CRITIQUE AND COMMENT

IN DEFENCE OF ONE TYPE OF RETRIBUTIVISM: A REPLY TO BAGARIC AND AMARASEKARA

R A DUFF*

[In 'The Errors of Retributivism' (2000) 24 Melbourne University Law Review *124, Bagaric and Amarasekara criticise four recent retributivist theories of punishment, and urge a return to a simple utilitarian theory. This paper responds to their criticism of one such retributivist theory, which justifies punishment as an exercise in moral communication aimed at inducing repentance, self-reform and reparation, and suggests that their defence of utilitarianism fails to confront the most serious objections to their utilitarian theory of punishment.]*

CONTENTS

I INTRODUCTION

A striking feature of penal philosophising during the last thirty years has been the revival of retributivism. What has been revived is not merely the kind of 'negative' retributivism which sets side-constraints on our pursuit of consequentialist aims that still provide the positive 'general justifying aim' of punishment,[1] but a 'positive' retributivism which finds at least a central part of the positive justification of criminal punishment in its character as a deserved response to past wrongdoing. In an article recently published in this journal,[2] Bagaric and Amarasekara identify four influential versions of positive retributivism, argue that none of them is tenable, and advocate a return to a simple utilitarian theory

* BA (Oxon); Professor, Department of Philosophy, University of Stirling.

1 On 'negative' as against 'positive' retributivism, see David Dolinko, 'Some Thoughts about Retributivism' (1991) 101 *Ethics* 537, 539–43; on the 'general justifying aim', see H L A Hart, *Punishment and Responsibility* (1968) ch 1 'Prolegomenon to the Principles of Punishment'.

2 Mirko Bagaric and Kumar Amarasekara, 'The Errors of Retributivism' (2000) 24 *Melbourne University Law Review* 124.

of punishment (simple because the theory is based on hedonistic act utilitarianism) — a theory which has been, they argue, unjustly maligned.[3]

I will not comment on Bagaric and Amarasekara's critiques of von Hirsch's retributivist theory (he can speak for himself), of the 'unfair advantage' theory (which has been comprehensively criticised by many theorists),[4] or of 'intrinsic' retributivism (which, despite Bagaric and Amarasekara's attempt to connect it to a non-cognitivist theory of moral judgement, hardly seems worth the attention they devote to it). However, I will respond to their criticisms of the communicative theory of punishment that I have defended, and I will comment briefly on their defence of utilitarianism.

II PUNISHMENT AS MORAL COMMUNICATION

I will not try to defend every aspect of the account that I offered in *Trials and Punishments*[5] against Bagaric and Amarasekara's criticisms — partly because that would be tedious, and partly because I no longer believe everything I argued then. However, I do still believe that criminal punishment should be justified as an exercise in moral communication which aims to bring offenders to face up to and repent their crimes, to reform themselves, and to make appropriate reparation to and seek reconciliation with those whom they wronged. I still believe, that is, that it should be justified as a species of secular penance, and I do not think that the objections offered by Bagaric and Amarasekara undermine this account.[6]

A *A Simple Example*

A defendant is convicted of a moderately serious crime — for instance of a series of burglaries of private homes. He is sentenced, let us imagine, under a 'combination order',[7] to a period of probation together with a requirement to undertake a specified number of hours of some court-approved community service. What should be the aims of this sentence? On my account, it should have several closely related aims.

First, it should aim to communicate to the offender, to bring home to him more forcefully than his conviction might have done, the censure that his crime deserves. He has committed a wrong — a wrong that properly falls under the

[3] Ibid.

[4] See, eg, R W Burgh, 'Do the Guilty Deserve Punishment?' (1982) 79 *Journal of Philosophy* 193; Jeffrie Murphy, 'Retributivism, Moral Education and the Liberal State' (1985) 4 *Criminal Justice Ethics* 3; R A Duff, *Trials and Punishments* (1986) ch 8; M Margaret Falls, 'Retribution, Reciprocity, and Respect for Persons' (1987) 6 *Law and Philosophy* 25; Barbara Hudson, *Justice through Punishment: A Critique of the 'Justice' Model of Corrections* (1987); Dolinko, above n 1; Jami Anderson, 'Reciprocity as a Justification for Retributivism' (1997) 16 *Criminal Justice Ethics* 13.

[5] Duff, *Trials and Punishments*, above n 4.

[6] For my more recent attempts to explain and defend (and revise) such an account, see R A Duff, 'Penal Communications: Recent Work in the Philosophy of Punishment' (1996) 20 *Crime and Justice: A Review of Research* 1; R A Duff, *Punishment, Communication and Community* (forthcoming, 2000); R A Duff, 'Penal Communities' (1999) 1 *Punishment and Society* 27.

[7] See *Criminal Justice Act 1991* (UK) s 11.

criminal law, as one that concerns not merely its direct victims, but the political community as a whole. His punishment should aim to bring him to hear and to recognise the community's condemnation of that wrong.

Second, it should aim to persuade him to *accept* that censure as justified — to face up to and to recognise what he did as wrong — which will also be to repent that wrong. This aim is internal to censure as a communicative enterprise: to criticise or censure another, to their face, for the wrong they have done is precisely to try (even if we have no real hope of succeeding) to bring them to recognise and repent that wrong.

