

## SECRECY IN CONSEQUENTIALISM: A DEFENCE OF ESOTERIC MORALITY

*Katarzyna de Lazari-Radek and Peter Singer*

### *Abstract*

Sidgwick's defence of esoteric morality has been heavily criticized, for example in Bernard Williams's condemnation of it as 'Government House utilitarianism.' It is also at odds with the idea of morality defended by Kant, Rawls, Bernard Gert, Brad Hooker, and T.M. Scanlon. Yet it does seem to be an implication of consequentialism that it is sometimes right to do in secret what it would not be right to do openly, or to advocate publicly. We defend Sidgwick on this issue, and show that accepting the possibility of esoteric morality makes it possible to explain why we should accept consequentialism, even while we may feel disapproval towards some of its implications.

### **1. Introduction**

In *The Methods of Ethics*, in the course of discussing 'whether exceptions should be permitted from ordinary rules on Utilitarian principles' Sidgwick famously divided society into 'enlightened utilitarians' who may be able to live by 'refined and complicated' rules that admit exceptions, and the rest of the community to whom such sophisticated rules 'would be dangerous.' Therefore, he concluded:

... on Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice and example.'  
[ME 489]<sup>1</sup>

<sup>1</sup> References in the text are all to Henry Sidgwick, *The Methods of Ethics*, 7th edition, (London: Macmillan, 1907).

What Sidgwick is here endorsing has come to be known as esoteric morality. Some key tenets of esoteric morality are worth mentioning at the outset:

- There are acts which are right only if no one – or virtually no one – will get to know about them. The rightness of an act, in other words, may depend on its secrecy. This can have implications for how often, and in what circumstances, such an act may be done.
- Some people know better, or can learn better, than others what it is right to do in certain circumstances.
- There are at least two different sets of instruction, or moral codes, suitable for the different categories of people. This raises the question whether there are also different standards by which we should judge what people do.
- Though the consequentialist believes that acts are right only if they have consequences at least as good as anything else the agent could have done, the consequentialist may need to discourage others from embracing consequentialism.
- Paradoxically, it may be the case that philosophers who support esoteric morality should not do so openly, because as Sidgwick said: ‘it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric’ [ME, 490].

Some people think that the fact that something *would* be wrong if it were done openly shows that it *is* wrong, even if done in secret. John Rawls endorses something like this view, as does Bernard Gert, for they both make it a condition of something being a normative theory that it be public. T.M. Scanlon argues that ‘justifiability to others’ is basic to morality, a claim that can be understood as ruling out esoteric morality. Bernard Williams rejects esoteric morality on more substantive grounds, calling it ‘Government House utilitarianism’ because it might justify the kind of paternalistic morality agreeable to colonial administrators.<sup>2</sup> Brad Hooker in *Ideal Code, Real World* seems to agree with both Rawls and Williams, for he holds that the right thing to do is what would

<sup>2</sup> These comments come from: Bernard Williams, ‘The Point of View of the Universe: Sidgwick and the Ambitions of Ethics’, *Cambridge Review*, May 1982, p. 189; Bernard Williams, *Ethics and the Limits of Philosophy*, (London: Fontana, 1985), p. 109; B. Williams, *Utilitarianism for and against* (Cambridge: Cambridge University Press, 1973), p. 139.

be in accordance with a moral code that, if widely accepted, would have the best consequences, and at the same time he describes Sidgwick's proposal as 'paternalistic duplicity,' saying that it would be morally wrong even if it maximized the aggregate good.

The main aim of this paper is to show that these objections fail. Esoteric morality is a necessary part of a consequentialist theory, and all of the points above can be defended. Thus we agree with Adrian Piper's assertion that 'secrecy is a necessary ingredient in a viable Utilitarian doctrine'<sup>3</sup> but in contrast to Piper we argue that it is a defensible aspect of any coherent and consistent consequentialism. (The exception would be a form of consequentialism that ranks openness as an overriding intrinsic value, but to our knowledge no consequentialist has defended such a view and we shall not consider it further.) We shall show why we should not accept any specification of the concept of morality that rules out esoteric morality. As for the accusation of paternalism and duplicity, we shall draw its sting by showing how difficult it is to maintain that paternalism or duplicity are always wrong. In response to Hooker's critique, we shall show that he himself cannot reject esoteric morality as categorically as he sometimes appears to think he can. Insofar as he can reject it at all, we argue, it is only because his rule-consequentialism is not really a form of consequentialism at all.

The idea that it is better if some moral views are not widely known was not invented by Sidgwick. In Plato's *Republic*, Socrates proposes that ordinary people be brought up to believe that everyone is born 'from the earth' into one of three classes, gold, silver or bronze, and living justly consists in doing what is in their nature. Only the philosopher-rulers will know that this is really a myth, a 'noble lie'.<sup>4</sup>

More surprisingly, perhaps, Catholic moral theology has also found that it cannot avoid the need for a doctrine that is plainly not intended to be widely known. This applies, for example, to the doctrine of 'mental reservation,' which holds that it is permissible to say something that misleads, and yet avoid the sin of lying by mentally adding information that would, if spoken, make the response truthful. For example, in Charles McFadden's *Medical Ethics*, a text written from a Roman Catholic perspective, doctors

<sup>3</sup> A. Piper, 'Utility, Publicity, and Manipulation', *Ethics*, 88, p. 201.

