Philosophical Perspectives on the Insanity Defense

1 The insanity defense and philosophy
More has been written on the insanity defense than on any other single topic in criminal law,¹ and yet statistically the plea of insanity is fairly rare, being considerably less common as a courtroom phenomenon than most laymen might expect.² A reason for this apparent antinomy is that the insanity defense raises assumptions and issues that are at the very heart, the central philosophical core of the criminal law³ and, in the view of some, challenges the conceptual basis on which the criminal law rests. At the present time the criminal law is coming under increasing fire from theorists in the social sciences and social technology, who argue that it is based on an outmoded, unscientific and irreparable conceptual scheme, that of 'autonomous man', man as a free agent. These theorists would have us get on with the business of scientifically shaping behavior rather than bothering to pursue unfathomable inquiries into 'the guilty mind'.⁴ Recent changes in legislation and precedents in the courts reflect a growing awareness and concern about the underlying philosophical conceptions that pose such a stumbling-block to clear and productive thinking on pressing issues related to the insanity defense, such as involuntary commitment, criminal responsibility, mental illness, psychiatric testimony and the role of experts in the law, the concept of the dangerous offender, and the McNaghten and Durham Rules in the courts.

The kind of philosophical work so badly needed in these areas is clear-headed enucleation and extraction of the essential philosophical principles, and their statement in terms understandable to those who are concerned with criminal responsibility in the social sciences and psychiatry, and with law and legislation. This is a difficult but sorely needed undertaking, and I believe that philosophers should expend greater efforts on it. I hope that this paper will be the right sort of stimulus in encouraging a greater degree of productive rapport between philosophers and others who share a concern with the problems raised by the insanity defense.

2 History of the insanity defense
There are basically four distinctly formulated standards of criminal responsibility used in the courts in cases where the insanity defense is raised. These are the McNaghten Rules, the Irresistible-Impulse test, the Durham (or product) Rule, and the test recommended by the American Law Institute's Model Penal Code.

A THE MCNAGHTEN RULES
In 1843 Daniel McNaghten, a Scot, shot Edward Drummond, principal secretary to the British Prime Minister Sir Robert Peel, believing that he had killed Peel. Drummond languished and died. McNaghten had exhibited a previous history of bizarre behavior, and was subject to delusions of persecution and hallucinations. At that time English law was in the process of developing a theory of criminal responsibility for the mentally ill. In a previous case, Regina v. Oxford (1840), where a man with a history of insanity, Edward Oxford, had shot at Queen Victoria, the jury was charged in the following terms: 'If some controlling disease was in truth the acting power within the defendant,
which he could not resist, then he will not be responsible.\textsuperscript{5} The case of McNaghten was the occasion of the formulation of a rule covering such cases, the issue being at that time quite controversial and causing heated debate, not only in the newspapers, but in the House of Lords and the House of Commons. Accordingly the fifteen judges of the common law courts were called upon by the House of Lords for information and provided answers to five questions put to them by the Lords. This set of five questions and answers constitutes the McNaghten Rules.

The first question asks for the point of law regarding crimes committed by persons with insane delusions. The response was that if the delusion is only partial then the person is punishable if he knew at the time of the act that he was acting contrary to the law of the land. The second and third questions ask in what form the issue should be presented to a jury. The criterion offered in response reads as follows: 'To establish a defense on the grounds of insanity it must be clearly proved that the accused was laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing or, if he did know it, that he did not know he was doing what was wrong.'\textsuperscript{6} Also included in this response are the points that a defendant is to be presumed sane until it is proven otherwise and that knowledge of right and wrong pertains to the specific act in question rather than to knowledge of right and wrong in general. Moreover the clarificatory point is made that actual knowledge of the law of the land is not essential to conviction (such knowledge being generally assumed under the law), so that the operative criterion is whether the defendant 'was conscious that the act was one he ought not to do'. The fourth question is whether a man who commits a crime under an insane delusion is thereby excusable - the reply is that he is to be considered as though the facts with respect to which the delusion exists were real. For example if a man kills another under the delusion that the latter is trying to kill him, he kills 'in supposed self-defense' and is thus excusable. The fifth question concerns the matter of expert (medical) testimony, and the response is that generally the medical man cannot be asked for his opinion on matters for the jury to decide, even though it may sometimes be convenient to put the question to the expert in the same general form as the question on which the jury is asked to decide.