Third, it should aim to persuade him to see the need to reform his future conduct so as to avoid such wrongdoing — and to help him begin to do so. Such a recognition of a need for self-reform is entailed by genuine repentance of past wrongdoing (to repent the wrong that I did is also to commit myself to trying to avoid repeating it). The help he might need could be provided, for instance, by discussions with the probation officer, by the experience of his community service, or by other more specialised programmes (for example, to help him to deal with addictions or to gain job-related skills) that the probation officer can offer.

Fourth, it should aim to provide some reparation to those whom he has wronged, and thus to reconcile him with them. What matters here is not *material* reparation for such material harm as might have been caused (punishable crimes do not always cause such harm; the harm they do cause is not always reparable; and the reparable harm they cause often cannot be repaired by the offender), but *moral* reparation for the moral wrong that was done. Central to such moral reparation is apology: the punishment the offender is required to undertake can be seen in part as a symbolic public apology that he is required to make. That apology is to be made to the direct victims of his crimes, and to the whole political community, whose public values he has flouted and who share in the wrong done to the direct victims.[8] Sometimes, as when an offender makes reparation to his victim through a victim–offender mediation programme, it is made immediately to the victim, and through the victim to the community; sometimes, as in the case of community service orders, it is made immediately to some part of the wider community, and thereby to the victim and to the community as a whole.[9]

Fifth, it should aim to reconcile the offender with those whom he has wronged, through this process of censure and symbolic apology. The offender, we hope, recognises those whom he has wronged as fellow citizens to whom he is bound; they recognise him as one who is still a fellow citizen despite his crime, and who has paid the punitive, apologetic debt that was due for his crime.

[8] See S E Marshall and R A Duff, 'Criminalization and Sharing Wrongs' (1998) 11 *Canadian Journal of Law and Jurisprudence* 7.

[9] On victim–offender mediation programmes, see Tony Marshall and Susan Merry, *Crime and Accountability: Victim/Offender Mediation in Practice* (1990); John Braithwaite, 'Restorative Justice: Assessing Optimistic and Pessimistic Accounts' (1999) 25 *Crime and Justice: A Review of Research* 1. On community service orders, see Ken Pease, 'Community Service Orders' (1985) 6 *Crime and Justice: An Annual Review of Research* 51; Norval Morris and Michael Tonry, *Between Prison and Probation: Intermediate Punishments in a Rational Sentencing System* (1990) ch 6.

Much more needs to be said about the details of this account: about its grounding in a particular kind of liberal-communitarian political perspective; about its portrayal of offenders and their relationship to the state and to their fellow citizens; about its implications for the sentencing of different types of crime; and about its relation to the actualities of our existing penal systems. Most of that will have to remain unsaid here,[10] but by responding to some of Bagaric and Amarasekara's objections I will be able to provide further clarification.

B *Punishment and Coercion*

Bagaric and Amarasekara ascribe to me the claim that 'punishment is not coercive'.[11] That would be a curious claim to make: it seems an obvious or even defining feature of criminal punishment that it is coercive, at least in the sense that it is imposed on offenders whether they consent to it or not. We can distinguish punishments that are simply *inflicted* on an offender (for example, she is taken to prison, or has a fine deducted from her salary), from those that are rather *required* of her (she is required to report regularly to a probation officer or to undertake community service). In the latter kind of case, the offender has an active role to play — she is not simply the passive victim of coercive force. It is still true, however, that it is not up to her whether she is punished or not, and that if she fails to undertake what is required of her, she will be liable to further, and in the end simply inflicted, sanctions.

However, I have never denied that punishment is coercive. What I *have* denied is that it should aim to coerce the offender's moral understanding or assent. Punishment must address the offender as a rational, responsible moral agent, and as a fellow citizen of the normative political community. This means in part that, whilst it should aim to *persuade* him to recognise and repent his wrongdoing, it must in the end be left up to him to attend or to refuse to attend to that moral communication, and to be persuaded by it or not. We must not seek to bully, terrorise or manipulate the offender into submission, but only to persuade him, as a moral agent, to recognise that he has done wrong. I am thus not committed to 'morals by force' — to an attempt to *beat* better moral values into offenders.[12] Punishment requires, and may in the end force, them to undergo what is intended to be a morally persuasive process and to *hear* a moral message, but it should not aim to force them to *accept* that message.

Bagaric and Amarasekara also object that, since punishment 'is imposed not negotiated', I cannot portray it as 'a communicative dialogue.'[13] Now, what is imposed on a person can still constitute an attempt to communicate with her; it can still address her understanding and seek a response mediated by that understanding. As to 'dialogue', there should certainly be room in the criminal process for the offender's voice to be heard: at her trial, in deciding her sentence,[14] and

[10] However, see above n 6 for further references.
[11] Bagaric and Amarasekara, above n 2, 171, 177.
[12] Ibid 178.
[13] Ibid 173.
[14] Duff, *Punishment, Communication and Community*, above n 6, ch 4.3.2.

through her response to her punishment. There are of course strict limits on what she will be heard to say and, in the end, neither verdict nor sentence are up to her: the court, speaking for the law and for the community whose law it is, claims the authority to determine these matters. However, enforced claims to effective authority do not make dialogue of a morally significant kind impossible. I can, for instance, engage in philosophical dialogue with my students, addressing and respecting them as rational beings, whilst still claiming the authority to require work from them, and to assess and grade that work.