<sup>4</sup> Plato, *Republic* 414b-415d.

and nurses are advised that if a feverish patient asks what his temperature is, and the truth would alarm him and make his condition worse, it is justifiable to reply 'Your temperature is normal today' while making the mental reservation that it is normal for someone in the patient's precise physical condition.<sup>5</sup>

## 2. Sidgwick's Defence of Esoteric Morality

It is, however, in the context of utilitarianism that most recent discussion of esoteric morality has taken place. In order to better understand why we are right to accept it, let us return to the passage from Sidgwick we have quoted. We will break it up into its distinct elements, and supply examples that help us to see what he means and why he is right. Sidgwick begins by saying:

*1) 'it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly'.*

On consequentialist grounds, it is plausible to believe that people who have more than they require to meet their basic needs ought to give everything they can spare to organizations like Oxfam that work to improve the lives of people living in extreme poverty in developing nations.<sup>6</sup> For present purposes, we can leave the notion of 'everything we can spare' undefined; what is important is that it is a highly demanding level of giving. But it is also plausible to believe that, given the way human beings are, very few of them will respond to an appeal to give away everything they can spare to help the poor. In that case, such an appeal will do little to help the poor. Perhaps advocating so demanding a standard will just make people cynical about morality as a whole: 'If *that* is what it takes to live ethically,' they may say, 'let's forget about ethics, and just have fun.' If, however, we were to promote the idea that living ethically involves donating, say, 10% of your income to the poor, we may get better results: many would give, and none would become moral cynics. Let's assume that this is the case, and the

<sup>5</sup> We owe this reference to Sissela Bok, *Lying*, (New York: Pantheon, 1978), p. 31. For more on the doctrine of mental reservation, see also *Catholic Encyclopedia*, <http://www.catholic.org/encyclopedia/view.php?id=7895>.

<sup>6</sup> Peter Singer, 'Famine, Affluence and Morality,' *Philosophy and Public Affairs*, I (1972), pp. 229–243; Peter Singer, *The Life You Can Save*, (New York: Random House, 2009).

total amount raised to help the poor by advocating that morality requires every comfortably off person to give 10% of their income will be much greater than would be raised by advocating that we should give everything we can spare. After we have done what we can to spread the message that every ethical person should give a tenth of their income to the poor, however, there remain many people in great need. Although I have myself given a tenth of my income to Oxfam, I still have enough left to dine out, go to movies, and take vacations abroad. I know that the consequences would be better if I were to stop doing this and give everything I can spare to Oxfam. Since I am a consequentialist, I know that that is what I ought to do.

Now Sidgwick's first element of esoteric morality applies. It would not be right to advocate openly that people ought to give everything they can spare to the poor, because that would be counterproductive. The right level of giving to advocate openly is 10%. On the other hand, assuming that I have no other relevant obligations that constrain me from giving more than 10%, I ought to give more, and in not giving more, I am doing something wrong. What if someone asks me, privately, how much she should give? If I know her well, and know that she is one of those rare people who would respond positively to a highly demanding standard, I should privately tell her what, as a consequentialist, I believe: that she ought to give substantially more than 10%.

2) *'it may be right to teach openly to one set of persons what it would be wrong to teach to others' . . .*

An antiterrorist officer teaches a specially selected elite how to break down terrorists psychologically, so as to get significant information from them that will make it possible to prevent further terrorist attacks. At the same time he teaches others, who cannot be trusted to use such methods only when necessary, that they must always treat prisoners humanely and that humane treatment rules out the very methods that he is instructing the elite to use.

Can such a distinction be defended? It is difficult for a consequentialist to defend the view that torture is always wrong – consider the famous 'ticking bomb' scenario in which torturing a terrorist, or even the terrorist's child, is the only way to prevent a nuclear bomb exploding in a city that will kill millions and maim or sicken millions more. But it is also plausible that without a general prohibition on torture, military and police personnel will

frequently torture people in a far wider range of situations, and that all of these instance of torture will be unjustifiable. Whether it is possible to find a few people who will, after the appropriate training, be able to restrict the use of torture or other inhumane techniques to the very rare occasions when it is clearly justified is an empirical question on which we do not need to offer an opinion. It is enough if we can show that, on the basis of facts that are not obviously false, it could be right to teach one group something that it would be wrong to teach another group.

- 3) *'it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world;'*

A soldier is brought before a general, charged with desertion in battle, for which the mandatory penalty is a long term of imprisonment. The soldier admits that he did desert, but begs for pardon, saying that he does not want his two small children to grow up without a father. Only one or two other people, who the general feels he can trust, know that the soldier deserted. The general assigns the soldier to duties behind the front line, telling him he must never say a word to anyone about deserting, or being charged with desertion. He tells his administrative officers to destroy all records of the charge and forget all about it.

Assume that if it were to become generally known that the fathers of small children may desert with impunity, this would materially hinder the ability of the army to fight, and that this would have bad consequences. Therefore it would have been wrong for the general to allow all the world to know of his decision regarding the deserter. It may still have been right for him to make that decision. Two children will have their father at home, the soldier and his wife will have a much better life, and the army's fighting ability will suffer no material harm.

- 4) *'and even, if perfect secrecy can be reasonably expected, [it may conceivably be right to do] what it would be wrong to recommend by private advice and example.'*

This is the most puzzling of Sidgwick's categories of esoteric morality. How is it possible that an act might be right to do, but not right even to recommend, in private, by advice or example? We offer the following case.

One of the most common objections to consequentialism is based on a hypothetical situation in which a surgeon has to do a delicate brain operation on a patient who happens to be the ideal organ donor for four other patients in the hospital, each of whom will die shortly unless they receive, respectively, a heart, a liver, and – for two of them – a kidney. The doctor is highly skilled, and is confident of her ability to carry out the brain surgery successfully. If she does, her patient will lead a more or less normal life. But because the operation is a delicate one, no one could blame her, or have any reason to suspect anything, if the patient were to die on the operating table. Moreover, the hospital is experienced in organ transplantation, and the surgeon knows that if the patient were to die, the recipients of the patient's organs would soon be able to go home and lead a more or less normal life. The surgeon knows no other details about her patient or the other patients, such as whether they are married, have children, or are about to discover a cure for cancer. In these circumstances, critics of consequentialism say, the consequentialist must think that the doctor ought to kill her patient, since in that way four lives will be saved, and only one lost, and this must be better than four dying and only one being saved. But, so the objection runs, it is obviously morally wrong for the surgeon to kill her patient, and any moral theory that says the contrary must be rejected.

We agree that the consequentialist must accept that, in these circumstances, the right thing for the surgeon to do would be to kill the one to save the four, but we do not agree that this means that consequentialism should be rejected. We think, on the contrary, that the appearance of unacceptability here comes from the fact that this is one of those rare cases in which the action is right only if perfect secrecy can be expected. Moreover, it is not an action that should be recommended to others. We realize that there is a paradox in saying, both that the action is right, and that it should not be recommended, since to say that it is right *is*, in a sense, to recommend it. We shall come back to this point shortly.

It is easy to see why in the doctor case, like the deserter case, it would be disastrous for it to become generally known that doing what the doctor did was regarded as right. The doctor's act violates some core principles of medical ethics. It is contrary to the ancient rule *primum, non nocere* ('first, do not harm') which tells doctors to give priority to not harming patients. That rule is said to encourage humility in doctors, reminding them of their fallibility and the power they have to cause harm. In contrast, leading

doctors to believe that they are justified in killing one patient to save others will only reinforce their occupational tendency to arrogance and may lead them to do what is wrong more often than it leads them to do what is right.

Killing a patient is the most flagrant breach imaginable of the ethics of the doctor-patient relationship, which is founded on the patient's trust in the doctor to put the interests of the patient first, and the necessity for the doctor to live up to that trust. If patients believed that doctors might kill them in order to help other patients, they would be unable to have confidence in their doctors, they would become fearful about going to hospital, and their health would suffer.

Given the widespread damage that would flow from general knowledge of a doctor killing a patient to benefit others, and the extreme rarity of the circumstances in which killing one patient could benefit several others, we think it obvious that a rule absolutely prohibiting such acts by doctors is justified. In this respect this case is like the ticking bomb justification for torture. Both are extremely hypothetical cases that are unlikely to arise. But would the doctor case also be one in which it is wrong to recommend, even privately, that someone should act in that way? One factor here is the damage that would be done if the act were to become public. Of our four examples, it seems likely that this would be greatest in the doctor case, whereas the benefit in that case would be much less than in the ticking bomb torture. And of course, whatever one may say about hypothetical case, in reality the more often a doctor kills one patient to save several others, the more likely it is that eventually this will become known. When considering the doctor case as a hypothetical case in which we can know that perfect secrecy will be preserved and the doctor is so self-disciplined that she will not be led by her success in this case to continue the practice of killing patients in other, more risky, situations, the consequentialist can hardly deny that what the doctor did was right. In the real world, however, we do not have such perfect foreknowledge, doctors may have flaws in their character, and it is likely to be better to say, not only in public, but also as a matter of private advice, that no doctor should ever contemplate killing a patient who wants to live.

As Sidgwick's examples show, esoteric morality takes distinct forms, requiring varying degrees of secrecy in different circumstances. We regard Sidgwick as correct in maintaining that consequentialists must accept some examples of esoteric morality as

justified in each of the categories that he describes, and we have sought to provide suitable examples. It remains to be considered whether the fact that consequentialism implies esoteric morality is a damaging objection to consequentialism, or simply an implication that consequentialists can and should embrace.

### 3. Objections

#### (i) *Rawls and Gert*

The ‘publicity condition’ is a well-known aspect of John Rawls’ theory of justice. He holds that the condition applies ‘for the choice of all ethical principles and not only for those of justice.’<sup>7</sup> He asserts that the publicity condition ‘arises naturally from a contractarian standpoint.’ Bernard Gert is another advocate of a publicity condition. Morality is, for him, ‘an informal public system that applies to all rational persons.’ By ‘public system’ he means that ‘all those whose behaviour is to be judged by the system, understand it, and know what kind of behaviour the system prohibits, requires, discourages, encourages and allows’ Gert seeks to justify ‘common morality’ – that is, the moral system that most thoughtful people use when making everyday, common-sense moral decisions – and believes that his position on publicity is shared by a wide consensus. He writes: ‘Hardly anyone denies that morality must be such that a person who adopts it must also propose its adoption by everyone’.<sup>8</sup>

We deny it, for the simple reason that such a definition of ethics, or morality, unduly narrows the most important practical question it is possible to ask. In the first sentence of *The Methods of Ethics* Sidgwick defines a ‘method of ethics’ as ‘any rational procedure by which we determine what individual human beings “ought” – or what it is “right” for them – to do, or to seek to realise by voluntary action.’ [ME 1] Note here the reference to ‘individual human beings’. It seems that Rawls and Gert want to define ethics as ‘any rational procedure by which we determine what a community of human beings ought to take as a standard of right or wrong

<sup>7</sup> John Rawls, *A Theory of Justice*, Revised edition, (Cambridge, Mass.: Harvard University Press, 1999), p. 112.

<sup>8</sup> Bernard Gert, *Morality: Its Nature and Justification*, New York: Oxford University Press, 1998, pp. 8–11.

for the voluntary actions of its members.’ That simply rules out of court what is surely a proper normative question: whether it is sometimes ethically justifiable to do what will, if and only if it remains secret, have best consequences. To answer that question we need substantive moral argument, not a definitional *fiat*.

Still, a defender of the Rawls-Gert definition might object, we can’t define concepts in whatever manner we like. Isn’t it just *true* that every society has its own morality, and as any anthropologist will tell us, morality is a social institution, not a matter of individual judgment? Here we have to be careful to distinguish descriptive claims from normative ones. Descriptively, it may be true that the morality of any society is its collective, shared code of behaviour. That’s the kind of thing an anthropologist may say and in that sense we have no objection to it. But if we are asking what we ought to do, we are asking a normative question that no anthropologist’s account of the nature of morality can answer. At most it can give us what R. M. Hare called an ‘inverted commas’ sense of what we ‘ought’ to do.<sup>9</sup> This no more determines what we ought to do – in the full-blooded, normative sense of the term – than rules of etiquette determine to whom we ought to speak without an introduction. We still need to decide whether we should act in accordance with that shared code.

To avoid misunderstanding, we will add that to reject the view that we ought always to act on a rule that can be part of a public system of morality is not to reject what Sidgwick calls the ‘principle of justice, that ‘it cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment.’ [ME 380] (This requirement is now more commonly known, following R. M. Hare, as the principle of universalizability.<sup>10</sup>) On the contrary, we regard the universalizability requirement as a way of distinguishing moral judgments from those that are merely self-interested. A person who wants to practice esoteric morality should ask himself if he is willing to agree that anyone else should do the same when in relevantly similar circumstances.

<sup>9</sup> R.M. Hare, *The Language of Morals*, (Oxford: Clarendon Press, 1952), pp. 124–126, 163–165.

<sup>10</sup> R.M. Hare, *Freedom and Reason*, (Oxford: Clarendon Press, 1963), pp. 10–13.

But one relevantly similar circumstance is the fact that the act will remain secret.

(ii) *Scanlon*

In *What We Owe to Each Other*, Scanlon does not endorse a publicity condition, as such, but rather the idea that ‘thinking about right and wrong is, at the most basic level, thinking about what could be justified to others on grounds that they, if appropriately motivated, could not reasonably reject.’<sup>11</sup> Does this conception of morality rule out esoteric morality? On one interpretation, it does not. After all, this article seeks to justify esoteric morality to others – namely you, the readers – and if our argument is as rigorous as we hope, you will not reasonably be able to reject this justification. It may well be, however, that when Scanlon writes about ‘justification to others’ – his emphasis – and distinguishes this from the idea that ‘a concern that one’s actions should be morally justifiable’ he is suggesting something more, and something that is incompatible with esoteric morality.<sup>12</sup> The general who allows the deserter to go free cannot justify his conduct to others, for to do so will lift the veil of secrecy that is essential to his action having the best consequences, all things considered.

Our response to this suggestion is along the same lines as that which we made to Rawls and Gert. To define ‘thinking about right and wrong’ as requiring justification to others in the sense specified is to exclude from the sphere of morality actions that might be morally justifiable, as long as the justification is kept to oneself, or to a very limited group. Whether or not it is wrong to do something that can only be justified in this silent or near-silent manner, is itself a substantive moral question. That question should not be ruled out of play by a definition that makes morality too narrow to include it.

(iii) *Hooker*

The opening sentence of Brad Hooker’s *Ideal Code, Real World* sets up a challenge to esoteric morality: ‘Shouldn’t we try to live by the

<sup>11</sup> T.M. Scanlon, *What We Owe to Each Other*, (Harvard University Press, Cambridge, MA, 1998), p. 5.

<sup>12</sup> *Ibid.*, p. 169.

moral code whose communal acceptance would, as far as we can tell, have the best consequences?’ Hooker apparently sees this as an open question, for here at least he eschews the easy answer ‘Yes, because that is what it means to act morally.’ Later in the book, when explicitly considering the challenge that Sidgwick’s account of esoteric morality presents to his moral theory, Hooker gives a moral, not a conceptual, response, saying ‘Such paternalistic duplicity would be morally wrong, even if it would maximize the aggregate good.’<sup>13</sup> Admittedly, in other places Hooker swings towards a conceptual answer to the question, referring to ‘the idea that morality should be thought of as a *collective, shared* code’ and citing Gert in support.<sup>14</sup> Since we have dealt with that claim, we will here focus on Hooker’s other arguments against esoteric morality.

In maintaining that elitism of the kind endorsed by Sidgwick is objectionable, Hooker goes so far as to say that he would stand by this view, even if the world consisted of a million imbeciles who could understand only ultra-simple rules, and one genius who could grasp much more complex rules, and better consequences would result from the genius following the more complex rules. Nevertheless, says Hooker, the genius should follow the same rules as the rest.<sup>15</sup> Exemption is only granted to the genius if following the simple rules would result in disaster. We’ll return to this exemption shortly. Meanwhile, why does Hooker think that the genius should follow the simple rules, although she could do more good by following the complex ones? He mentions the convenience of having just one code for internalization by everyone, but that can’t be the major reason, for we can suppose that the good the genius could do by acting on the more complex reasons would outweigh the inconvenience she would suffer by having her own set of rules. At this point Hooker swings towards a conceptual answer to the question, referring to ‘the idea that morality should be thought of as a *collective, shared* code.’ He acknowledges that this thought is not a ‘peculiarly consequentialist one’ – indeed, we would say that it is not a consequentialist one at all – but he adds that ‘we may favour one version of rule consequentialism over any other because this version coheres best

<sup>13</sup> B. Hooker, *Ideal Code, Real World*, (Oxford: Clarendon Press, 2002), p. 85.

<sup>14</sup> *Ibid.*, p. 87.

<sup>15</sup> *Ibid.* p. 86. Hooker attributes the objection to Margaret Little.

both with our *general* beliefs about morality and with our beliefs about what morality requires *in particular cases*.<sup>16</sup>

Hooker is here referring to his endorsement of reflective equilibrium. In the introduction to *Ideal Code, Real World* he lists it among the methodological criteria on which he will base his argument, and states it as follows: 'Moral theories must cohere with (i.e. economically systematize, or, if no system is available, at least endorse) the moral convictions we have after careful reflection.'<sup>17</sup> A few pages later he restates this as 'We should evaluate rival moral theories in terms of their ability to cohere with the convictions in which we have the most confidence after due reflection.'<sup>18</sup>

Who are the 'we' to whom Hooker is referring? There are two possibilities. One is that it refers to the general public – in other words, that we should evaluate rival moral theories in terms of their abilities to match what Sidgwick called 'commonsense morality' or what philosophers nowadays sometimes call 'folk morality.' The other possibility is that 'we' refers to those of us who are evaluating various moral theories in order to decide which one to accept.

If Hooker is referring to the general public, we reject the idea that the test of a sound moral theory should be how well it coheres with the moral convictions of the general public, even after 'careful reflection.' Encouraging members of the general public to reflect carefully on their moral convictions is all very well, but it is unlikely to diminish to any significant extent the influence of a variety of factors that are irrelevant to the soundness of a moral theory, including false religious beliefs, cultural and ethnic prejudices, and the innate predispositions that are a legacy of the process of millions of years of evolutionary selection.<sup>19</sup>

If, on the other hand, Hooker is referring to those of us who are evaluating the various moral theories – in other words, people who read and discuss *Ideal Code, Real World* – then the appeal to reflective equilibrium is otiose. *Of course* we should evaluate rival moral theories in terms of their ability to cohere with the convictions in which we have the most confidence after due reflection. The due

<sup>16</sup> Ibid p. 88.

<sup>17</sup> Ibid p. 4.

<sup>18</sup> Ibid p. 12.

<sup>19</sup> For further discussion see Peter Singer, *The Expanding Circle: Ethics and Sociobiology*, (New York: Farrar, Straus and Giroux, 1981) and Peter Singer, 'Ethics and Intuitions', *The Journal of Ethics*, 9, (2005), pp. 331–352.

reflection is what we are doing right now, and for philosophers and anyone else capable of considering the arguments we are presenting, to say that we accept a moral theory just *is* to say that it is the one that, after going through this process, coheres with the convictions we have most confidence in. Because we are philosophers discussing what fundamental normative theory we should accept, everything is at least open to discussion. If our firmest conviction is that we should do whatever will have the best consequences, and after due reflection this conviction remains firmer than the idea that we should always avoid paternalism and duplicity; and if in particular cases where duplicity has better consequences than honesty, we still think we should do what has the best consequences, then Hooker's appeal to reflective equilibrium gives us no reason to accept his version of rule-consequentialism nor to reject esoteric morality. We are not trying to match our theory with common moral judgments: it's only our own convictions that have to be coherent. Hooker may, of course, have different convictions. To the extent that he does, he is not a consequentialist and that is a fundamental point of difference between us.

We want to forestall one possible response. In saying that the best normative theory does not have to match the moral convictions of the general public, we are not saying that these convictions are irrelevant to moral decisions, or can generally be ignored. Consequentialists cannot ignore anything that may have an impact on the consequences of their actions, and the moral convictions of ordinary people will have an impact on the consequences of our actions in a wide variety of fields. We entirely agree with Hooker that in proposing or promoting a moral rule for general acceptance in a society, it is vital to know whether it coheres with the prior moral convictions of most people, and hence has good prospects of easy acceptance, or clashes with these prior moral convictions, and so runs a high risk of rejection. In that sense, these moral convictions are data that we must take into account in deciding what we ought to do. But that is a very different thing from giving them probative force in deciding which normative theory we should accept.

As we already noted, Hooker's version of rule-consequentialism allows us to breach rules when doing so is the only way to avoid disaster.<sup>20</sup> Indeed, Hooker believes that one of the most important

<sup>20</sup> See Brad Hooker, *Ideal Code, Real World*, pp. 86, 98–9.

rules that should be in a moral code is a rule telling people to prevent disasters. We need to know, of course, what is to count as a 'disaster.' In order to keep rule-consequentialism distinct from act-consequentialism, Hooker needs to maintain some conceptual space between his 'avoid disaster' rule, and the act-consequentialist rule that we should break all other rules any time doing so would maximize the good. In responding to an objection from Richard Arneson, Hooker shows that he is prepared to shrink this space significantly. Here is Arneson's objection:

In war, soldiers fighting for a just cause ought to stand by their post when attacked, unless outnumbered by attacking enemy so that even stout defence would be futile. Suppose this rule, followed by nearly everybody, would produce ideal results. But the rule in fact is not internalized by the military forces fighting for a just cause in a particular war. The enemy have attacked and most of your fellow troops have run away. You can stand and fight, in conformity with the ideal rule, or you can run and live to fight another day. The consequences of conformity to the rule would not be disastrous, but would be decidedly negative. You will die and gain very little if anything for your side . . . Common sense morality, which holds that the obligation to obey hypothetically useful rules is sensitive to the actual degree to which others are complying here and now, surely says one should run and live to fight another day. Act consequentialism to its credit says the same. Rule consequentialism, even sophisticated rule consequentialism with the disaster avoidance proviso added, would have to hold that one ought to stand and fight and die. So much the worse for sophisticated rule consequentialism.<sup>21</sup>

In his reply, Hooker admits that his account of what is a disaster is vague, but suggests that 'Perhaps . . . you are not required to stand by your post when this would do your side no good but would be disastrous for you.'<sup>22</sup> That seems reasonable, in itself, but what does it do to Hooker's objection to esoteric morality? Let's

<sup>21</sup> Cited by Brad Hooker, 'Reply to Arneson and McIntyre,' *Philosophical Issues*, 15 (2005), p. 270–271. Richard Arneson, 'Sophisticated Rule Consequentialism: some simple objections', *Philosophical Issues*, 15 (2005), p. 239.

<sup>22</sup> See Brad Hooker, *Ideal Code, Real World*, p. 133–4; Brad Hooker, 'Reply to Arneson and McIntyre,' *Philosophical Issues*, 15 (2005), p. 273.

combine Arneson's example with the case of the million imbeciles and the one genius. Regrettably the imbeciles are too stupid to tell when stoutly defending their posts has some chance of producing good results, and when it is futile. Attempts to get them to internalize a 'prevent disaster' rule have proven unsuccessful, because if they are given permission to use their own judgment about when a rule needs to be broken in order to prevent a disaster, the slightest danger leads them to sense that disaster is imminent, and they abandon their posts. The only way in which the nation can defend itself against unjust aggression from a tyrannical neighbour is to get its overwhelmingly imbecilic army to internalize the rule: 'Stand by your post at all times, unless you are ordered to retreat.'

Serving among these imbeciles, however, is the genius, who can be highly confident of her judgment about the utility, or futility, of standing by one's post when under attack. One day the platoon of imbeciles with which she serves comes under attack. Correctly judging resistance futile, she runs away, thus saving her life. When she reaches safety, she happens to encounter a television crew who ask her what happened to her comrades. She says that they bravely followed the rule 'Stand by your post at all times, unless you are ordered to retreat' and since no retreat was ordered, presumably have all been killed. The journalist then asks her why she did not follow the same rule. What would Hooker have her say? That she followed the disaster avoidance rule? But for her to say this on television would fatally weaken the internalization by virtually all of the population of the 'Stand by your post' rule and this would itself have disastrous consequences. So perhaps, in order to prevent disaster, the genius should lie, and invent some plausible story about her escape? While doing so she can tell herself that she is only doing this because the audience will not be able to understand the more complex moral rule on which she rightly acted. But now she is practicing esoteric morality.

Although the example of the single genius and the million imbeciles is far-fetched, the point is a general one: in order to prevent a disaster, we may have to do something that, if it were to become a disaster, would itself bring about a disaster of a different kind. So the 'avoid disaster' rule may *require* us to practice 'paternalistic duplicity.' Hooker could stand by his original claim that the esoteric morality Sidgwick supports 'would be morally wrong, even if it would maximize the aggregate good' but he needs to add 'unless it is necessary to prevent disaster.' He will, however, have

to withdraw his claim that his form of rule-consequentialism ‘rules out the objectionable elitism and duplicity from the start’ because that implies that elitism and duplicity can never be right.

So Hooker’s approach can justify esoteric morality, but only to prevent disaster, not simply to promote the good. How big a difference does this leave between his rule-consequentialism and act-consequentialism? Remember that on Hooker’s view the death of a single innocent person may be a disaster, so there are many acts, especially on the level of government actions – which is the level at which those despised paternalistic and duplicitous colonial administrators act – that can be construed as acts that are necessary to avoid disaster. And on the other side, as we shall see in more detail shortly, act-consequentialists have many good reasons for favouring transparency in most situations, and hence for not accepting esoteric morality when – taken apart from the risks of not acting and living in a transparent manner – it produces a modest net benefit. We think that in practice the difference between the circumstances in which the two theories permit esoteric morality will be quite narrow.

When Hooker offers examples of how his rule-consequentialism applies to problems in the real world, he focuses first on the issue that we used in our first example: our obligation to give aid to those living in poverty. Before concluding our discussion of Hooker, we want to comment on the differences between our views and his on this point.

Hooker acknowledges that we have obligations to help those in great need. This obligation applies, in his view, ‘even if the personal sacrifices involved in helping them add up to a significant cost’ but they do not require us to go beyond this threshold. Importantly, the ‘significant personal cost’ threshold is aggregative – even if you could help someone in great need at a very small personal cost, you are not obliged to do so if that small cost, added on to other costs you have already incurred to help others, takes you over the threshold of ‘significant personal cost.’<sup>23</sup> Consistently with Hooker’s general approach, the grounds on which he defends this moderately demanding principle are that, given human nature as it is and is likely to remain, of all the rules a society might adopt to prescribe how much people ought to give to those in need, this rule, if adopted by the vast majority of the

<sup>23</sup> Brad Hooker, *Ideal Code, Real World*, p. 166.

community, can be expected to have consequences at least as good as those of any other rule. The costs of trying to get a more demanding rule accepted and acted upon are just too high. About that, Hooker might be right.

It is worth noting that Hooker explicitly rejects one possible way for a rule-consequentialist to avoid highly demanding obligations in this situation. He does *not* say that our obligations are only to act in accordance with the rule that would have best consequences, if the vast majority of the society were to act upon it. That might result in an undemanding rule, because if *everyone* were giving aid, a relatively modest donation from each person could provide enough aid to relieve the needs of those living in extreme poverty. Hooker observes, rightly in our view, that if very few people are acting in accordance with the rule, we cannot ignore this fact and meet our obligations by giving only the relatively modest amount that would be sufficient if everyone were giving.

Nevertheless, Hooker thinks that the question ‘What moral rule would have the best consequences?’ can give us the answer to the question ‘What ought I to do?’ We, on the other hand, consider the questions distinct, and likely to have different answers. Hence the possibility that, as we have already mentioned, the right thing for a consequentialist to do is promote the rule that would have best consequences – very plausibly, as Hooker suggests, a moderately demanding one – while holding herself obliged to do much more than the rule requires.

#### 4. A Consequentialist Approach to Publicity

We have rejected the idea that ethics is necessarily public, and we have shown that Hooker’s defence of rule-consequentialism does not provide a sound basis for rejecting esoteric morality. We have offered examples of where we believe esoteric morality is justifiable. Nevertheless there are good reasons why consequentialists should share in the broad support for transparency in ethics, and hence should avoid esoteric morality in most circumstances.

##### *a. The benefits of a shared code*

It is plausible to hold that if a society is to work well, it needs to have a shared moral code that its members can internalize, follow

and expect others to follow. Many studies indicate that trust is an important factor in ensuring social welfare.<sup>24</sup>

*b. The benefits of open discussion*

Transparency permits open discussion and criticism of rules and policies that are being considered for implementation. To accept a morality that is only for the elite implies that we are permitted to manipulate those who are not part of the elite, in order to produce the best consequences. When we do so, we are unable to seek the opinions of those who we are manipulating on the policies we are actually implementing. This is the essence of Williams' objection to 'Government House utilitarianism'. We imagine the white colonial administrators sitting around in their cane arm-chairs under the ceiling fans, discussing how best to rule the natives. They may discuss their policies among themselves, and with the imperial government back home, but not with those who are most directly affected by them, the natives themselves. Under these circumstances they will have a tendency to convince themselves that what is in the best interests of the imperial power is the right thing to do. The danger is great that it will all go wrong because of the absence of exchange of ideas that could have happened if the policies had been transparent.

*c. The dangers of elitism*

Even if the lack of transparency does not lead to evils in any way comparable to those of oppressive colonial regimes, there are good grounds for objecting to dividing society into an elite and the masses. Whether it is nobles over peasants, whites over blacks, capitalists over workers, bolsheviks over the masses, or men over women, we know that those who are part of the elite will feel superior and have no difficulty in justifying, in their own terms, giving themselves privileges that in no way benefit – and often grievously harm – those they consider beneath them.

*d. The public nature of education*

We must also remember that morality is, at least in part, a social institution that exists only because each generation of children is

<sup>24</sup> See, for example, Francis Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity*, (New York: Free Press, 1995).

educated to accept it. Since education is a public process, this cannot be education in an esoteric morality, at least not unless the children of a special elite were to be educated in secret, which would have the undesirable implications just mentioned. So a large part of morality must consist of rules or principles that are known by everyone, including teachers and children.

Yet at the same time the fact that it is children who we must educate in the moral principles society accepts imposes some limits on transparency. We need to give children reasons for acting in accordance with moral rules and principles that they can easily understand and reliably apply. This will limit the kinds of reasons we can give them. At an early age, we may offer them self-interested reasons, or we may simply tell them that 'rules are rules' and must be obeyed. If we are not deontologists of this simple sort, when we say this to children we are failing to be transparent with them. And it is certainly not true of every society that children are encouraged, as they mature, to reflect on and discuss the basis for these rules. We may be influenced by the consequences of inviting children to question the justification for obeying moral rules, and if the consequences are going to be sufficiently bad, we may condemn those who, like Socrates, encourage the young to question the basis of morality.

When we deceive people about the reasons why they should act in a certain way, we make it impossible for them to develop their critical capacities, at least in respect of those reasons for action about which they are being deceived. The ideal kind of political entity, we may well think, is one in which all citizens are capable of deliberating on the reasons for acting and for adopting particular policies. If they are unaware of the true reasons for the principles and policies they are following, they will not learn these habits of deliberating, or will not learn them well.

*e. Respecting Preferences*

Our final reason is not one that Sidgwick would have supported. In his view, only states of consciousness have ultimate value. [ME 396]. Some contemporary utilitarians reject this view, seeing the ultimate good in terms of the satisfaction of preferences. 'Satisfaction' here need not refer to experienced satisfaction. Suppose that we prefer that our friends always give us their honest opinion about our work, never sparing our feelings, but they fear that if they tell the truth, we may become discouraged, so occasionally

they are dishonest. We never find out about this. Nevertheless, our preference that they be honest with us has not been satisfied. On a preference utilitarian view, the fact that this preference was not satisfied means that our lives went less well, other things being equal, than they would have if our friends had told the truth.

It is reasonable to assume that most people do not like the idea that others would tell them one thing while believing another – and that they especially would not want people in positions of authority and trust – like governments, or doctors – to deceive them about the principles they follow. If this is so, then for a preference utilitarian, it would count as another important reason against practising esoteric morality. Whether it is a decisive reason will depend on how most people would rank this preference against other preferences that it may only be possible to satisfy if some people do practise esoteric morality.<sup>25</sup>

The five reasons given above will count for both act-consequentialists and rule-consequentialists. Consequentialism generally accepts that it is desirable for a society that there is a publicly accepted set of rules or principles that people internalize and generally follow. Consequentialists can agree that it is important for people to be able to rely on the moral rules and to know that others will follow them – society will function better if there is a generally accepted set of rules than if there is not. The two forms of consequentialism differ on when exceptions from these rules can be allowed, and whether such exceptions should be made public or not.

## **5. Consequentialism as a Criterion of Right Action versus Consequentialism as a Guide to Action**

As we have acknowledged at the outset, consequentialists may, to bring about the best consequences, need to discourage others from embracing consequentialism. Though they are ready to justify their criterion of moral rightness – the best possible consequence – and the correctness of consequentialism as a criterion of right action, they may state that people should adopt other criteria (or *also* other criteria) as a guide to action as that will produce the best consequences.

<sup>25</sup> We thank Ronald Dworkin for alerting us to the significance of the distinction between hedonistic and preference utilitarianism on this issue.

Though we do believe that act consequentialism is the right moral theory, we tend to agree with Brad Hooker that ‘Maximize the good’ is not the best decision procedure, at least not always and not for every person, since it may bring about bad consequences such as a lack of trust, or something even worse.<sup>26</sup> This may be a problem of human nature and its tendency to think of oneself rather than of others – once people are allowed to break widely accepted moral rules in order to maximize the general good, it is entirely possible that they will also break them to achieve their own ends. There is also a problem with calculating what the result will be. Some people are able to calculate well, while some, whether because of a lack of intelligence, or of time, or for other reasons, are likely to miscalculate what would be best in a given situation. All this brings a consequentialist to the conclusion that we should encourage people to keep to a publicly known set of rules, to be truthful, to improve their character, and not to focus on maximizing the good all the time. But it may not be enough to say: ‘What you should aim at is to maximize the good, but in order to achieve it do not think about that all the time, try to follow the rules, work on your character, etc.’ It is possible that in order to achieve better results we have to keep the consequentialist aim itself secret.

To this line of thought, Bernard Williams objects: ‘If utilitarianism . . . determines nothing of how thought in the world is conducted, demanding merely that the way in which it is conducted must be for the best, then I hold that utilitarianism has disappeared, and that the residual position is not worth calling utilitarianism. If utility could be globally put together at all . . . then there might be maximal total utility from the transcendental standpoint, even though nobody in the world accepted utilitarianism at all.’<sup>27</sup>

Along with Derek Parfit we do not believe that Williams’ critique can do severe harm to consequentialism. There are two

<sup>26</sup> Brad Hooker writes: ‘if we had just the one rule “Maximize the good” sooner or later awareness of this would become widespread. And becoming aware of this would undermine people’s ability to rely confidently on others to behave in agreed-upon ways. Trust would break down. In short, terrible consequences would result from the public expectation that this rule would prescribe killing, stealing, and so on when such acts would maximize the good.’ *Ideal Code, Real World*, p. 94.

<sup>27</sup> Bernard Williams, in J. J. C. Smart and Bernard Williams, *Utilitarianism For and Against*, p. 135. See also Samuel Scheffler: ‘There is a persistent feeling of discomfort generated by the idea of a moral theory which is willing to require widespread ignorance of its own principles’ in *The Rejection of Consequentialism*, (Oxford: Clarendon Press, 1994), pp. 48–9.

main reasons for that. First, as Parfit observes even if a theory is self-effacing (that is, it tells us ‘that we should try to believe, not itself, but some other theory’<sup>28</sup>), this does not make it self-defeating, because

It is not the aim of a theory to be believed. If we personify theories, and pretend that they have aims, the aim of a theory is not to be believed, but to be true, or to be the best theory. That a theory is self-effacing does not show that it is not the best theory.<sup>29</sup>

A second reason for rejecting Williams’ critique is that it is unrealistic to think that it would follow from a consequentialist theory that no one should try to act in accordance with its criterion of rightness. Again along with Parfit, we believe that a consequentialist theory is likely to be only partially self-effacing (though it is also partly esoteric).<sup>30</sup> There are, and will be, people who believe in the principle of consequentialism and its rightness and who try to act to maximize general good. We doubt that the world would be a better place if no one was a consequentialist. Utilitarians have made the world a better place: Bentham, Mill, and Sidgwick are among these, along with many other utilitarian reformers.<sup>31</sup> In addition, as Peter Railton argues, consequentialism is more flexible than many deontological theories. It allows us to avoid any sort of ‘self-defeating decision procedure worship’ by taking into account the consequences of using particular decision procedures.<sup>32</sup>

## 6. Different Codes, Different Moral Judgments

Before we conclude, we should address the issue of how we should regard a person who does what is wrong because he follows

<sup>28</sup> Derek Parfit, *Reasons and Persons*, (Oxford, Oxford University Press, 1984), p. 40.

<sup>29</sup> *Ibid.*, p. 24.

<sup>30</sup> According to Parfit, consequentialism could be a partially self-effacing theory, if it prescribes that some people reject it, and it could also be esoteric, if it tells those who still believe in it that they should not ‘enlighten the ignorant majority’. See *Reasons and Persons*, pp. 41–42.

<sup>31</sup> For an explanation why it could be worse if we were all made to forget about consequentialism and made to believe in some kind of Common Sense theory see Parfit, *Reasons and Persons*, p. 41.

<sup>32</sup> Peter Railton, ‘Alienation, Consequentialism, and the Demands of Morality,’ *Philosophy and Public Affairs* 13 (1984), p. 156.

common morality when a sounder, consequentialist view of what to do in his circumstances has been kept secret. For example we may believe that generally it is a good thing to act according to the rule 'Do not lie' and so we encourage someone to follow this rule. Since we are afraid that if we tell him to break it sometimes in order to maximize the good, he will make too many mistakes in his judgments, we tell him to follow the rule under all circumstances. As a result he doesn't lie even in order to save an innocent person's life. Did he do wrong? It seems that a consequentialist has to judge him by whether he obeys the principles intended for him. But then we have to agree that we have different standards of judgment for those who can think critically about the nature of morality, and those who can't, or perhaps shouldn't. And that could even include all of us, in some circumstances. On the other hand we may say that there is only one standard – everyone does what is best in the circumstances in which he or she is. In any case, we should keep in mind that for a consequentialist, each instance of praising or blaming someone is itself an act that should be assessed in terms of its consequences.

## 7. Conclusion: The Paradoxical Nature of Esoteric Morality

Sidgwick says that his conclusions are of paradoxical character:

... the Utilitarian conclusion, carefully stated, would seem to be this; that the opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric. [p. 490]

In other words, in public, esoteric morality should be disavowed. Given that, it is not surprising that the idea of esoteric morality should meet with resistance, for to resist it *is* to publicly disavow it. To defend this position in public is therefore to be involved in a paradox. Most of you reading this may be philosophers, but you are also members of the public, and your resistance is therefore, the 'right' response, in the sense that it is good that you should have that response. You should be reluctant to embrace esoteric morality, and you should feel strongly that there is something

wrong with our conclusion. Nevertheless, sometimes we are right to do in secret what it would be wrong to do, or to advocate, in public.<sup>33</sup>

*Katarzyna de Lazari-Radek*  
*University of Lodz*  
*Kopcińskiego 16/18*  
*90-242 Lodz, Poland*  
*klazari@mm.com.pl*

*Peter Singer*  
*Princeton University,*  
*Princeton, NJ, USA 08544*  
*& University of Melbourne,*  
*Parkville, Vic. Australia 3010*  
*singerp@gmail.com*

<sup>33</sup> We thank Ronald Dworkin and Thomas Nagel for the opportunity to present an earlier version of this paper to their seminar at New York University. The comments we received on that gruelling but intellectually challenging occasion have helped us to strengthen the paper, although we doubt that this revised version will satisfy either our hosts or our other critics in the audience. We have also benefited from comments received when the paper was discussed at the University Center for Human Values, Princeton University, and at the Department of Ethics, University of Lodz.