WHO MAY CARRY OUT PROTECTIVE DETERRENCE?

BY MICHAEL SPRAGUE

Anthony Ellis argues that institutional punishment occurs automatically in a way analogous to mechanical deterrents, and given that issuing real threats is justified for self-defence, institutional punishment, intended to protect society via deterrence, can be justified without violating the Kantian constraint against using persons as means only. But institutional punishments are in fact not executed automatically: they must be carried out by moral agents. Ellis fails to provide a basis for those agents to justify the performance of their legal duties.

Anthony Ellis proposes a deterrence theory of punishment which, he says, can avoid the common objection that punishing for deterrence commits us to using persons unacceptably. Deterrence theory holds that punishment is justified if it will deter potential criminals from committing crimes in the future (and if the harm prevented by deterrence outweighs the harm caused by the punishment). Kantians and others have objected that deterrence theory violates the constraint against using persons as a means only, for the punished offender is treated only as a means to preventing future crimes. Ellis aims to avoid this charge without abandoning consequentialism. While his theory may avoid the constraint against using persons, it suffers from another difficulty: while it seems (at first) to permit legislators to issue deterrent threats (in the form of laws), it does not provide justification for agents to carry out the punishments required by those laws.

Ellis presumes that we have the right to use force only in cases of self-defence (construed to include third-party defence of others), and that we also have the right to threaten the use of force in order to deter others from acting aggressively. After all, he says (p. 338), ‘it must surely be better to try to prevent aggression rather than have to deal with it forcefully when it occurs’. The difficulty, then (p. 339), is in justifying the execution of a threat that fails to deter aggression:

The justification for threatening is that it will prevent aggression (and any harm involved in a self-defensive response). If the threat does not work, however, that justification will not carry over into a justification for carrying out the threat, for carrying out the threat will, by hypothesis, do nothing to prevent these harms.

Ellis' attempt to avoid this difficulty rests on drawing an analogy between institutional punishment and automatic retaliation systems. On the premise that automatic systems are justifiable, he concludes that institutional punishment is justifiable.

The notion of automatic retaliation systems has been explored by Warren Quinn.\(^2\) He imagines a situation in which we can construct machines that will automatically execute retaliatory harms against anyone who violates our rights. Once activated, the machines cannot be deactivated. ‘M-punishment’, as he calls the automated retaliation, is therefore out of the control of any agents as soon as the initial choice to activate the system has been made. According to Quinn (p. 339), this reduces the choice of whether to impose retaliatory harms (when threats fail) to a derivative choice, derived from the ‘basic’ choice to set up the threat. He goes on to argue that we have ‘both practical and moral reasons to create a real threat if we create any’, because a deceptive threat would be unlikely to work, and because it would be ‘morally insupportable’ for civic authorities to ‘practise wholesale deception in a matter as vital as this to each citizen’s interest’. But (p. 339–40)

It does not ... follow from the fact that our threat is real and will, therefore, almost certainly lead to m-punishments that our initial justification [for activating the system] must appeal to the anticipated deterrent effects of those m-punishments. If such an appeal were necessary, then the future m-punishments themselves, along with the threat of them, would enter our scheme of justification as means of protection. But the justification we seek makes no such appeal. We see that each m-punishment will occur as the result of the previous existence of a real threat, and we insist that each such prior threat can be completely justified by reference to the protection that it can be expected to create.

In other words, only one decision is made by each agent – to create a real threat by activating the system, and this decision is justified by its deterrent effects. The m-punishments themselves do not need individual justification, for they proceed automatically from a failure of the threat.

Granting for the sake of argument that Quinn’s moral conclusions about m-punishment are correct, I can now examine Ellis’ protective-deterrence theory. Ellis sees institutional punishment as a sort of ‘semi-automatic’ system. Statutes serve as deterrent threats, and if the threat is ignored, punishment ‘goes forward fairly automatically’ (Ellis, p. 341):

... one agent might be authorized to apprehend suspected aggressors, another might be authorized to decide whether they really were aggressors, and another might be authorized to administer punishment if appropriate. But no one would have authority to deactivate his part of the system except in special circumstances.

(The ‘special circumstances’ alluded to here seem to be simply cases of limited discretion, which, as Ellis rightly points out, occur at all levels of the punishment...

systems of his model cases, the US and the UK. This is what makes punishment
‘semi-automatic’, rather than automatic like m-punishment.)

This may seem like automation from the viewpoint of the legislator who
initially makes the deterrent threat, but it is not automation in the way Quinn’s
m-punishment is. Quinn cautions us (p. 358) to remember that his ‘automatic de-
vices, while marvellously sophisticated, are not persons responsible for authentic
moral choices’. It is precisely this consideration that Ellis has overlooked. For the
agents carrying out punishment, unlike Quinn’s machines, are persons responsible
for moral choices, and although no agent would have legal authority to deactivate his
part of the punishment system, this in itself does nothing to establish his moral
authority to execute his legal responsibilities.

The essential question, then, is this: by what justification may an agent, acting on
behalf of the state, inflict the harms required by law on a convicted offender? (I shall
henceforth call this agent ‘the corrections officer’.) Ellis’ protective deterrence theory
does not seem to offer an answer. The corrections officer cannot be justified by the
potential deterrent effects of the punishment, for then the protective deterrence
theory would fail to avoid the constraint against using persons. And, although a
legislator might, perhaps, be justified in making a real threat (in the form of law)
because subsequent enforcement of the threat is apparently automatic, the correc-
tions officer who must physically place an offender behind bars cannot similarly
justify his act, for his act is not automatic.

The only other option to rescue the theory seems to be to argue that the cor-
rections officer is justified in executing punishments by some considerations not
related to punishment at all, such as an obligation to the state to carry out his duties.
This strategy also fails, however. Clearly not just any promise to act will morally
oblige an agent to act – a hit-man is not justified in killing someone by the fact that
he has promised to do it. So to make sense of this justification for the execution of
the punishment, one would have to argue that the corrections officer is under
obligation to carry out his duty to the state at least in part because the state’s order is
itself justifiable. But by Ellis’ conception, the state’s order is justifiable only in so far
as the retaliation it threatens is inflicted automatically when the threats are ignored.
This leads to the absurd conclusion that the corrections officer is morally permitted
to inflict the punishment at least in part because that very punishment is inflicted
automatically.

Ultimately, assuming the premise that lawmakers are justified in making de-
terrent threats only if the punishment of offenders will proceed automatically (when
those threats fail) entails, by modus tollens, the conclusion that lawmakers are not
justified in making deterrent threats – because the further premise, that punishment
does in fact proceed automatically, is false.

Florida State University

© The Editors of The Philosophical Quarterly, 2004
Copyright of Philosophical Quarterly is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.