C *The Ideal and the Actual*

It should be clear that my account is not offered as an explanation or justification of our existing penal practices. It would be absurd to suggest of many of the punishments which offenders now suffer (especially imprisonment) that they aim or serve to induce repentance, reform and reconciliation, and hard to deny that they often amount (in fact if not by design) to precisely the kind of oppressive bullying or terrorising that my account forbids. However, I have always insisted that what I offer (as any normative theory of punishment offers) is an account of what punishment *ought to be*, not a comforting justification of the penal status quo. If criminal punishment is to be adequately justified, I argue, it must be a communicative process of the kind I describe: the fact that the punishments imposed by our existing systems are not of that kind shows not the inadequacy of my account as a normative theory of punishment, but the radical imperfections of those penal systems. Similarly, a utilitarian who argues that punishment should be a cost-effective way of achieving certain social benefits is not defeated by evidence that our existing penal systems are not cost-effective means to such benefits: she will simply and rightly argue that that evidence shows our existing systems to be inadequate.

It *might* be an implication of my account that criminal punishment is not justified in our existing political, legal and penal circumstances, though I am now more inclined than I was to think that some kind of penal practice can (imperfectly, hesitantly, qualifiedly) be justified, even in our present situation.[15] Even if that is an implication, it does not show the account to lack 'practical relevance'.[16] Of course, if we believed that punishment, in something like the forms in which it is currently imposed and administered, *must* be justifiable, and that the task of normative penal theory is to find that justification, an account which held that it was not justifiable would thereby rule itself out of consideration. We have, however, no reason to believe that: we cannot dismiss a priori either the absolute abolitionist claim that punishment cannot in principle be justified, or the contingent abolitionist claim that it cannot be justified in our present situation.[17] Nor indeed do such claims lack 'practical relevance', since an

[15] Contrast Duff, *Punishment, Communication and Community*, above n 6, ch 5, with Duff, *Trials and Punishments*, above n 4, ch 10.

[16] Bagaric and Amarasekara, above n 2, 172.

[17] See also Duff, 'Penal Communications', above n 6, 67–87.

argument which shows some existing human practice to be unjustifiable has obvious practical relevance.

What *is* true is that any normative penal theorist must have something to say about the practical implications of her theory, and that if it is an implication of her account that present penal practices are unjustifiable, she must have something to say about what should be done — about the direction in which those practices should be reformed, or about what should replace them if they cannot be so reformed as to become justifiable. In *Trials and Punishments*, I suggested pessimistically that we might not be in a position from which we could directly aim to institute a system of communicative, penitential punishment, and that, if we also think that we must still punish criminals, we should perhaps use punishment as a rational deterrent subject to rigorous side-constraints.[18] Even then, the ideal of punishment as an exercise in moral communication would have 'practical relevance': partly because it should, by reminding us of the radical imperfection of our penal practices, induce a salutary caution, humility and restraint, and partly because it should motivate us to work towards the kinds of radical moral and political change which would make it possible for punishment to become what it ought to be. As I noted above, I am now more optimistic about the extent to which a conception of punishment as moral communication could directly guide our penal practices. Even without such optimism (which I still sometimes think is unfounded), however, my account would have the kind of 'practical relevance' that normative theories should have: as a basis not necessarily for justifying, but for judging, our existing practices; as a critical standard against which they must be assessed; and as an ideal to which we should in the end aspire.

D *Rights and Autonomy*

Respect for autonomy — for our fellow citizens as rational, autonomous moral agents — is central to a familiar kind of liberal political perspective.[19] For those who take seriously the demands of autonomy (that we both respect and foster the autonomy of our fellow citizens), a central question about criminal punishment must be whether it can be consistent with (or even, more ambitiously, expressive of) a proper respect for the autonomy of those who are punished or threatened with punishment.

One simple answer to that question, to which Bagaric and Amarasekara seem drawn,[20] is that punishment is consistent with a *due* respect for the autonomy of those who are justly punished, since those who break the law thereby forfeit their claim to our respect,[21] and we may therefore treat them in whatever way will

[18] Duff, *Trials and Punishments*, above n 4, 291–9.

[19] See Duff, *Punishment, Communication and Community*, above n 6, ch 2, where I sketch a species of liberal communitarianism which recognises autonomy as a central value.

[20] Bagaric and Amarasekara, above n 2, 174.

[21] For versions of this argument, see Alan Goldman, 'Toward a New Theory of Punishment' (1982) 1 *Law and Philosophy* 57; Christopher Morris, 'Punishment and Loss of Moral Standing' (1991) 21 *Canadian Journal of Philosophy* 53. Cf Phillip Montague, *Punishment as Societal Defense* (1995) 2–4; Duff, *Punishment, Communication and Community*, above n 6, ch 1.3.1.

cost-effectively prevent further crimes. Now, if punishment is to be justified at all, the culpable commission of a crime must obviously make *some* difference to the offender's moral standing, rights and entitlements. We must also ask whether there are any wrongs which are so destructive of the bonds of human community that the wrongdoer no longer has any claim on our respect or concern. It does not seem plausible, however, that the commission of *any* crime, even a serious crime, should thereby forfeit the offender's basic claim to be respected as an autonomous and responsible agent.

That is why I have argued that punishment, as an exercise in moral communication, must be consistent with continuing respect for the offender's autonomy. This requires me to argue that neither the kind of coercion which punishment involves, nor the attempt to persuade the offender to repent of his crime which punishment as moral communication involves, necessarily infringes autonomy. The argument about moral persuasion was sketched above in Part II(B). The argument about coercion appeals in part to an analogy between punishment and the use of force in self-defence, and to the crucial moral difference between using force to prevent a person carrying through a wrongful attack in which he is already engaged, and using force to prevent a future attack on which we believe he will embark. The former respects his autonomy, since it is justified as a defensive response to what he has already, as a responsible agent, begun to do: he has, we could say, brought it on himself. The latter, by contrast, infringes his autonomy, since it does not leave him free to decide for himself whether or not to embark on the attack.

To appeal to this analogy is not to justify punishment on the model of self-defence; such a justification is not plausible.[22] It is rather to show that coercion need not infringe autonomy if it can be justified as an appropriate response to the other's actions. Punishment, I argue, can be justified in precisely this way if it is a matter of forcefully censuring the offender, of requiring him to undertake (or if necessary imposing on him) the apologetic reparation that he owes to those he has wronged, and of trying to persuade him to recognise and repent of his crime.

Another way to reconcile punishment with respect for the offender's autonomy would be to argue that offenders have a *right* to be punished, or that it accords with what they ('really') want.[23] Bagaric and Amarasekara exaggerate the significance of my comments on the idea that offenders 'want' to be punished[24] — those comments were not intended to *justify* punishment by an appeal to the offender's wishes, but only to explain the particular sense that that idea

[22] For different versions of such a justification, see Laurence Alexander, 'The Doomsday Machine: Proportionality, Punishment and Prevention' (1980) 63 *The Monist* 199; Daniel Farrell, 'The Justification of General Deterrence' (1985) 94 *Philosophical Review* 367; Warren Quinn, 'The Right to Threaten and the Right to Punish' (1985) 14 *Philosophy and Public Affairs* 327; Montague, above n 21. For criticism, see Duff, *Punishment, Communication and Community*, above n 6, ch 1.3.2.

[23] It is not enough to argue that punishment aims to benefit offenders, since that would not by itself avoid the charge of autonomy-violating paternalism. Although I did argue this in *Trials and Punishments* (above n 4, 250–66; see Bagaric and Amarasekara, above n 2, 179–80), and do still believe that punishment should benefit the offender by reconciling him with the community to which he belongs, I no longer see this as a proper *justifying aim* of punishment: see Duff, *Punishment, Communication and Community*, above n 6, ch 3.4.1.

[24] Bagaric and Amarasekara, above n 2, 176.

can make *within* a conception of punishment as penance.[25] I should, however, say something here about the idea of a right to be punished, which they also criticise.[26]

To talk of a 'right' to be punished is to talk of punishment as something that is owed to the offender (not just, for instance, to the victim or the wider community), and as something that is supposedly for her own good (not just as something that she deserves). Perhaps it is also to imply that punishment is something that the offender would claim for herself, if she realised the truth. That might indeed sound strange. However, it will seem less strange when we look at the alternatives to punishing the offender.

One alternative would be simply to ignore her crime, but that is not a morally available option. For if her crime is, as crimes under a morally justified system of criminal law must be, a public wrong that properly concerns the community as a whole, to ignore it would be to condone it. We owe it to the victim (when there is one), to ourselves and to the values by which our political community is supposedly defined (values flouted by the crime), not to condone such wrongs. We could also say that we owe it to the offender to take her seriously as a responsible moral agent, which includes censuring her wrongdoing; to ignore or to condone those wrongs would be implicitly to deny her standing as a responsible moral agent.

Another alternative would be to subject her to some kind of measure which simply aimed to prevent her from repeating such wrongdoing — for instance by deterring her, by re-forming her, or by incapacitating her. Such measures might or might not be classified as punishments: that would depend in part on the procedures through which, and on the criteria by which, they were implemented. However, a central argument in favour of retributivism in general, and of the communicative version of retributivism that I espouse in particular, is that other methods of coercive, preventive treatment for offenders (whether or not they count as 'punishment') fail to respect their status as responsible moral agents,[27] whereas retributive punishment respects that status. If that argument is sound, and if ignoring the offender's crime is not an option, then we can say that we owe it to the offender, out of respect for her as a responsible moral agent, to punish her for her crime — which is to say that punishment is her right.

I do not claim to have said enough here to show that we can properly talk of a right to be punished: I have tried only to indicate the kind of argument that would make legitimate sense of such talk. It might still be objected, however, that at best I could hope to show that censure or criticism is the offender's right, not that punishment, involving penal 'hard treatment', is her right. This brings us to the final criticism from Bagaric and Amarasekara that I want to discuss — their charge that my account offers no adequate justification for penal hard treatment.

25 Duff, *Trials and Punishments*, above n 4, 269–71.

26 Bagaric and Amarasekara, above n 2, 174.

27 See Duff, 'Penal Communications', above n 6, 9–12, and further references given there.

E *Justifying Hard Treatment*

Censure can be communicated by direct speech, for instance, by what the judge says to the convicted offender. It can also be communicated by purely symbolic punishments, which are burdensome or painful *only* by virtue of their condemnatory meaning. For example, a pickpocket whose punishment was to have a sign saying 'I stole from my fellow citizens' stuck to his bathroom mirror for a fixed period would be pained by it only if he were pained by the condemnation it implied.[28] Censure *can* also be communicated through 'hard treatment' punishments — punishments that are burdensome or painful independent of their condemnatory meaning. Imprisonment, fines, community service or probation orders, and other familiar sentences are burdensome or painful even for someone who is unmoved by or unaware of their condemnatory meaning as *punishments*, but they can also be intended and understood as methods of communicating the censure that the offender's crime deserves. However, as Bagaric and Amarasekara point out,[29] a communicative theorist must explain why the message to be communicated must be communicated through hard treatment punishments, rather than just through direct speech or through a system of purely symbolic punishments.

On my account, the hard treatment dimension of punishment serves the communicative aims of punishment in three related ways.[30]

First, it is a way of focusing the offender's attention on her crime, of trying to bring her to face up to her crime and its implications, and of overcoming our familiar tendency to turn our attention away from matters that are uncomfortable or that we do not care about as we should. The aim, or the hope, is that this will bring the offender to recognise and to repent of her crime as a wrong, to which Bagaric and Amarasekara object that I cannot justify punishing offenders who have already repented, or those who are 'beyond saving'.[31] However, apart from the fact that undergoing punishment can deepen and strengthen a repentance that might otherwise be shallow or incomplete, the punishment of an already repentant offender can assist in her self-reformation, and in reconciling her with those whom she has wronged (see below). As for the offender who, we are sure, will not be brought to repentance, I have claimed that we should never give up on a fellow citizen as 'beyond saving'. Bagaric and Amarasekara argue that this commits me to an 'intrinsic retributivism' which seeks no end beyond the infliction of punishment.[32] This is to misinterpret my view. Punishment is justified as an attempt to bring the offender to repent, and my claim is that we should make that attempt even if we are sure that, given the offender's intransigence, it will fail: we owe it to the offender to continue to treat her as someone who is within the reach of moral communication, and not to dismiss her as

[28] Of course, once punishments are public, it is harder for them to be *purely* symbolic, for the offender is then likely to suffer through the responses of others who become aware of his crime through his punishment.

[29] Bagaric and Amarasekara, above n 2, 180.

[30] Duff, *Punishment, Communication and Community*, above n 6, chh 3.4–3.6.

[31] Bagaric and Amarasekara, above n 2, 181.

[32] Ibid.

beyond redemption. Bagaric and Amarasekara might not agree that we owe offenders this continuing respect and concern. Nonetheless, by trying to ascribe to me an intrinsic retributivism as my 'real justification' for punishment,[33] they ignore the role that that conception of the offender's moral standing, as someone who is redeemable, plays in my account.[34]

Second, suitable hard treatment punishments can assist the process of moral self-reform which communicative punishment also aims to become: they can be vehicles through which offenders can come, and be helped to come, to see how they can so reform themselves as to avoid such wrongdoing in the future. Bagaric and Amarasekara object that such claims lack 'empirical support': that the 'high rate of recidivism amongst those who have experienced hard treatment suggests that hard treatment is more likely to cause anger, frustration and a regression in one's moral health rather than repentance and reform'.[35] Such an objection would have force if my claim were that the familiar kinds of hard treatment punishment which are salient in our existing penal systems actually serve to induce repentance and self-reform, but that is not my claim (see Part II(C) above). My claim is rather that suitably designed and administered kinds of hard treatment punishment *should*, and *in principle* could, serve those aims. Whilst that claim can find impressionistic or anecdotal support in accounts of some probation programmes, some kinds of community service order and other kinds of penal measure, such as the CHANGE project for violent men,[36] it does not depend on proof that our existing penal systems serve those aims.

Third, hard treatment punishments can serve as reparation for the moral wrong that was done, made to those whom the offender wronged. They constitute a kind of enforced or required apology that is given greater weight by being thus expressed and, therefore, also serve to reconcile offenders with those whom they have wronged. In response, Bagaric and Amarasekara argue that sincere apology in itself should suffice to provide the necessary reparation; that to demand hard treatment as well is to fall back again on an intrinsic retributivism which simply demands that the offender suffer more pain;[37] and that hard treatment punishments are ill-suited to the expression of the sincere apology that reconciliation requires, since we cannot tell whether the offender is sincerely apologetic.[38] The answer to the first objection is that something more than a merely verbal apology is needed to make it clear to the victim, and to the wider community, that the wrong done is being taken seriously — just because a merely verbal apology can all too easily be superficial, glib or less than serious. The answer to the second

[33] Ibid.

[34] For more on the repentant and the defiant offender, see Duff, *Punishment, Communication and Community*, above n 6, chh 3.7.3–3.7.4.

[35] Bagaric and Amarasekara, above n 2, 181.

[36] Duff, *Punishment, Communication and Community*, above n 6, ch 3.5. The CHANGE project aims, through confrontational group work, to challenge domestically violent men, to bring them to accept responsibility for their violence and to help them to change their ways: see R Emerson Dobash and Russell Dobash, *Women, Violence and Social Change* (1992) ch 7; Jonathan Scourfield and Russell Dobash, 'Programmes for Violent Men: Recent Developments in the UK' (1999) 38 *Howard Journal of Criminal Justice* 128.

[37] Bagaric and Amarasekara, above n 2, 183.

[38] Ibid 182.

objection is that, with a proper liberal respect for the offender's privacy (for the privacy of his conscience), the state should not try to determine whether the offender's purported repentance is sincere. If the offender has undertaken or undergone the requisite kind of penitential punishment, the state and his fellow citizens should accept that apology without inquiring more closely into its sincerity.[39]

I am painfully aware that my replies to the objections raised by Bagaric and Amarasekara have often consisted largely of gestural sketches of arguments, together with references to my more recent publications on punishment. However, it is not really possible to do more than this in a short reply. I hope that I have said enough to give readers a sufficiently clear idea of my account of punishment for them to decide whether there might be more to say for it than Bagaric and Amarasekara allow, and whether it would be worth their while to read more of it. I turn now to another aspect of Bagaric and Amarasekara's article: their advocacy of a strictly utilitarian theory of punishment.

III PUNISHMENT AND THE MAXIMISATION OF PLEASURE

Bagaric and Amarasekara ground their theory of punishment on hedonistic act utilitarianism,[40] the doctrine that an action is right so long as it maximises pleasure (the only intrinsic good) and minimises pain (the only intrinsic evil). I will comment briefly on the character of this justification of punishment; on their response to the familiar objection that utilitarians must sanction the deliberate punishment of the innocent; and on their failure to respond to the objection that utilitarian theories of punishment also deny the *guilty*, and all those who are threatened with punishment, their proper moral standing.

A *Justifying Punishment*

If a hedonistic utilitarian is to justify a particular action or practice, she must be able to argue not only that it is an *effective* means of causing pleasure or preventing pain, but also that it is at least as efficient a means of producing such good or preventing such evil, in the prevailing circumstances, as any available alternative.

Bagaric and Amarasekara ground the justification of punishment in its crime-preventive efficacy, in particular in its efficacy as a general deterrent.[41] They are appropriately sceptical about its efficiency as a method of incapacitation, rehabilitation or specific deterrence. Now it is indeed plausible that many systems of punishment are *effective* as general deterrents: there are some who refrain from crime because of the threat of punishment, and who would commit crimes were that threat removed. It is also plausible that systems of punishment could provide *efficient* general deterrence, at least in the minimal sense that their

[39] Duff, *Punishment, Communication and Community*, above n 6, chh 3.4.2, 3.6.1.
[40] Bagaric and Amarasekara, above n 2, 130.
[41] Ibid 139.

crime-preventive benefits can outweigh their costs.[42] However, a utilitarian needs to do more than appeal to such common sense beliefs about punishment as a deterrent.

First, she needs to be able to work towards an account of what kinds of conduct should be criminal, of what modes and degrees of punishment should be threatened against each of those kinds of conduct and (a distinct question) imposed on those who are proved to have engaged in them. She must also consider what level of resources should be devoted to the detection, prosecution and punishment of crime, the procedures through which, and the criteria by which, guilt should be determined, and so on. She needs, in effect, to work towards a complete account of a justified system of criminal law and punishment.[43] That is of course true of any theorist of criminal law and punishment, but it presents particular difficulties for a utilitarian, especially for a hedonistic utilitarian. For she needs to find a way of working out how much pleasure, and how much pain, various possible practices and policies are likely to produce — and then somehow add the pleasures, add the pains, and weigh them against each other. However, when we try seriously to think what this would involve, we will realise that such a utilitarian calculus is a fantasy. It is not just one that it would in practice be very difficult to carry out — it is one that it would be absurd even to think of trying to carry out.[44]

Second, utilitarians must also ask whether we should have a system of criminal law and punishment at all. *Given* a system of criminal law that operates in part as a deterrent, we can agree that its immediate abolition — the sudden removal of the threat of punishment — would probably be disastrous,[45] but that is not enough to justify its maintenance. Utilitarians must ask whether there are other possible methods of preventing socially harmful conduct which could replace, and be more efficient than, punishment.[46] To fail to address this question, to assume that we must maintain, and must thus be able to justify, a system of criminal law and punishment, is to be guilty of 'begging the institution'.[47] Again, a version of this question faces any normative theorist of criminal law, but it poses a particular problem for the hedonistic act utilitarian, since answering it requires a calculus of pleasures and pains that it is, to put it mildly, hard to envisage as being possible even in principle.

[42] For a useful survey of recent evidence, see Andrew von Hirsch et al, *Criminal Deterrence and Sentence Severity* (1999).

[43] John Braithwaite and Philip Pettit, *Not Just Deserts* (1990) 12.

[44] It is also worth noting another implication of hedonistic act utilitarianism of the kind that Bagaric and Amarasekara espouse, which can be illustrated by the following (extreme) example. Imagine two rapes. In one, the rapist acts out of anger or despair, and in fact gains no pleasure from his crime; in the other, the rapist does gain pleasure from his crime; in both, we can suppose, the effects on the victim and others are roughly similar. A hedonistic act utilitarian must apparently judge the second rape to be morally preferable to the first, since he must say, 'At least that rape brought about some good — the rapist's pleasure'. Are we really to accept a moral view according to which the pleasure a rapist gains from his crime is, in itself, a moral consideration in favour of his action?

[45] Bagaric and Amarasekara, above n 2, 140.

[46] Duff, *Trials and Punishments*, above n 4, 104, 164–72.

[47] See M M Mackenzie, *Plato on Punishment* (1981) 41.

B *Punishing the Innocent*

Bagaric and Amarasekara offer a two-pronged response to the familiar objection that utilitarians must sanction the deliberate punishment of the innocent when this would serve the preventive aims of punishment. First, they 'outsmart' the critics by arguing that the deliberate punishment of an innocent *would* be justified if it really would maximise utility.[48] Second, they argue that retributivists must also, if they are to justify punishment at all, justify systems which will (by mistake) punish some innocents, and that retributivist attempts to avoid the objection which they claim is fatal to utilitarianism are doomed to failure.[49]

As to the first point, the simple counterargument is Hart's: even if we would, in some extreme situation, recognise that an innocent 'must' be punished, we would also realise that this involved a significant moral cost, a significant wrong done to the innocent person, which the utilitarian cannot recognise.[50] More generally, the objection to utilitarianism on this score is not so much that utilitarians would in the end be committed to actions that should not be done (even in emergencies), but rather that they display a morally inadequate grasp of the *reasons* that bear on such actions. Since pain is the only intrinsic evil, they cannot recognise the intrinsic wrong of injustice that is done to an innocent scapegoat for what it is. They must also count the chance that the scapegoating will be found out as a relevant (and sometimes no doubt conclusive) reason against it, whereas the critic argues that the wrongfulness of the injustice does not depend on whether it is found out or not.

As to the second point, Bagaric and Amarasekara rightly note that a retributivist response must appeal to some version of the 'doctrine of double effect':[51] the retributivist must claim that the punishments of the innocent under a retributivist system are foreseen but not intended, and that the system can therefore be justified as being just — whereas the deliberate punishment of the innocent to which utilitarians must be committed is categorically unjust. This is not the place for a full discussion of that principle; nor do I (nor need retributivists) support all the ways in which it has been used, many of which are certainly spurious. However, it is worth pointing out, first, that even if there are some cases in which it is not only hard in practice to determine what an agent intended, but in principle impossible to draw a clear and uncontroversial distinction between intention and foresight,[52] this does not show that the doctrine of double effect

[48] Bagaric and Amarasekara, above n 2, 141–4. See J J C Smart, 'An Outline of a System of Utilitarian Ethics' in J J C Smart and Bernard Williams, *Utilitarianism: For and Against* (1973) 1, 69–72; and D C Dennett (ed), *The Philosophical Lexicon* (8th ed, 1987): 'outsmart, v. — To embrace the conclusion of one's opponent's *reductio ad absurdum* argument. "They thought they had me, but I outsmarted them. I agreed that it *was* sometimes just to hang an innocent man"', available at <http://www.blackwellpublishers.co.uk/LEXICON/default.htm> at 1 August 2000 (copy on file with author).

[49] Bagaric and Amarasekara, above n 2, 144–9.

[50] Hart, above n 1, 81.

[51] See R A Duff, 'Retributive Punishment — Ideals and Actualities' (1991) 25 *Israel Law Review* 422, 435–41.

[52] Bagaric and Amarasekara give no indication as to why we should think that there is such a difficulty in principle: see, eg, Bagaric and Amarasekara, above n 2, 147. For an account of the

cannot be properly applied when that distinction can be drawn. Second, whilst the doctrine does serve to allow deontologists to attend to consequences in their practical reasoning, this does not make their views consequentialist,[53] since consequentialists hold that *only* consequences *ever* matter. Third, there does seem to be a significant moral difference between the deliberate punishment of someone who is known at the time to be innocent (this being what utilitarians are accused of being ready to sanction), and the mistaken punishment of someone who is in fact innocent, but has been proven guilty by procedures which include reasonable safeguards against mistaken convictions (this being what retributivists must accept as a feature of any human system of punishment). The retributivist tries to do justice, whilst recognising that we will sometimes fail; furthermore, she is ready to recognise, to admit and to make reparation for such mistakes as occur, if they can be detected (whereas a utilitarian committed to punishing an innocent for the sake of the greater good would presumably also be committed to trying to prevent detection or publicity). The utilitarian, critics argue, is not even trying to do justice.

C *Doing Justice to the Guilty — and Others*

The revival of positive retributivism brought with it a new kind of objection to utilitarian and other consequentialist theories of punishment. This objection applied even to 'mixed' theories which built in side-constraints forbidding the deliberate punishment of the innocent, since the objection concerned the treatment of the *guilty* and of other citizens as well.[54]

The objection takes slightly different forms in relation to the different ways in which, for a consequentialist, punishment can bring its justifying benefits — by incapacitation, by reform, or by deterrence. Since Bagaric and Amarasekara focus on deterrence, I will also focus on the version of the objection which applies to a deterrent system of punishment — even to one that builds in side-constraints forbidding the deliberate punishment of the innocent (or the excessively harsh punishment of the guilty).

Suppose we start with the moral demand that we should respect each other, and that the state should respect its citizens, as responsible, autonomous moral agents. Suppose we also accept that we, and the state, will sometimes want to bring others to behave, or not to behave, in certain ways — in particular, not to commit crimes. What constraints does that moral demand set on the ways in which we, or the state, can properly try to achieve that aim?[55]

distinction between intention and foresight, and of the rare cases in which it is in principle difficult, see R A Duff, *Intention, Agency and Criminal Liability* (1990) chh 3–4.

53 Bagaric and Amarasekara, above n 2, 145.

54 See, eg, Jeffrie Murphy, *Retribution, Justice, and Therapy* (1979) 94; Herbert Morris, 'Persons and Punishment' (1968) 52 *The Monist* 475; Andrew von Hirsch, *Censure and Sanctions* (1993) ch 2; Duff, *Trials and Punishments*, above n 4, ch 6; Duff, *Punishment, Communication and Community*, above n 6, chh 1.2–1.3.

55 For simplicity's sake, I talk here as if the demands of respect for responsible autonomy set side-constraints on our pursuit of the (consequentialist) goal of modifying behaviour. In the end, however, I think we should understand those demands as helping to determine the very ends we should pursue, and not just the means by which we may pursue them; Duff, *Punishment, Communication and Community*, above n 6, ch 3.2.2.

It requires first that we seek to modify citizens' future behaviour only by offering them *reasons* to act as we want them to act or think they should act. We must not modify their behaviour by mere force (by incapacitating them, for instance), or by manipulating their attitudes or desires in ways that seek to bypass, rather than appeal to, their capacities for practical reasoning (as some kinds of 'reform' might do). The capacity to guide our own actions in the light of our grasp of reasons for action is central to our character as responsible, autonomous agents: to treat and respect another as a responsible, autonomous agent therefore requires us to respect that capacity rather than to undermine or bypass it. A deterrent system of punishment satisfies *this* requirement, since it offers potential offenders reasons to refrain from crime, which it hopes they will find persuasive. It is effective just insofar as potential offenders see the threat of punishment as a good reason to refrain from crime.

A second, related requirement is honesty — that we do not try to deceive the other person about the reasons she has for (or against) acting as we think she should act. This is because deception is a kind of manipulation, which seeks to subvert the other's rational capacities. A strictly consequentialist system of deterrence faces a problem here, because a strict consequentialist (and in particular a hedonistic utilitarian) cannot see honesty as having more than a contingent and instrumental moral importance. Suppose, for instance, that by suitable manipulation or presentation of data, we could make it appear to the public at large that the chances of being caught, convicted and punished for a crime were much higher than they actually are. This looks like a tactic that a utilitarian should seriously consider, since it seems to offer a cost-effective way of increasing the law's efficiency as a deterrent,[56] but it flouts the moral demand that we treat and respect others as responsible, autonomous agents. It seeks to subvert the rational capacities of those who are threatened with punishment: rather than enabling or helping them to decide, on the basis of accurate information, how they are to behave, we try to manipulate them by feeding them false information.

Even without such straightforward dishonesty, however, the use of punishment and the threat of punishment as a deterrent are arguably inconsistent with the demand that we respect others as responsible, autonomous agents. Hegel claimed that such a penal system treats one who is punished, or threatened with punishment, 'like a dog instead of with the freedom and respect due to him as a man'.[57] One way to explain this thought is by arguing that to respect another as a responsible, autonomous agent, it is not enough that we try to modify her conduct (that is, to persuade her to modify her own conduct) only by offering her reason to do so: the reasons we offer her must also be appropriate and relevant

[56] Alternatively, perhaps we could make it appear that the sentences served by convicted offenders were much harsher than they actually are: cf Braithwaite and Pettit on the difference between 'the *Realpolitik* of implementation' and the 'symbolic politics' of 'denunciation and reprobation': above n 43, 176–7. As Bagaric and Amarasekara note, however, it is far from clear that increasing (or appearing to increase) penalty levels improves the law's deterrent efficacy: Bagaric and Amarasekara, above n 2, 137–8; see also von Hirsch et al, above n 42.

[57] Georg Hegel, *Hegel's Philosophy of Right* (first published 1821, 1942 ed, T M Knox trans) 246. See also Duff, *Trials and Punishments*, above n 4, 178–86, and Duff, *Punishment, Communication and Community*, above n 6, chh 1.3.2, 3.1.2, 3.2.2, and 3.3.1.

reasons. Now the appropriate and relevant reasons for refraining from crime are connected with the wrongfulness of the criminal conduct in question, and with the legitimacy of the law's claim on our obedience and allegiance. It is for those reasons that we think others should refrain from crime, and those reasons that justify our or the state's attempts to persuade them to do so. A deterrent system of punishment does not appeal to those kinds of reasons for refraining from crime; rather, it creates a new and purely prudential reason, the threat of punishment, for so refraining. However, that is to fail to treat those who are thus threatened as responsible, autonomous agents: it is to fail to address them as moral agents, in terms of the relevant moral reasons which should motivate them not to commit crimes, and to address them instead in amoral terms as merely self-interested beings.

This objection, which accuses consequentialist theories of punishment of failing to give due weight not merely to the rights of the innocent, but to the moral standing of the guilty (and indeed of all those who are threatened with punishment) has not, admittedly, persuaded everyone; nor do I have space to do justice to it here. My point is simply this: that the moral objections to consequentialist (including side-constrained consequentialist) penal theories do not focus only on the rights of the innocent, but also concern the way in which the guilty are seen and treated — and, more generally, the way in which the state addresses and treats all its citizens through a system of criminal law and punishment. An adequate defence of a consequentialist theory of punishment must at least address this kind of objection; Bagaric and Amarasekara simply ignore it.

My conclusion is therefore that there is more to be said for a retributivist account of punishment as an exercise in moral communication than Bagaric and Amarasekara allow, and more to be said against their purely utilitarian account of punishment than they recognise.