There is an incredibly voluminous literature criticizing the McNaghten Rules for a great variety of reasons.\textsuperscript{7} Many of these emotionally charged criticisms have exhibited the most elementary kinds of misunderstanding of the statements of the chief justices. To clarify some of the chief misunderstandings, a few points should initially be made. First, the Rules are not meant to be of the 'if and only if' type.\textsuperscript{8} That is, the Rules are not, as the historical context of their formulation clearly shows,\textsuperscript{9} meant to cover every possible kind of exemption from criminal responsibility related to mental disease. The five questions were specifically addressed to the issue of insane delusions. No mention is made of other exculpatory grounds related to insanity. This does not, I think, indicate rejection of all other grounds, but indicates simply that they are not being considered. Second, the operative criteria are expressed in non-technical terms familiar to the layman, inviting broad interpretation. Nowhere is it specified that knowledge of the 'nature and quality of the act' is 'cognitive', or excludes 'emotional' awareness of the nature of the act. Thus a narrow interpretation, restricting such knowledge to 'cognitive' factors, is gratuitous. Third, the medical expert is not required to state his conclusions in terms of 'knowledge of the nature and quality of the act'. On the contrary he is specifically enjoined to eschew this form of conclusion in his testimony. It is, according to the Rules, the function of the jury to rule on this question. Thus a great many of the historical criticisms and denunciations of the McNaghten Rules are, as I see it, simply beside the point.

\textbf{B THE IRRESISTIBLE-IMPULSE TEST}

According to the American Bar Foundation Study,\textsuperscript{10} McNaghten is the sole test of criminal responsibility in fewer than half the states, but in at least fifteen states it is accepted in conjunction with the Irresistible-Impulse Test. It is \textit{prima facie} plausible that a person may lack control over his action because of mental illness even though he knows the nature and quality of that action.\textsuperscript{11} Certain schizophrenics, for example, are quite acutely cognizant of what they are doing, having an accurate awareness of the nature of the consequences of their criminal act, and may be perfectly aware of its
criminal wrongfulness. Indeed certain persons of this type evidence knowledge of the nature and quality of their act by very deliberate and careful planning. Plainly there are many such cases where, despite such knowledge, we want to say that the individual did not have control over his action because he was suffering from acute mental illness of such a nature as to deprive him of free control and hence responsibility.¹¹a Thus it seems implausible that the McNaghten Rules can cover all cases of lack of capacity due to insanity, and hence they appear to require supplementation by an additional test or rule. The Irresistible-Impulse Rule fills this lacuna. Unfortunately this rule, like the McNaghten Rules, has been the object of an avalanche of hostile criticism, based to a great extent on elementary misunderstandings. First, the term ‘irresistible impulse’ is an altogether unfortunate one, appearing to require that the crime, to be exculpated, must emanate from some sort of mysterious, interior ‘irresistible impulse’.¹² Such impulses, understandably, have proved to be impossible to define or identify. Second, the term ‘irresistible impulse’ appears to denote an inner, violent, sudden, passionate explosion, and thus appears to be an inappropriate exculpatory ground for crimes that are not sudden but are the result of brooding or reflection. But the latter, as we noted, are the very type of crime that such a rule is needed to cover. Historically a failure to agree on any univocal statement of a control rule of this type has added to the general confusion, with the result that a clear and agreed-upon statement of the rule has not emerged. The Report of the Royal Commission on Capital Punishment reflects this ambivalence, discrediting the concept of ‘irresistible impulse’ while proposing a control rule to supplement the McNaghten Rules. The Commission concludes: ‘The concept of the “irresistible impulse” has been largely discredited ... it is inherently inadequate and unsatisfactory.’¹³ Yet the Commission recommended (the minority’s primary recommendation and the majority’s secondary recommendation) that the McNaghten Rules should be supplemented by an alternative clause removing responsibility if the defendant ‘was incapable of preventing himself from committing the act’.¹⁴

C. THE DURHAM RULE

In response to increasing criticism of the McNaghten and Irresistible-Impulse Rules, the Durham Rule was established in the District of Columbia Court by Judge David Bazelon in 1953.¹⁵ The Durham Rule originates, however, with rulings in the New Hampshire Supreme Court affirming the inadequacy of the McNaghten Rules and stating that ‘the verdict should be “not guilty by reason of insanity” if the killing was the offspring or product of mental disease in the defendant’.¹⁶ The major impetus to enact the Durham Rule has arisen from problems of expert psychiatric testimony in the courts, the claim being that the outmoded ‘moral’ terminology of McNaghten inhibited expert witnesses on mental illness for whom the terms ‘right and wrong’ and other expressions in the McNaghten Rules had no exact scientific meaning. Increasing use of psychiatric experts in cases of the insanity defense began to create acute problems of courtroom procedure, and according to an increasingly prevailing view, the ‘old-fashioned’ McNaghten Rules had the detrimental effect of unduly inhibiting and obstructing the optimal use of psychiatric evidence. Hence the emergence of Durham Rule was much heralded as a significant advance. The Rule, as stated by the Committee on Criminal Responsibility of the Bar Association of the District of Columbia in 1959, reads as follows:

The accused is not responsible for a criminal act if such act was the product of a mental disease or mental defect. A mental disease is a diseased mental condition which may get better or get worse; a mental defect is a diseased mental condition which cannot get better and cannot get worse. The criminal act was the product of the mental disease or mental defect if the act would not have occurred except for the disease or defect; and that is so whether the disease or defect was the only cause of the act, or the principal one of several causes, or one of several causes.¹⁷

The Durham Rule has not been widely adopted,¹⁸ but its history in the courts has been followed with great interest. It is probably fair to say now that the Durham Rule has largely been a failure as an attempt to ameliorate the problem of expert testimony. Briefly, three main difficulties have emerged.
The concept 'mental disease' has been shown to be notoriously unclear by skeptics - as a technical psychiatric term, it has proved to be overwhelmingly difficult, if not impossible, to give a useful, agreed-upon explication of this term. Second, the term 'product' is highly ambiguous - Judge Bazelon has called this concept the 'albatross' of Durham. Third, by phrasing the criterion in psychiatric technical or quasi-technical terms, the test has encouraged the experts to offer 'conclusory' evidence of responsibility, thus confusing the functions of the jury and the experts and obstructing rather than freeing the channels of communication between expert and jury.

D THE ALI MODEL PENAL CODE
In 1962 the American Law Institute (ALI) proposed the Model Penal Code, including a section on the Defense of Insanity containing the following test:

(1) A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality and wrongfulness of his conduct or to conform his conduct to the requirements of law.

(2) As used in this Article, the terms 'mental disease or defect' do not include an abnormality manifested only by repeated criminal or otherwise antisocial conduct.

Note that (1) is very reminiscent of McNaghten and the Irresistible-Impulse Test, except that the phraseology permits a somewhat broader interpretation of these tests, purporting to alleviate some earlier problems introduced by excessively narrow interpretations. The ALI test is gaining increasing acceptance and it is especially notable that Durham was overturned in the District of Columbia in 1972 in favor of the ALI test in the case of Brawner v. US.

3 Justification of aversive measures
The following austere sketch, risking over-simplification to achieve pellucidity and pungency, will crystallize the central moral issue of criminal responsibility. The fundamental point separating the two classes of individuals at issue in the Insanity Defense is that individual A, the criminal offender, has freely and knowingly broken the law and it is therefore 'fair' for the community to apply measures against his wishes. This is a matter of equity or justice. By contrast, an individual B's action, not done knowingly or in control, is not equitably culpable and therefore it is not 'fair' for the State to impose measures on B against his wishes. The point is that B had no opportunity to avoid breaking the law and he is therefore not morally open to aversive measures by the State. Thus any aversive measures imposed on B in the form of incarceration, compulsory treatment or whatever cannot be justified on grounds of fairness, but solely on utilitarian grounds; for instance isolation from the community might be justified on established grounds of potential danger to the community. The rights of these two classes of individuals are therefore significantly different from a moral point of view. Aversive measures may be fairly imposed on A, in accord with the severity of his crime. But this is not the case with B. B has not voluntarily given cause for abrogation of any of his rights. His interests are to be equated with those of any citizen, and his actions may be circumscribed only in so far as they might jeopardize the best interests of the community. In so far as B's actions seem likely to remain beyond his control, his conduct may need to be regulated by the State. But in such regulation the rights of the individual must be completely and fully observed. Thus, as I see it, from a moral point of view the coercive, humiliating, penal attitude and measures that are all too obviously in effect against many mental patients at the present time are outrageously morally abhorrent and completely unjustified. The criminal has sacrificed some of his rights, but the mental patient has not. His rights, I believe, also include the so-called 'right to treatment', flagrant violations of which are at present in great evidence. The point is that if we are to distinguish between A and B, as I think we can and must, and as we do under the criminal law, then the distinction must be consistently adhered to in all its implications, and not clumsily conflated and blurred with the effect that the treatment of innocent persons as criminals or 'quasi-criminals' is condoned and even encouraged.
The grounds upon which the State has a moral right to sanction measures on an individual against his wishes in the cases of A, the criminal offender, and B, the 'insane' person who is potentially dangerous, are logically and morally quite distinct. The case A comprises those persons who have voluntarily, that is, knowingly and freely, undertaken to break the law and are therefore responsible for their actions, according to mens rea. The case of B comprises that class of persons judged to be dangerous to the community, who may have acted in such a way that their action would have been criminal except that i) it is ascertained that it was not within their control to refrain; or that ii) they did not appreciate the nature of the action. Now it is a moot point just how clearly and exactly these two classes of individuals can be separated in thought, and what further means may be used for this separation, and it is also debatable, I believe, as to whether we ought to continue to pursue this distinction and use it to differentiate sharply between penal and treatment methods. But morally and logically a working distinction can be made, and such a distinction is an underlying basis of the criminal law at the present time.

4 A thought experiment
Suppose we have an individual, A, who voluntarily undertakes to harm another individual, B. Suppose also there exists a fair and impartial judicial body, J, representing the interests of the community to which A and B belong. Now it seems clearly reasonable that in so far as A has freely and voluntarily undertaken to violate a just law of J in harming B, he may be asked to take, or even coerced into taking, a course of action that deviates from his own best interest or choice of action in a degree commensurate with the degree of harm he has set out to cause, in so far as this course is estimated by J to be in the best interests of the community. We can summarize the situation by saying that A, in voluntarily transgressing a fair law, has sanctioned the abrogation of his own rights or freedom in the matter, and thus made it fair and reasonable for the community to importune him in some degree by coercing him into deferring to its collective interest.

In this abstract thought-experiment are incorporated the following assumptions.

1) The expression 'A voluntarily undertakes to harm B' is an expression to which a specifiable meaning may be attached. That is, we presuppose a concept of morality whereby certain actions are morally wrong. Also, we presuppose that an explication may be given to the sense in which human action may be said to be free, voluntary or within the control of the agent.

2) We presuppose that there is, at least in theory, such a concept as a 'fair and impartial judicial body'. It is absolutely crucial to the thought-experiment that there should exist such an intermediary between A and B. Were B to react to A's harm by individual unregulated retaliation without J, the fairness of B's act might be highly questionable on the grounds of lack of impartial examination and regulation of the various factors involved. Without the intermediary J, the language of 'revenge' or 'retaliation' may be appropriate.

3) J can fairly estimate the best interests of the community in such cases.

4) A just law has been violated, i.e., the particular law is not morally exceptionable.

5) There is a degree, corresponding to the nature of the crime, to which a path of action corresponds, that is not in the best interests of A, beyond which it is unfair to subject A and up to which it is fair to subject A, and this degree is commensurate, in some fashion that can be estimated, with the degree of harm that A set out to cause.

6) It is not fair for the community to coerce A into bending to its interests, at least on the grounds being considered here, if A's action was not voluntarily undertaken.

Each of these assumptions is a mare's nest of moral, legal, psychological and political problems. Yet for the thought-experiment to take place, one must be able to make all six assumptions provisionally, as though it were logically possible and conceivable for all of them to obtain jointly.

5 Mental disease and control
Why are persons suffering from 'mental illness' exempted from moral responsibility for their actions?
Quite clearly 'mental illness' itself is not always a reasonable and morally acceptable ground for the exemption of blame. The only plausible answer is our belief that mental illness can deprive an individual of control or appreciation of his action in such a way as to negate responsibility. What is appealed to is a principle that runs through the entire fabric of western thought, culture and morality, namely that a man is excused from blame for an otherwise blameworthy act that was committed unknowingly or that it was beyond his control to avoid. I should like to maintain that, despite the gargantuan difficulties in implementing this principle and in explicating it, it retains a basic intuitive moral reasonableness for most of us, and is an aspect of the cherished tradition of fairness that our culture has evolved. It is a corollary to our belief that a man is responsible for all consequences that he can reasonably foresee as flowing from his free actions, and only for those.

Now there is no such reasonable or accepted moral principle that asserts or implies that a man is excusable from actions that are a product of mental disease. Though a man might be suffering from 'mental illness', he might reasonably be expected to exert himself to the extent of resisting and refraining from carrying out untoward actions towards which his mental illness might steer his motivations. It is only where such resistance is impossible that we are inclined to exempt his behavior from responsibility. In other words, it is not morally sufficient for exemption from blame that the action is a product of mental illness, if this expression is understood in the sense of cause as one of a set of necessary conditions for the action. Only if 'product' is understood to mean 'sufficient condition', in the sense that the behavior in question was a necessary outcome of mental illness in such a way that the agent had no control, is it reasonable to excuse him on the ground of mental illness. Thus the crucial factor is not whether the individual simply suffered from some form of 'mental illness', but whether such considerations warrant our judgement that he lacked control over his actions at the time or whether he acted through delusion and unknowingly.

6 Retribution and justice
A variety of principles, of one form or another, have come under the heading of 'retributive'. But from this set of principles, two are especially important to distinguish.

1) If an individual freely and knowingly violates a just law then the State has an obligation to ensure that this individual is the recipient of aversive measures commensurate with his crime.

2) If an individual freely and knowingly violates a just law then the State has the right to coerce that individual to be the recipient of aversive measures up to a limit commensurate with his crime, to the degree to which such measures are estimated by fair judicial process to be in the best interests of the community, and not to a degree beyond that limit.

We might call the former the Principle of Retribution (PR) and the latter the Principle of Justice (PJ). The principles of this family have been associated with the terms 'punishment' and even with the stronger term 'revenge'. However it is clear that while PR may appropriately commend the use of such terms, PJ has nothing at all to do with revenge, except in the most accidental way, and is even sufficiently weak to be devoid of the retaliatory connotations of the unfortunate and very confusing and ambiguous term 'punishment'. PJ does not license aversive measures for their own sake or as retaliation, but only for the protection of the general interest, whether the form such measures might take is psychiatric treatment, isolation from the community, traditional penal incarceration or whatever.

Though PR has attracted numerous influential adherents historically, and indeed represents a tradition in so-called 'theories of punishment', I believe that fewer enlightened and reasonable persons would be inclined to accept it nowadays. Though PR is no doubt still very attractive to many, and even, it can be argued, represents certain elements of the criminal law at present, I believe that enlightened reform is in a direction away from PR. On the other hand PJ could be offensive only to the strong opponent of the principles of mens rea, and is free of many of the objections that could be brought against PR or even stronger principles of this type. Historically one may find a continuum of gradations between PR and PJ and even extensions stronger than PR and variants weaker than PJ. But
PR and PJ represent two poles in the debate and it is useful to focus on them in order to show that a reasonable principle of fairness or justice can be distinguished from a stronger form of retributive principle that opponents of the former are only too ready to conflate with it, so as to tar PJ with emotional connotations of 'revenge' and 'an eye for an eye'.

The entire question of criminal responsibility is a complicated skein of medical, legal and moral issues. Continued insistence in the courts, however, that the ultimate judgement of criminal responsibility must rest with the jury, and not with psychiatric or other experts, indicates that there is an underlying moral element in such decisions. Such an element is defined '... by the totality of underlying conceptions of ethics and justice shared by the community, as expressed by its jury surrogate'. PJ is best seen as a rational reconstruction of the underlying moral concept of fairness or equity, shorn of the religion-originated overtones of retribution-punishment-revenge.

7 An Aristotelian view
Acceptance or rejection of the principle P, below, is neither a psychiatric nor a purely legal question, but rather a matter of morality and public policy.

$P$: If agent $A$ is responsible for action $A$ then $A$ did $A$ knowingly and $A$ was in $A$'s control.

The psychiatrist, or at any rate the social-science expert, may be useful or even decisive in providing evidence in favor of or against $P$ in a given case. $P$ is itself, however, a meta-legal principle underlying mens rea, and cannot be accepted or rejected on purely legal or psychiatric grounds. $P$ is a moral principle. Many behavioral scientists today deny that there are moral principles, or that such principles are scientifically coherent or meaningful at all. Yet in so far as it is possible for us to have knowledge of moral principles at all, $P$ surely comes under this heading. If moral principles are logically incoherent then $P$ also is incoherent. $P$ is an ancient principle, first clearly formulated by Aristotle: 'Those things are thought involuntary which take place under compulsion or owing to ignorance; and that is compulsory of which the moving principle is outside, being a principle in which nothing is contributed by the person who is acting or is feeling the passion, e.g. if he were to be carried somewhere by a wind, or by men who had him in their power.' The concepts contained in $P$ ('responsible', 'agent', 'control', and 'know') are problematic and not well understood from a scientific point of view, although they have been much discussed by philosophers. However a plausible case can be made out for the usefulness of such concepts in areas of adjudication of moral responsibility, and our considerations on the insanity defense suggest that explications of these concepts should be even more resolutely pursued.

8 Control and the jury
The matter of who is supposed to establish what conclusion in the insanity defense between the jury and the panel of psychiatric experts is confusing and complicated. But two competing theories emerge from the spectrum of possible shades of opinion as focal points for consideration.

$T(1)$: The Jury-Control Decision: according to this view of the matter, the psychiatrists are to present evidence bearing on the matter of the defendant's control of his action at the time of the act. It is the task of the jury to decide whether this evidence is sufficient to warrant the conclusion that the defendant was not in control of his action at that time.

$T(2)$: The Jury-Responsibility Decision: according to this view, the psychiatric expert is to present evidence and a conclusion as to whether or not the defendant controlled his action. It is the function of the jury to decide whether the lack of control is sufficient to negate responsibility.

The crucial difference between $T(1)$ and $T(2)$ is as follows: according to $T(1)$ the jury rules on the issue of control, and the issue of responsibility thereby follows automatically - no control, no responsibility, but according to $T(2)$, we can separate the concepts of control and responsibility so that the implication

\[
\text{responsibility} \quad \longrightarrow \quad \text{control}
\]
does not obtain automatically. Here it may be concluded by the experts that the defendant lacked control, but the jury need not automatically conclude, in turn, that the defendant is not responsible. The additional question as to whether the degree and nature of the lack of control is adequate to negate responsibility needs to be settled.

According to some descriptions of the task of the jury, the issues of control and responsibility are separate and both need to be decided by the jury. Here we have a third alternative.

T(3): The Dual Jury-Decision: according to this view, the jury is faced with two logically independent tasks: i) deciding on whether control obtained; and ii) resolving the question of responsibility.

The Dual Jury-Decision account of the function of the jury may be found in *Eichberg v. US*:

In the first place it [the jury] measures the extent to which the defendant's mental and emotional processes and behavior controls were impaired at the time of the unlawful act. The answer to that question is elusive, but no more so than many other facts that a jury must find beyond a reasonable doubt in a criminal trial.

The second function is to evaluate that impairment in light of community standards of blameworthiness, to determine whether the defendant's impairment makes it unjust to hold him responsible. The jury's unique qualification for making that determination justifies our unusual deference to the jury's resolution of the issue of responsibility.43

Here two separate decisions are described, the extent of impairment, and whether such impairment negates responsibility, and both are regarded as jury-decisions. This point of view takes the emphasis off control as the decisive factor, allowing independent grounds for deliberation on the question of responsibility.

It is interesting to note that this very bifurcation was rejected by the ALI on the ground that the jury should not be left at a loose end on deciding responsibility, but should confine themselves to the ostensibly factual considerations of control and knowledge. We remember that the ALI test requires substantial capacity to appreciate conduct or to conform conduct to law. A minority of ALI draftsmen had suggested that this should be modified to read that a person 'is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect his capacity either to appreciate the criminality of his conduct or to conform his conduct to the requirements of law is so substantially impaired that he cannot justly be held responsible'.44 This charge was rejected on the grounds that in form at least, the inquiry should be confined to fact, and that the proposed modification might set the jury at large to evolve its own legal rules and standards without the crystallization provided by legal requirements. Thus the decisive advantage of T(1) is destroyed or eroded in T(2) and T(3).

One point requiring clarification in comparing T(1), T(2) and T(3) is whether the decision on control is simply a factual-diagnostic question, best settled by experts, or whether this decision is also partly a moral question, on which the experts should not be allowed to rule at all. This is a very difficult and important philosophical question that has not been settled, and on deliberating on it it would be a great error to be dogmatic.45 But it seems reasonable that the decision on control is best seen as both a moral and a factual-diagnostic question mixed, and that, for this reason, it is best left to a jury to decide, although the jury should be as well and clearly informed as possible, by experts, of psychiatric or other relevant scientific evidence.

Judgements on control exhibit considerable slippage in their confirmation, and it seems plain that, given the present state of the behavioral sciences, there is no very exact or highly calibrated method, no decisive data-processing procedure on control. We all have a firm intuitive grasp of the notion of control, based on our own internal insight into our own actions. We have, at least, a paradigm of control. We also have various intuitive means, of an inductive nature, of estimating in certain cases whether another person was in control of his action at a given time. But these criteria are not in themselves well understood, nor can they be applied decisively in many borderline cases. Thus in many cases there is considerable indeterminacy in deciding on the issue of control. This, of course, is reflec-
Now in criminal cases it is necessary and useful for us to make decisions on the question of control. Indeed at times we are under a moral and civic obligation to make such decisions. But how to make a reasonable and fair decision in the face of this indeterminacy? The answer is to be found in the jury system. Where such slippage exists on the question of control, and a decision must be made, that decision can be made only within the confines of the rules of evidence of the criminal law, and if any moral elements enter into the decision the underlying moral principles of the community must be reflected in them. While it is difficult to say just how moral elements enter into decisions of control, it would certainly be foolhardy to deny that moral elements do often somehow make their way into them, given that judgements of control in legal contexts appear to be based on considerations of possibly value-laden estimations of, for example, the natural and expected probable consequences of an action. The point is that if such slippage is filled in by moral considerations, as hardly seems avoidable in practice, a jury ought to make the decision, not a person who might not reflect the underlying values of the community. In a way the jury provides a kind of safeguard against the abuse of the slippage involved in control decisions.

Thus T (1) is preferable to T (2). Further, however, an important semantic distinction might clarify some of the confusion in which T (1) and T (2) are embedded. We might distinguish, after Rescher, between control and influence. Control is a yes-or-no affair, whereas influence may be said to admit of degrees. That is, influence is an intrinsically probabilistic concept, whereas control is not, although, epistemically speaking, our estimate as to whether control obtains in a given case may be probabilistic.

It is important to draw a distinction between ... control on the one hand, and what I propose- somewhat arbitrarily - to term influence upon the other. Essential in the idea of control is a condition of definiteness: the controller(s) can definitely make something happen or definitely preclude its happening. But there is also the prospect of what - by way of contrast with control - we shall call influence, viz., the capacity to make something’s happening more likely or less likely. For example, the taking of vitamin pills may render it less probable that I shall contract a common cold. The pills do not give me control-not even incomplete control - over my catching colds: the connection is merely one of influence in the specified sense (i.e., taking the pill `influences' whether or not I shall catch cold).

Thus the very definiteness of control negates responsibility. The question of whether and how we know that an agent lacked control aside; if (hypothetically) an agent lacks control, then it follows logically that he is not responsible for his action. The question of influence, and with it the very difficult notion of diminished or partial responsibility, must be distinguished from the concept of control. Now we can see T (2) in a different light. What the psychiatric experts are required to establish is evidence relevant to the questions of the degree of influence that the agent himself exerted over his action, and the degree that may be attributed to other factors, and in particular to 'mental disease'. The jury must decide, on the basis of its understanding of that evidence, whether the degree of influence of the agent over his action was so minimal that a decision of 'no control' is warranted. Indication of the nature of the strength of evidence required to warrant a verdict of 'no control' is a legal question of burden of proof on which juries are instructed according to the prevailing rule, as I understand it. Thus as the logic of the concept of control begins to emerge, the appropriateness of T (1) and the untowardnesses of T (2) become evident.

9 Functional identity of productivity and control
Judge Bazelon suggests that in practice, the difference between the Durham Rule and the ALI test is minimal. So despite the semantical differences in formulation. Whatever the merits of this claim as a prediction about jury behavior, it needs to be clearly appreciated that there is a significant semantic difference between the two rules that cannot and, I think, should not be obliterated. The ALI test makes the deciding factor control - did the defendant control his action, did he have 'substantial
capacity' to appreciate his conduct or conform it to law? It is also stipulated that such lack of control must be the result of mental disease or defect. According to the Durham Rule, the deciding factor is simply whether the action was a product of mental disease or defect - there is no requirement of control as such. Thus in the ALI test the jury's attention is focused on the central moral question of whether the defendant had an opportunity to evade the action in question, whereas in the Durham Test attention is on productivity as the central issue to be decided. One test focuses on control and the other on the factor, 'mental disease or defect' supposed to be operative in blocking control. Conversely, however, the element of control is present in Durham to the extent that productivity by 'mental disease or defect' negates control, and productivity is indeed present in the ALI test in the expression 'as a result', as noted by Judge Bazelon. The significant difference is that in the ALI test control is the central issue and productivity is merely a secondary consideration, whereas in Durham the reverse obtains. Now it seems to me that no amount of fiddling with the ambiguous expression 'product' can entirely obliterate the logical distinction between these two tests. It is conceivable that an individual's act might be a product of mental disease (in the sense of necessary condition, as ruled in Carter), and at the same time that this individual might be in control of his action and fully aware of what he is doing. An individual might have a strong wish to indulge in some bizarre act stemming from mental disease or defect, while at the same time maintaining sufficient control to refrain from translating wish into action. Such an act having been accomplished, we might accurately say of it that mental disease was a necessary component in it while not being sufficiently dominant definitely to exclude control, in the sense that the person could have refrained.

Now it seems to me that the fundamental locus of concern on deciding moral responsibility in such cases is whether the individual controlled the action, whether he had the chance to avoid doing as he did. If not, it is unfair to hold him responsible. Whether his action stemmed from 'mental disease' is a secondary question, having to do with the classification and confirmation of loss of control. In adjudicating the question of control we ask: 'How do we know control was absent?', and also: 'What was the operative factor in lack of control-external countervailing forces, lack of appreciation of the act, "internal" mental countervailing forces?'. Trying to answer these questions helps us to estimate whether control was really absent in a given case, by asking what kind of factor was operative in the control situation, so that we might look in that general direction for evidence. If the operative factor appeared to be 'mental disease' then we need to call in experts in this particular area. But the fundamental issue is not the cause of the lack of control, or the cause (necessary condition) of the action, but whether control was or was not operative at the time. Thus the question of productivity is important, but secondary, and it is crucial that it should be relegated to a secondary position in the statement of a rule for a jury.

Judge Bazelon appears to want us to believe, on the contrary, that productivity and control really amount to the same thing. They are, he writes, functionally identical. He illustrates the claim of functional identity by asking how Brawner would have fared under the ALI test: '. . . I have little doubt that the government would seek to introduce expert testimony, as it did under Durham, that Brawner committed this act not because of his personality disorder, but rather because he wanted "to get even with somebody who broke his jaw.".' 51 Thus, concludes Judge Bazelon, whatever this line of testimony might be called, it is functionally identical to the productivity question under Durham.

These tests are by no means logically equivalent, as we have seen, and any 'functional' equivalence that might be said to obtain between them could only be in virtue of independent instructions appended to them that might cause them to be interpreted as calling for the same test of responsibility. Such an equivalence might obtain where 'product' is understood in such a strong sense as always to negate control. 52 But this does not seem to be the sense that Judge Bazelon has in mind, judging by the Carter interpretation of product as sine qua non condition. 53 The same testimony might be invoked, but in one case the testimony tends to support the assumption of control, whereas in the other it may tend to support the defense's claim of productivity. As it happens, in Brawner's case the same evidence may seem to support both hypotheses in equal degree, but as we saw previously, we can easily imagine a case where this equality of degree of support does not obtain. Simply because it
appears to obtain (though even this is not easy to confirm) in this particular case does not imply that such an equivalence need apply in all cases. Even if Brawner's action was the product of an explosive personality disorder (in the sense of *sine qua non*), a jury might still reasonably conclude that his action was not beyond his control in the sense that he had the capacity to refrain. Note that describing his mental condition as 'an inability to deal with provocation' (*Brawner*, p.114) has the effect of illicitly presupposing a control criterion. Describing the mental illness as an 'inability' already forecloses the issue of control if we understand 'inability' as negating control. In this case there is indeed a functional equivalence achieved by conflating the two tests in a fashion incommensurate with the standard view of Durham-Carter. Yet on Judge Bazelon's favored interpretation, where evidently an 'inability' is not understood in such a fashion as automatically to negate control in a particular case, the testimony is functionally distinct from that under the ALI test, since a jury might find productivity but not overrule control. Thus Judge Bazelon's tendency to convey the impression that Durham and the ALI test really amount to the same test in practice is quite misleading.

### 10 Causation and productivity

The expression 'product of mental illness' is an unfortunate one in legal or moral contexts, aside from any technical meanings it might be given in various branches of the social sciences, because not only has the expression 'mental illness' provided enormous difficulties in explication, but the term 'product' ('result' or 'effect' are alternatives) is ambiguous in at least three ways - it may refer to a necessary or sufficient condition relation, or to the more usual legal concept of 'cause' whereby the cause of an event is one of a set of necessary conditions such that i) taken together with other necessary conditions a set sufficient for the event is formed, and ii) this particular necessary condition is distinguished from the remainder in the set by some factor such as voluntariness or abnormality. These are the three primary legal senses of the terms 'cause'. Now these factors, such as voluntariness and abnormality, that single out a 'cause' from other necessary conditions of an event or action appear to vary from context to context in a manner that appears to make their isolation something of a pragmatic matter. In any more or less applied area of inquiry, such as law, psychiatry and medicine, the criteria of causation are tied to isolating factors that can be changed or manipulated. In searching for the 'cause' of a disease, a medical researcher is searching for some variable that can be controlled by human beings, so that by manipulating this control parameter we can eliminate the disease.

Now unfortunately for the Product Rule, the factors relevant to criteria for causation may vary considerably from context to context - they may well be quite different in psychiatry and in the criminal law. It would therefore be unwise to assume that the expression 'product' carries a univocal meaning across interdisciplinary lines. And in fact this word 'cause' has, historically speaking, been the cause of great perplexity and semantical confusion in science and philosophy.

Unfortunately the attractiveness of the Durham Rule appears to have been partially due to equivocation on the term 'product'. We have seen that a 'mental disease' plausibly negates responsibility only where it is a sufficient condition of the negation of control of the action. If 'mental disease' is merely a necessary condition of the action, surely the presence of 'mental disease' does not by itself negate responsibility. Whether any other meaning of 'product' might be given to salvage the Durham Rule from falling to pieces on this Scylla and Charybdis is the next question. That there is an acceptable answer forthcoming, however, seems to me unlikely, since the real issue of responsibility lies in the element of control, the 'product' criterion being simply ill suited to its purported role. In interpreting the ALI Rule, therefore, it might be well to leave the element of 'result of mental disease or defect' lurking in the wings where it belongs, and not bring it back to center stage in an attempt to resurrect the Durham criterion.

### 11 The elimination of responsibility

One very influential, widespread and pervasive school of thought currently advocates the complete abolition and rejection of the concept of responsibility and allied concepts associated with *mens rea*. According to this view the criminal law should be solely concerned with rehabilitation and not at all
with punishment, the latter being a relic of the prescientific approach to penology. The 'eliminate-responsibility' adherent suggests that an enlightened legal system should be based on the 'forward-looking' aims of rehabilitation and social protection, whereas the notions of responsibility and affixing blame are essentially 'backward-looking' in their concern with elements of the mental state of the offender at the time of his criminal act. Thus mens rea is associated with punishment, a comparison to the once predominant Christian concept of sin, now morally otiose and without force in the consciousness of the general public. Lady Barbara Wootton, a forceful and articulate spokesman of this view, argues that the trend to strict liability away from mens rea in the criminal law in Great Britain is to be welcomed as a reflection of the trend away from traditional concern with responsibility.

Thus it is by no means a trivial thesis to argue that there is an absolutely crucial distinction between the individual who has freely and knowingly undertaken to inflict violence on another person, and the individual who has brought about exactly identical harmful consequences while being in the grip of a mental condition precluding his control or appreciation of his act.58 That such obvious and important moral considerations can be overlooked by the 'eliminate responsibility' proponents would be incredible were it not for the explanation that the philosophy of science and social technology that provides the conceptual background for such views is a deterministic one that consigns morality, along with the other accoutrements of 'freedom and dignity', to the realm of prescientific pseudo-knowledge. According to this view, the social sciences put a premium on the scientific, rational shaping of behavior by the intelligent manipulation of physiological-environmental variables, rather than by indulging in the curious metaphysical exercises of blaming, allocating responsibility, trying to fathom intentions and to determine the 'guilty mind', arbitrarily pronouncing actions to be right or wrong and punishing wrongdoers. All such activities, based on outmoded faculty-psychology and other obsolete views of behavior, are in turn based on the fallacy of free will. Thus it is not surprising that 'Eliminate Responsibility' theorists fail to mention the principles of fairness and morality that seem obviously to be the foundations of mens rea. for it is a systematic consequence of the determinist view of the social sciences that such principles are meaningless.

DOUGLAS WALTON, Ph.D., formerly Assistant Professor, University of Winnipeg; Canada Council Leave Fellowship, