remarking, for example, that we are not without reason 'afraid of trusting the magistrate with so unlimited an amount of power over individuals'.<sup>25</sup> Sixth, even if such a system did not degenerate into outright tyranny, there is reason to suppose that it is not in general a good idea to have the law 'interfere with the whole detail of private life'.<sup>26</sup> Indeed, *On Liberty* is rife with arguments for this conclusion. In light of these considerations and the seemingly negligible and speculative benefits of such a scheme, the utilitarian is justified in thinking that we ought not adopt a scheme which individuates actions in terms of motives.

What about making motives relevant in the other way suggested above – not by punishing people for bad motives, but by not punishing them so long as they have good motives? It would seem equally apparent, upon reflection, that such a system is unlikely to maximize utility, once again granting certain relatively uncontroversial empirical assumptions. What benefit would there be in such a system? There would presumably be fewer people punished, which is good on utilitarian grounds, given that punishment reduces the happiness of the punished. Other than this there is no apparent benefit to such a system. There are, however, considerable costs. First, there would again be the diversion of resources inherent in sorting out people's motives, which almost certainly would be greater than the savings in reduced costs of incarceration. Second, such a system would presumably reduce the deterrent effect of punishment substantially, as you could hope to do the wrong thing but convince a jury, or those people who might impose more informal, private sanctions against you, that you did it from a good motive. Further, there is the need to deter moral fanatics who are wrongly convinced that morality requires something perverse, say, that they exterminate the members of some despised minority group. Such a person might believe that exterminating others would maximize utility, in which case their intention is such that if their belief was not formed in a negligent way (hard to imagine in this particular case unless the subject is mentally impaired) they would not be punished even by the kind of conventions Mill would defend. However, they need not think that the killing of the members of the relevant minority group would maximize utility; they might simply believe that it is good in itself in some perverse way. Such people may very well be acting from a good motive, in the sense that their motive is to do what is morally right, though they happen to have

<sup>25</sup> Mill, *Utilitarianism*, p. 47.

<sup>26</sup> Ibid.



a false view of what morality demands.<sup>27</sup> It would seem that not punishing such people, and thereby deterring such fanatics could have deeply dangerous consequences. Indeed, in *An Examination of Sir William Hamilton's Philosophy*, Mill is keen to emphasize how it 'is impossible to assert the justice of punishment for crimes of fanaticism, on any other ground than its necessity for the attainment of a just end. If that is not a justification, there is no justification'.<sup>28</sup> Mill takes it as given that such fanatics ought to be punished and argues that utilitarianism provides the only plausible explanation of this judgement. Indeed, Mill cites his father's views with approval here in a way that is very revealing of his hostility to making punishment sensitive to motives for just this sort of reason:

He [James Mill] regarded as an aberration of the moral standard of the modern times, compared with that of the ancients, the great stress laid upon feeling. Feelings, as such, he considered to be no proper objects of praise or blame. Right and wrong, good and bad, he regarded as qualities solely of conduct – of acts and omissions; there being no feeling which may not lead, and does not frequently lead, either to good or bad actions; conscience itself, the very desire to act right, often leading people to act wrong. Consistently carrying out this doctrine, that the object of praise and blame should be the discouragement of wrong conduct and the encouragement of right, he refused to let his praise or blame be influenced by the motive of the agent.<sup>29</sup>

Third, a scheme making punishment sensitive to motives in this way is likely to be unpopular in virtue of the widespread belief that these first two considerations are sound, undermining respect for the law and morality.

It seems, then, that Mill is on safe ground in assuming that our jurisprudential practices should individuate actions in terms of intentions rather than motives. Not only does common sense affirm this conclusion, it can be given a powerful utilitarian justification. My argument has focused on external (legal and informal), rather than internal sanctions (of conscience), but this is fine. For on Mill's definition it is sufficient for an act to be wrong that it is appropriately punished, and it is sufficient for an act to be right that it is appropriately compelled. Hence to determine whether an action is right or wrong it is always relevant to determine the agent's intentions; otherwise we could not be sure that it does not fall under some appropriate possible legal rule. This is perfectly compatible with allowing that being appropriately subjected to internal sanctions is also sufficient

<sup>27</sup> Though this is perhaps a strained sense of 'good motive', given the dangerousness of moral fanatics. Thanks to Jonathan Dancy for discussion here.

<sup>28</sup> *CW*, ix. 461 f.

<sup>29</sup> *Autobiography*, *CW*, i. 51. Thanks to an anonymous referee for bringing this very helpful passage to my attention.

for an action's being wrong. A problem would arise for my interpretation only if the appropriateness of internal sanctions was sensitive to motives. Fortunately, the many (if not all of the) arguments rehearsed above against making motives relevant in the case of external sanctions have relatively obvious analogues in the case of internal sanctions as well, and I have focused on the external case only for ease of exposition. For example, resource-diversion is still an issue, since a person's own motives are famously not transparent even to the agent herself. Also, the risk of rationalization would be increased if people could avoid the pangs of conscience by convincing themselves simply that their motives were pure in spite of their bad intentions. So if we read the chapter V definitions of 'right' and 'wrong' back into chapter II, then we have a plausible explanation of why Mill would say what he does say.

My suggestion, not surprisingly, is that we should read the chapter V definitions of 'right' and 'wrong' back into the discussion of intentions and motives in chapter II. This interpretation seems justified on at least three grounds. First, and most importantly for present purposes, it seems the only apparent way to make Mill's overall view on motives and intentions intelligible, so charity speaks in favour of this interpretation. Second, it is only in chapter V that we really get Mill's considered and more refined view – the characterization of rightness and wrongness at the beginning of chapter II is deliberately rough and meant only to give the reader the general idea of utilitarianism. There is no apparent reason to think that in the discussion of intentions and motives Mill would not have relied upon his more considered view.<sup>30</sup> Third, there are Mill's previously quoted remarks on his father's discussion of intentions. That is the only other context (of which I am aware) in which Mill makes exactly the same point about rightness and wrongness always depending only upon the agent's intention and not upon his motive, and right after he makes this point in that context he immediately moves (within the same paragraph and in the very next sentence) to a discussion of when an agent's intention makes him guilty or worthy of acquittal. In that context, Mill seems to see the question of rightness or wrongness and the question of guilt or worthiness of acquittal, as one and the same question, as he moves from the one to the other without any sign whatsoever that he thinks he is changing the subject. In the passage in chapter II of *Utilitarianism* and in the discussion of intentions and motives in his commentary

<sup>30</sup> Nor should one worry that he had not yet formulated that view, for chapter V was actually composed before chapter II. See Berger, *Happiness, Justice, and Freedom*.

on his father's work, Mill has in mind the more refined, chapter V accounts of moral rightness and wrongness.<sup>31</sup>

Mill makes a strong claim about the connection between intentions and the rightness or wrongness of action and denies that there is any such connection between motives and morality. Taken in isolation, such claims are puzzling, especially coming from a utilitarian. However, when we put Mill's striking claims in the context of his larger theory, and in particular his account in chapter V of *Utilitarianism* of rightness and wrongness, as well as the context of his connection to Bentham, Mill's remarks are intelligible and plausible. Admittedly, the specific details of my reconstruction of how Mill might have defended his claims are somewhat speculative, and perhaps Mill would not have endorsed them all. The crucial point is that once one sees just what Mill meant by 'right', 'wrong', 'intend' and their cognates, all he must do to defend his claims is show that it is sensible to adopt practices of blame and punishment that individuate actions in terms of intentions rather than motives. Even if Mill would not (or should not) adopt the specific arguments for such practices canvassed here, there may well be room for a range of other arguments for that conclusion. In any case, the main burden of my argument has been to show that Mill's claims can be made intelligible and plausible, though whether they are sound will depend upon empirical questions. We may, on empirical grounds, disagree with Mill about the relative importance of intentions and motives for a utilitarian, but we should not find his account of their relative importance philosophically puzzling.<sup>32</sup>

mr ridge@staffmail.ed.ac.uk

<sup>31</sup> It might be objected that my proposal blurs the distinction between act-evaluation and agent-evaluation, since calling an action right on my account amounts to claiming that agents ought to be compelled to perform such actions. If this is a problem, then I think it pretty clearly really is a problem for Mill, given his claims in Chapter V. In fact, though, there is still room for an important distinction between act- and agent-evaluation. Calling an action right is to evaluate the action, and to say that it the action is of a type which ought, in general, to be compelled. The point of telling someone an action is right is to get them to do it. Similarly, the point of telling someone what they did was wrong is to get them not to do it in the future, and to do so by making them feel guilty. By contrast, when we tell someone she is a bad person, as when we call someone a coward, we are telling her not that any particular action of hers is such that she should feel guilty about it, but that she should try to be a different kind of person. In this case, the emotion we aim to produce is shame, rather than guilt, and the main point is to criticize the motives which typically lead the agent to act, rather than to criticize some particular action of hers. For a classic discussion of the distinction between guilt and shame (and one which does not match exactly with my own), see John Rawls, *A Theory of Justice*, Cambridge, MA, 1971, pt. III.

<sup>32</sup> Many thanks to Geoffrey Sayre-McCord and Jonathan Dancy for helpful comments and encouragement.