PHILOSOPHY

THE JOURNAL OF THE ROYAL INSTITUTE OF PHILOSOPHY

VOL. XXIX. NO. 111 OCTOBER 1954

"THE JUSTIFICATION OF PUNISHMENT"

ANTONY FLEW, M.A.

I

I WANT to discuss philosophically, to glance at the logic of, the parts of this expression "the justification of punishment" and then to draw from this discussion one or two morals for discussions of the justification of punishment. This paper is based on one originally given to the Scots Philosophy Club at its Aberdeen meeting in 1953, as the third part of a symposium on *The Justification of Punishment* (no inverted commas).

II

(*a*) *Punishment.* (i) This term is both vague and 'open-textured' (Waismann).[1] Vague; because in several directions there is no sharp line drawn at which we must stop using it: when does punishment of the innocent or illegal punishment cease to be properly called punishment at all? (Here we must beware the scholasticism which F. P. Ramsey attacked in Wittgenstein, when the latter insisted that we *cannot* think illogically. For though there does come a point at which 'thinking' is so illogical, 'punishment' so wayward, that we should refuse to call them thinking or punishment at all, there is nevertheless a wide margin of toleration. And of course, as usual when we say there comes *a point at which*, what we really mean is that there comes *a twilight zone after which*.) Open-textured; because many questions of its applicability could arise over which even full knowledge of current correct usage might leave us at a loss. Not because this case fell within a more or less recognized No-Man's-Land of vagueness across which no sharp line had been

[1] *Logic and Language.* (Edited Antony Flew), Vol. I, pp. 119–20.

drawn but because it was of a sort which had simply not been envisaged at all: would it be punishment if no effort was made or even pretended to allocate the 'punishments' to the actual offenders, but only to ensure that the total of hangings, say, balanced the total of murders; irrespective of who was hung? (See Ernest Bramah[1] on thus "preserving equipoise within the Sacred Empire.") A third feature, which partly overlaps both the other two, is that several logically independent criteria are involved. Ideally these are all simultaneously satisfied, but there is no strict unanimity rule here to paralyze action: so the word may be applied, and correctly, where one criterion is definitely not satisfied (and not merely where, through its vagueness, there is doubt as to whether it is or is not satisfied).

Once these features, which this concept shares with so many others, are recognized it becomes clear that it would be well as a prelude to possible discussions of the ethics of punishment to list the criteria. This usually useful preliminary is in this case exceptionally important. Both because there is—since ideas here have certainly developed and are still changing and controversial—every reason to expect there are minority users of the word "punishment": that some people will insist that certain elements are essential which others will not so regard. And because in ethical controversy the temptation to produce from up one's sleeve at later stages in the argument apparently decisive definitional jokers is very strong: and can only be removed by making clear from the start what is and what is not to be involved in the central notions.

In listing the criteria satisfied by what, without honorific intentions,[2] we may call a standard case of punishment, in the primary sense of the word, we have to realize: both—as we have already mentioned—that there are some non-standard users with their private variations on this primary use; and that there are secondary uses of the word (with which those intending to discuss the ethics of punishment are not directly concerned); and that it is correctly applied even by standard users in its primary sense to non-standard cases of punishment, i.e. cases in which not all the criteria are satisfied, but which because of its vagueness the word can cover.

I am going to present my remarks here as *proposals*. Not because I regard them as arbitrary; for on the contrary they are based on what I take to be the general or at least the dominant tendencies in current usage; though I shall not give as many illustrations as would otherwise be necessary, because Mabbott has already done a large part of this work, in his "Punishment" (in *Mind* 1939: this is quite

[1] *The Wallet of Kai Lung*, *Kai Lung's Golden Hours*, and *Kai Lung Unrolls His Mat*.

[2] There is, for instance, nothing honorific in saying that this car is a standard model; not 'custom built' or bespoke.

the most valuable article I know on this subject). But because it needs sometimes to be emphasized that no philosophical analysis of the meaning of any term worth so analysing can ever leave things exactly as they were—however conservative the intentions and protestations of its protagonists. For it must necessarily tend to change the meaning for us and our usage of the terms it analyses—ideally by precisifying it and making clear its implications.

(ii) I *propose*, therefore, that we take as parts of the meaning of "punishment", in the primary sense, at least five elements. *First*, it must be an evil, an unpleasantness, to the victim. By saying "evil" —following Hobbes—or "unpleasantness" not "pain", the suggestion of floggings and other forms of physical torture is avoided. Perhaps this was once an essential part of the meaning of the word, but for most people now its employment is less restricted. Note in this connection the development of an historically secondary use of the word; as applied first to a battering in boxing, then extended to similar situations in other sports where there is no element of physical pain (e.g. as an equivalent to "trouncing", of bowling in cricket).

Second, it must (at least be supposed to) be for an offence. A term in an old-fashioned public school, though doubtless far less agreeable than a spell in a modern prison, cannot be called a punishment, unless it was for an offence (unless perhaps the victim was despatched there for disobedience at home). Conversely, as Mabbott most usefully stresses, if a victim forgives an offender for an injury which was also an offence against some law or rule, this will not necessarily be allowed as relevant to questions about his punishment by the institution whose law or rule it is. A mnemonic in the 'material mode of speech': "Injuries can be forgiven; crimes can only be pardoned".

Third, it must (at least be supposed to) be of the offender. The insistence on these first three elements can be supported by straightforward appeal to the *Concise Oxford Dictionary*, which defines "punish" as "cause (offender) to suffer for an offence".

Notice here that though it would be pedantic to insist *in single cases* that people (logically) cannot be punished for what they have not done; still a *system* of inflicting unpleasantness on scapegoats—even if they are pretended to be offenders—could scarcely be called a system of punishment at all. Or rather—to put it more practically and more tolerantly—if the word "punishment" is used in this way, as it constantly is, especially by anthropologists and psychoanalysts,[1] we and they should be alert to the fact that it is then used in a

[1] Cf. e.g. J. C. Flugel, *Population, Psychology, and Peace*, pp. 70–1. "Another germane example is the stigma of 'illegitimacy', and this example illustrates the important fact that the punishment or suffering in question need not necessarily be endured by the culprit" and "There is such a thing as vicarious punishment".

metaphorical, secondary, or non-standard sense: in which it necessarily has appropriately shifted logical syntax (that is: the word in this case carries different implications from those it carries in a standard case of its primary sense). A likely source of trouble and confusion.

Fourth, it must be the work of personal agencies. Evils occurring to people as the result of misbehaviour, but not by human agency, may be called penalties but not punishments: thus unwanted children and venereal disease may be the (frequently avoided) penalties of, but not the punishments for, sexual promiscuity. To the extent that anyone believes in a personal God with strong views against such sexual behaviour, to that extent he may speak of these as divinely instituted punishments (though, allowing the linguistic propriety of this, the fact that so often the punishments fall on the innocent and can be escaped by the guilty should give pause still).

(Note here. First the distinction often and usefully made—but rarely noticed even by those making it—between the 'natural' penalties *of* and the prescribed penalties *for* such and such conduct (gout *of* port-bibbing: free kicks *for* fouls). Second, that the expression "to the extent that" is peculiarly appropriate to beliefs about God and a quasi-personally sustained moral order in the universe: for with most people these meander somewhere between complete conviction and complete disbelief; and hence to offer the present distinctions between the uses of "punishment", "penalty of" and "penalty for" as if these were already completely given in present (correct) usage would be seriously to misdescribe the confused situation which actually confronts us.)

Fifth, in a standard case punishment has to (be at least supposed to) be imposed by virtue of some special authority, conferred through or by the institutions against the laws or rules of which the offence has been committed. Mabbott brought this out clearly. A parent, a Dean of a College, a Court of Law, even perhaps an umpire or a referee, acting as such, can be said to impose a punishment; but direct action by an aggrieved person with no pretensions to special authority is not properly called punishment, but revenge. (Vendetta is a form of institutionalized revenge between families regarded as individuals.) Direct action by an unauthorized busybody who takes it upon himself to punish, might be called punishment—as there is no unanimity rule about the simultaneous satisfaction of all the criteria—though if so it would be a non-standard case of punishment. Or it might equally well be called pretending (i.e. claiming falsely) to punish. The insistence on these fourth and fifth criteria can be supported by appeal to the *Oxford English Dictionary* which prefaces that same definition as that given in the C.O.D. with "As an act of superior or public authority".

Besides these five positive criteria I *propose* negatively that we

should not insist: *either* that it is confined to either legal or moral offences, but instead allow the use of the word in connection with any system of rules or laws—State, school, moral, trades union, trade association, etc.; *or* that it cannot properly be applied to morally or legally questionable cases to which it would otherwise seem applicable, but instead allow that punishments, say, under retrospective or immoral laws may be called punishments, however improper or undesirable the proceedings may be in other respects. Laxity in both these directions conforms with normal usage; while in the second it has the merit of separating ethical from verbal issues.

I shall say nothing about "collective punishment" except that: while no doubt the original unit on which punishment was inflicted was not the individual but the family, tribe, village, clan, or some other group; nevertheless for most of us today "collective punishment" is somehow metaphorical or secondary: it is a matter of regarding a group as an individual, for certain purposes.[1]

(*b*) *Justification.* This term is multiply relational. A justification has to be of A, rather than B, against C, and to or by reference to D; where A is the thing justified, B the possible alternative(s), C the charge(s) against A, and D the person(s) and or principle(s) to whom and/or by reference to which the justification is made. The variables may have more than one value even in one context: there may, for instance, be more than one charge. But they do not all have to be given *definite* values *explicitly*. Indeed the point of saying all this lies precisely in the fact that in most cases of justification the values of some of the variables are given only implicitly by the context, and perhaps rather indefinitely too: hence, just as in the notorious case of motion, it is possible to overlook (some of the implications of) the relational nature of the concept. The alternative(s), for instance, may be unstated and even very hazily conceived: but that there must be at least one alternative is brought out by considering that "There is no alternative" is always either a sufficient justification or a sufficient reason for saying that the question of justification does not arise. Finally, and most important —again compare the case of motion—the reference point, the fourth variable (the person(s) or principle(s) by reference to which justification is made) has the same rather indefinite value implicit in most actual contexts: the value "(whom I consider) reasonable people, and what (*fundamentally*) they agree on".

[1] Compare here (*a*) K. R. Popper's examination of the appropriateness and limitations of the metaphor involved in regarding rogue states as criminals: *Open Society*, Vol. I, pp. 242 ff. (*b*) the Nuremberg Trials: a colossal effort to discover what the guilt of the Nazi régime amounted to in terms of the particular guilt of particular Germans. This point I owe, like so much else, to Mabbott, privately, and later to his *The State and the Citizen*.

Presumably this has contributed to the use of "justification" as a near synonym for "reason for". Which in turn has been a very minor determinant of the modern fashion—for which there is much to be said—of presenting moral philosophy as an enquiry into what are and are not *good reasons* in ethics. But note here. First, that the word "justification" may, even in contexts where all the variables have the same values, be used in two relevantly different ways: either implying that the proposed justification is or not implying that it is morally or otherwise acceptable to the user; in the latter case if it is *very* unacceptable the word may be put in protest quotes. (*Mutatis mutandis* the same is true of "reason".) Second, that this mode of presentation tends to conceal the existence of really radical ethical disagreements. This point has been pressed in a critical notice of S. E. Toulmin's *The Place of Reason in Ethics* by J. Mackie (*Australasian Journal of Philosophy* 1951). Presumably Toulmin would answer, on Kantian lines, that he was elucidating the *nature of ethics as such*: from which the fact that certain reasons were relevant and good, and others irrelevant and bad, followed necessarily. This would imply that those who seem to be doing ethics but admit different reasons to that extent cannot be doing ethics at all, or at least are very unreasonable people: by definition. Which is perhaps fair enough: providing that steps are taken to bring out just what must be involved in rejecting the definitions implicitly accepted by Toulmin, and that many would reject some of the implications of these definitions, and that their position in so doing is monstrous. To say this last is abandoning pure analysis to take sides, as all men must, in a struggle: making a normative, participant's utterance; and not a purely analytic, neutral's observation.

(*c*) *The.* The assumption behind the use of the definite article is that there is one and only one (unless the whole expression is interpreted, as it rarely is, as strictly equivalent to "justifying punishment"). This is questionable twice over, at two levels: first, because the variables admit of various values—what would serve as justification against one charge and for a Roman Catholic could be simply irrelevant against another and for an atheist humanist; and, second, because in any one context (i.e. where the same values are given to all the variables) there may be two logically separate acceptable justifications both independently sufficient. And surely this is not merely possible but likely: for the fields of human causation, motivation, and justification are precisely those in which 'overdetermination' (Freud) is most common. (An action is said to be *overdetermined* when at least two motives were at work to produce it, either of which alone would have been sufficiently strong to do so separately. The concept, *mutatis mutandis*, obviously can and should also be applied to matters of causation and justification.)

III

We come now to some applications of ideas suggested in our outline examination of the parts of the expression "the justification of punishment" to ethical discussions of the justification of punishment. We shall do this mainly by reference to Mabbott's paper, mentioned above; but we shall have an eye to other contributions especially those made to the Aberdeen symposium, also mentioned above.

(*a*) *Theorizing and justifying.*—Traditionally, views about the circumstances, the severity, and the forms in which punishment is justified—and why—have been presented as *theories* of punishment. Sometimes the metaphor has been developed by speaking of justification as providing "a theory to square with the facts (our moral convictions)".[1] This idiom seems to me radically misguided. First, because it conceals the essentially relational character of justification, which makes it an entirely different sort of thing from theorizing in the positive sciences (see II(*b*) above): suggesting, for instance, that the work of justification *could* be completed finally, for ever, and for everyone.

Second, because it misrepresents questions of value as questions of fact or philosophy: and insofar as these 'theories' are intended as justifications it is wrong to present them as enquiries into why we do, or what are our reasons for doing, what we do. Mabbott begins "I propose . . . to defend a retributive theory and to reject absolutely all utilitarian considerations from its justification" (p. 152). In supporting this he writes: "The view that a judge upholds a bad law in order that law in general should not suffer is indefensible" (p. 157). This piece of judicial psychology is irrelevant: surely not "upholds" but "should uphold" is meant. It is quite difficult enough to disentangle factual, ethical, and analytical questions without adopting an idiom which obscures the differences between these, and encourages such confusing slips as this one.

Third, because it conceals the dynamic character of fruitful ethical discussion about justification. It suggests that two men embarking on such a discussion must start from a static, given, unalterable set of "facts" (their moral principles or convictions) and go on thrashing out together the implications in this field of these principles. Whereas in fact these "facts" (convictions) are often modified in the course of the discussion itself: just because that brings out unacceptable implications of or reveals unnoticed inconsistencies between these "facts" (convictions). (Compare the mistake made by Aristotle in *Nichomachean Ethics*, III. iii. (1112A18ff.), where his account of deliberation is similarly static and over-formal:

[1] One of the Aberdeen symposiasts.

tending, except at § 13 (1112B25–27), to overlook that men, being imperfect seers of implications and not omniscient, are often led to modify their objectives in the course of and in the light of deliberation about the means to reach these.)

Fourth, because it embodies and hides certain questionable assumptions: that "our moral convictions" are in agreement, and that they are unchanging. Both are false. Though in both cases, of course, it depends a lot on whom "our" is referring to. But with the former assumption, even taking British professional philosophers as the us-group, it is difficult to believe that debate about the ethics of contraception, abortion, homosexuality, and suicide, would not reveal differences both in the weight given to different admitted *prima facie* obligations, and even perhaps in what were admitted as obligations at all. For instance, if I may follow one bit of unconventionality, the mentioning of such subjects in connection with philosophical ethics, by another, the mention of religious gulfs—the Roman Catholics among us are surely committed to the (at least almost) unconditional repudiation of all these as always morally wrong; and this not as purely religious tabu but as a matter of natural (as opposed to revealed) obligation. In the case of the latter assumption, one would hope that some of the moral convictions of us especially are open to alteration by argument.

Both assumptions are peculiarly questionable in connection with punishment. Not merely do the ethical views of different people differ pretty considerably, as we can see from the continuing arguments about the proper purpose, justification, and reform of punishment; but the same person often holds inconsistent views at different times or even at the same time. Who of us can manage to be consistent about the relative weight of retributive and utilitarian considerations: and this not merely because we are always being tempted to give more weight to the former than we should in a cool hour adjudge proper, whenever we are emotionally involved against the offender; but also because even in cool hours it is so hard to be sure and steady about difficult particular concrete problems of conduct.

(*b*) *Punishment as necessarily retributive.*—"Why should *he* be punished?" asks about the punishment of a particular person on a particular occasion. This is the type of question with which Mabbott was concerned: he formulated his problem as "Under what circumstances is the punishment of some particular person justified, and why?" (p. 152); and gave the answer "The only justification for punishing any man is that he has broken a law" (p. 158). This is supposed to be a retributive view and to be an ethical matter: as is made clear in the sentence already quoted "I propose to defend a *retributive* theory, and to reject absolutely all utilitarian considera-

tions from its *justification*" (my italics). Which would lead us to expect arguments, or at least assertions, to the effect that the vicious (or the criminal) deserve, and ought to be made, to suffer for their wickedness (crimes): without regard to the public advantage of such a system. But—to my mind fortunately—this is not at all what we are given. Roughly: insofar as Mabbott's view can be called retributive it is not a justification (satisfactory or otherwise);[1] and insofar as any sort of justification is offered it is ideal (not hedonistic) utilitarian.

(i) Unfortunately, Mabbott never makes clear how far he is appealing to the meanings of words and how far to "our moral convictions": the confusion is easy for the latter are often incapsulated in the former; and expressions like "what we should say" are ambiguous as between moral and linguistic propriety. But it does seem as if his answer to his main question is intended to depend upon the very meaning of the word "punishment". Yet insofar as this is so he is not really offering a justification, based on retributive ethical claims; but a necessary truth drawn from, and elucidating the meaning of, "punishment".

It is interesting to compare here the position of F. H. Bradley: "Punishment is punishment only where it is deserved. We pay the penalty because we owe it, and for no other reason. If punishment is inflicted for any other reason whatever than because it is merited by a wrong, it is a gross immorality, a crying injustice, an abominable crime." (*Ethical Studies*, pp. 26–7). This is similar to but not the same as Mabbott's position. Not the same; because Mabbott "dissents from most upholders of the retributive theory—from Hegel, from Bradley, and from Dr. Ross" (*loc. cit.*, p. 154) on the grounds that the essential link is between "crime" and "punishment" and not between "doing wrong" and "punishment". Similar; because both are confusing necessary truths with ethical claims: if Bradley's first and second sentences express necessary truths then the ethical claims made in the third are out of place; for if this is so, then however gross the immorality, crying the injustice and abominable the crime it *cannot* be punishment at all if it is not both deserved and paid because owed.

Furthermore, not only is Mabbott's "retributive justification"—insofar as it rests on an appeal to the meaning of "punishment"—not what it pretends to be, but this appeal itself cannot be made out completely. First, because while a *system* of 'punishing' people who had broken no laws could not be called a system of punishment, the term "punishment" is sufficiently vague to permit us to speak

[1] Significantly, C. W. K. Mundle at Aberdeen and in a revised version of his paper to be published in the *Philosophical Quarterly* was unable to recognize Mabbott as a fellow retributionist.

in single cases and providing these do not become too numerous (if they do become too numerous then *ipso facto* the use, the meaning, of "punishment" has changed) of punishing a man who has broken no law (or even done no wrong). This objection might be met by saying that the term already is more precise than we have allowed, or by now deciding to make it so. The effect of adopting the latter alternative—like that of all such manœuvres with meanings—would be of course to shift and not to solve the *ethical* problems. (Consider the Rousseauian dialogue: Q. I know how to recognize the General Will: but is it always morally sound? A. The General Will is always upright—by definition. Q. But now how do I recognize it—so defined?) Second, because Mabbott wrote that "the *only justification* for punishing any man is that he has broken a law" (my italics); and it is surely impossible to draw the italicized point out of the present meaning of "punishment". One can only say that the question "Did he do it?" is always relevant to the question "Ought he to be punished?", or even that it is the only question which is always, because *necessarily*, relevant: appealing for this to the meaning of "punishment" as "an evil inflicted on an offender for an offence". But we cannot maintain that there is a contradiction involved in attempting to justify a single punishment for any reason other than that the victim has broken a law. Of course, this objection too could be met by a suitable adjustment of the meaning of "punishment"; though this adjustment would be both greater and harder to defend than that required to meet the former objection; and it would likewise involve a (not necessarily undesirable) shifting of any *ethical* problems, which cannot be dealt with by any manœuvres with definitions.

(ii) Insofar as Mabbott's solution to his problem is intended as an ethical claim it is open to grave objections. Taken in this way it is a claim that, while it does make sense to speak both of punishing those who have broken no law and of doing so for other reasons than this (with other justifications than this); the *sole* and *always requisite true* justification is that the victim has committed an offence. (Mabbott did not add "sufficient": presumably in order to allow for the possible justifiability of pardons and unjustifiability of enforcing certain laws.)

The objections both depend on what might be canonized as The Principle of the Multiplicity of Ethical Claims. Everyone who is not a one-track fanatic—which Mabbott emphatically is not—recognizes as *prima facie* valid grounds of obligation several ethical rules or claims: (indeed it might be said that a creature recognizing one and only one was not doing ethics at all, even fanatically). But to recognize several rules or claims which are logically independent is to open the way to both over-determination and conflict: for it is to concede that it is at least logically possible that there can

be circumstances both such that in them one claim reinforces another; and such that obeying one claim means disobeying another, satisfying one involves overriding another. And, fortunately in the former case, notoriously in the latter, with the set of claims which people actually do recognize, such circumstances do occur.

The objection to saying that the *sole* justification for punishing someone is that he has committed an offence is that Mabbott and almost everyone else would in fact allow that a punishment in certain circumstances was overdetermined in its justification—was justified twice over. Certainly: because, though Mabbott claims to "reject absolutely all utilitarian considerations from its justification", he is prepared to appeal to these to justify *systems* of punishment. But if a *system* is to be justified even partly on such grounds, some cases within that system must be partly justifiable on the same grounds: the system surely could not have effects which no case within it contributed. (See III(*c*) below.)

The objection to saying that this justification is *always* requisite is that, if you allow any moral claims other than that only offenders ought to be punished, then in certain circumstances one of these might be conceded to have overriding force; and hence to justify the punishment of an innocent man; and hence to show that this retributive justification is not always requisite. Unless, of course, you are prepared to insist that no one who has not committed an offence ought ever to be 'punished' for it: though the heavens fall. And though it is probably true that the evil effects of such injustice will almost always outweigh the good; still, unless we can believe in some Providential guarantee that this is *always* so we cannot accept such a claim, and at the same time accept other claims which logically might sometimes override it,—the claim for instance to do what one can to prevent the fall of the heavens or some similar catastrophe.

We shall not argue further for these objections or provide illustrations: because Mabbott would no doubt be willing to admit them; and because this paper is in any case primarily concerned only with the prolegomena to ethical discussion. But one easy but mistaken assumption calls for notice: to show that something is just (or unjust) is not always and necessarily to show that it is justified (or unjustified); in spite of the common root and similar appearance of the words "just" and "justification". The sentence before "The only justification for punishing any man is that he has broken a law" was "Any criminal punished for any one of these reasons (*pour encourager les autres*, etc.—A.F.) is certainly unjustly punished". But the former is no sort of restatement of the latter: something may be just but open to all sorts of (moral or legal) objections; (and of course something may be just and it may be

https://doi.org/10.1017/S0031819100067152 Published online by Cambridge University Press

possible to justify it on further grounds). Again on p. 154 Mabbott seems to be making the same assumption, where he argues that even the most excellent consequences cannot "prove the punishment was just": for though this is true, they may nevertheless be conceded to *justify* the injustice.

(iii) A third interpretation of Mabbott's solution, which is perhaps nearest to what he actually had in mind when he wrote (as opposed to what he perhaps, or really, or ought to have, meant) is that "The only justification for punishing any man is that he has broken a law" amounts to an assertion that, if we are considering the punishment of a particular person on a particular occasion *and are accepting that the penal system is generally all right*, the sole consideration relevant to the question "Would it be justified or not?" is "Did he commit an offence for which this is the penalty prescribed?" (Omitting for the sake of simplicity and with Mabbott, questions of extenuation and excuse.) Whereas if we are considering the merits of a particular law, or the advantages of having a system of laws, or of a penal enforcement of those laws (as opposed perhaps to psychiatric enforcement by the compulsory treatment of offenders); then other and, particularly, utilitarian considerations are relevant.

This interpretation involves the most important distinction, which it is one of the chief merits of Mabbott's paper to underline, between systems and particular cases within those systems.[1] He criticizes, for instance, a writer who "confuses injustice within a penal system with the wrongfulness of a penal system" (p. 160) and later maintains that "it is essential to a legal system that the infliction of a particular punishment should *not* be determined by the good *that particular punishment* will do either to the criminal or 'society'. . . . One may consider the merits of a legal system or of a credit system but the acceptance of either involves the surrender of utilitarian considerations in particular cases as they arise" (pp. 162–3).

The first objection to the position thus interpreted is that what Mabbott is saying in the passage just quoted is only *necessarily* true in the great majority not in all, particular cases: for there is no *contradiction* involved in saying that you accept a system but propose nevertheless to allow an occasional exception. This notion of "necessary truth in a great majority of cases" sounds scandalous and contradictory. It is perhaps a stumbling block; but it is not contradictory. It is just what is often needed in and appropriate to the "informal logic"[2] of the vague, elastic, concepts of everyday

[1] cf Hume *EPM* App. III 256.

[2] This phrase is borrowed from G. Ryle, "Ordinary Language", in *Philosophical Review*, April 1953: an invaluable exposition of the actual views and assumptions of those philosophers who "care what dustmen say".

discourse (as opposed to the strict rigid "p" and "q" and "triangle" of the formal disciplines of mathematics and symbolic logic). It might have been used in II(*a*) above to make the point that, in the present meaning of "punishment" it is not incorrect to speak of punishment where one or other criterion for what we called a standard case of punishment is not satisfied, so long as these exceptional cases remain exceptional: we might have said that it is necessarily true in the great majority of cases that a punishment satisfies each (and all) of the criteria listed.

This being so, it is not possible to say that anyone accepting that there ought to be a system has committed himself to saying that utilitarian considerations are *necessarily* irrelevant to this or that particular case. The most that can be said—and this is very well worth saying—is that such a person is thereby committed to the surrender of utilitarian considerations in particular cases *as a rule*, a rule to which there can be exceptions. But, unfortunately, exceptional cases do not arrive labelled as such: so it is always open to someone to say that *this* is the case where an exception ought to be made.

Mabbott's short way with misplaced utilitarianism, therefore, will not work. His mistake here can be seen as yet one more example of our perennial and pervasive failure to realize that the concepts of everyday discourse have not, and, if they are to do the sort of jobs they are needed to do, cannot have, the sharp outlines and rigidity of the concepts of the calculi of mathematics. Since this short way will not work perhaps it would be wise to re-examine, in a parallel excursus of our own, the sort of move which he handles so toughly in his *Excursus on Indirect Utilitarianism*.

"When I am in funds and consider whether I should pay my debts or give the same amount to charity, I must choose the former because repayment not only benefits my creditor . . . but also upholds the general credit system" (pp. 155–6). "The view that a judge upholds ("ought to uphold"?—A.F.) a bad law in order that law in general should not suffer is indefensible" (p. 157). Mabbott has two arguments against these appeals to the *indirectly* utilitarian (ethical) advantages of keeping promises, enforcing laws, repaying debts, in cases either where the *direct* results seem unlikely to be the best from a utilitarian point of view or where the *indirect* benefits may increase the balance of *direct* advantage. First, with Ross, he claims that the indirect consequences of single breaches of rules have been exaggerated. Second, with great emphasis as supposedly decisive, he claims that the indirect disadvantages of breaches are the consequences of people getting to know of the breaches, not of the breaches as such. "It follows that indirect utilitarianism is wrong in all such cases. For the argument can always be met by 'Keep it dark' " (p. 157).

But this, too, is too short a way with dissenters. The first claim is perhaps all very well; but, of course, a series of exceptions in single cases can add up to the effective abandonment of a general rule: and Mabbott himself is emphatic about the possible utilitarian advantages of having even a bad system of laws rather than no system at all. A beach is made up of single grains of sand. The second claim too is sound enough, but again the "indirect utilitarian" could argue that while the occurrence of single exceptions can, as a matter of fact, be kept dark (although it is almost impossible in *this* particular case to be sure that we shall succeed in suppression) it is, as a matter of fact, impossible to hush up the occurrence of a large number of exceptions: and large numbers are made up by the accumulation of units.

Of course such indirect utilitarian arguments[1]—depending as they do on *contingent* facts, about the actual effects upon the maintenance of a system of particular breaches of the prescriptions of that system, and about the possibilities of hushing up the occurrence of such breaches, cannot show that it is never and *necessarily* impossible to justify, upon utilitarian principles, the making of an exception by a breach of the prescriptions of a system, which can itself be justified upon utilitarian principles. To adapt one of Mabbott's own examples: suppose someone, having willed his property to be disposed of by a friend in accordance with private oral instructions, makes that friend promise to devote the money to a futile purpose; then, especially if the friend were the only witness of the making of that promise, even when all possible weight had been given to indirect utilitarian considerations of the importance of maintaining the system of promise-keeping and confidence in promises made; still it might be that the balance of good, on utilitarian principles, was overwhelmingly on the side of breaking this particular promise, diverting the money to a beneficent purpose, and keeping the matter dark. Mabbott takes this as the *reductio ad absurdum* of any utilitarianism (ideal or hedonistic or what have you). Perhaps it is paradoxical, and perhaps it is repugnant to the present moral convictions of many or even most of us. But still a utilitarian might very well accept and even glory in the paradox: insisting that this was precisely the sort of case in which "our moral convictions" ought to be reformed.

"Reform" is surely a key word here: for at any rate the classical Utilitarians were concerned not merely with the reform of institutions but with the reform of ethical ideas and ethical reasoning.

[1] It is the possibility of such indirect utilitarian arguments which ensures that properly thought out utilitarianisms are self-regulating doctrines. See I. M. Crombie "Social Clockwork" in D. M. MacKinnon's *Christian Faith and Communist Faith* (Macmillan 1953), especially pp. 109 f.

Mill's *Utilitarianism* is devoted to deciding "the controversy respecting the criterion of right and wrong" ("Everyman", p. 1): and phrases like "a test of right and wrong" (p. 2), "the moral standard set up be the theory" (p. 6), and "the utilitarian standard" significantly recur again and again. "Though the application of the standard may be difficult, it is better than none at all: while in other systems, the moral laws all claiming independent authority, there is no common umpire entitled to interfere between them" (p. 24). Mill refuses "to enquire how far the bad effects of this deficiency have been mitigated in practice, or to what extent the moral beliefs of mankind have been vitiated or made uncertain by the absence of any distinct recognition of an ultimate standard" (p. 8): but his phrasing makes clear his belief that there was a need for reform which he was hoping to meet. The burden of his charge against "the intuitive school of ethics" and of the corresponding claim for its "inductive" rival rests on the failure of the one and the success of the other in producing the test or standard required. And the acceptance of this alone makes possible a corresponding reform in ethical reasoning: "Whether happiness be or be not the end to which morality should be referred—that it should be referred to an *end* of some sort, and not left in the dominion of vague feeling or inexplicable internal conviction, that it be made a matter of reason and calculation, and not merely of sentiment, is essential to the very idea of moral philosophy; is, in fact, what renders argument or discussion of moral questions possible" (Essay on *Bentham*).[1]

To return to the objections to Mabbott's solution in its third interpretation. The second objection is that on examination this resolves itself into one or the other of, or a confusion of, the other two. For either it amounts to saying that punishment is necessarily retributive, in the sense in which this thesis has already been examined and explained (in III(*b*) (i) above): the assertion that the sole consideration relevant to the question "Would it be justifiable to punish him?" is "Did he commit an offence for which this is the penalty prescribed?"; depending—allegedly—upon the very meaning of "punishment". Or it amounts to the same thing as the ethical

[1] Perhaps we should emphasize here that, of course, Mill (though not Bentham) was very insistent indeed about the great importance of "secondary principles", and the proposed reform did not consist in or include anything so monstrous as the suggestion that we should "endeavour to test each individual action directly be the first principle" (p. 22). There is no need to expatiate here, in view of J. O. Urmson's recent powerful attack on these and other popular misconceptions of Mill (*Philosophical Quarterly*, 1953). But for a grim warning of the results of accepting this suggestion which Mill was not making, see Arthur Koestler's *Darkness at Noon*, on the Party "travelling without ethical ballast". Here the self-regulator—see footnote above—was not in use.

claim (examined in III(*b*) (ii) above) that the sole and always requisite true justification for punishing a man is that he has committed the appropriate offence: but masquerading as a piece of logic (see II(*b*) above on *good reasons* in ethics). Decisions as to what is and is not to be allowed to be relevant in ethical discussions can—when the consideration concerned is not *necessarily relevant* (cf. III(*b*) (i) *ad fin*)—themselves be value decisions. Consider how my (really ethical) refusal to take into account the effects on, say, Egyptians or negroes of my actions could be given a logical look by saying that these are *irrelevant*: by refusing to admit as any sort of *reason* in any discussion about the right thing to do any statement of the form "That would harm E, an Egyptian, or N, a negro".[1]

To sum up this section (III(*b*)): the upshot is that in general there must be a 'retributive' element in punishment, inasmuch as punishments to be punishments must be of an offender for an offence; though this is not a matter of universal logical necessity but only of 'necessary truth in the great majority of cases', since there can be occasional exceptions. Mabbott's main mistakes were, I think: first, to insist that punishment is *necessarily and always* 'retributive' in this (non-ethical) sense; second, to think that such a supposedly necessary truth about punishment (or anything else) could constitute a retributive or any other sort of moral justification of it (or anything else). To attempt to justify from the concept alone is like trying to prove existence from the concept alone—The Ontological Argument. Though there is this at least to be said for it: that if "our moral convictions" are to be accepted as the arbiter, then the attempt to justify from the concept alone will amount to an appeal to the popular moral convictions incapsulated in ordinary language.

(*c*) *Overdetermination.*—Sometimes it seems to be assumed that there must be an inconsistency in justifying the adoption and enforcement of a law or a system of laws by *both* utilitarian *and* retributive appeals. But there is no necessary inconsistency in this, any more than there is in having advocated the nationalization of coal but now opposing the nationalization of chemicals or cement: though of course the people making these appeals or combining these policies may say other things, or offer supporting reasons, which do involve them in inconsistency. Thus, having a law against murder, or against any other sort of behaviour which is reckoned to be wrong whether it is made illegal or not, can be defended *both* on the grounds that this makes for commodious living and against lives nasty, brutish, and short *and* on the grounds that it makes for

[1] "Mr. Lyttelton has helped to force through a federation based on the admitted policy of regarding African political opinions as irrelevant" (*Observer*, 23/8/53).

wicked men getting their deserts. (I suspect that not only would most people resort to both sorts of arguments, but that Mill too would have done so—though he would have said that any "secondary principle" of retribution for ill-desert had *ultimately* to be justified by reference to the "first principle", the "Greatest Happiness Principle". I cannot hope to make good the historical claim here: but I refer again to Urmson's most excellent paper; and to the fact that Mill made "the turning point of the distinction between morality and simple expediency" that for wrong-doing "a person ought to be punished" (*loc. cit.*, p. 45).)

Perhaps philosophers have been misled into this assumption by undertaking to find a general comprehensive justification for all justified punishments; deceptively described as "a theory of punishment". This must commit them to producing: either an ethical claim which is to be insisted on in every case and to which no exceptions whatever will be admitted; or a necessary truth which obtains the universality required only at the cost of ceasing to be any sort of ethical justification.

IV

We have tried in this paper to bring out the main features of the logic of the expression "the justification of punishment"; and to apply the lessons to one outstanding paper about the justification of punishment. We have not attempted, except insofar as was absolutely necessary, incidentally, to do any actual justification on our own account. Indeed the entire paper may be considered as prolegomena only: except that they are prolegomena of the sort which suggest that the latter enterprise—in the form of a comprehensive and universal enquiry as opposed to either a series of piecemeal jobs done in particular contexts or the attempt to find generally useful principles to which there will be occasional exceptions—was misconceived.

Kings College, Aberdeen.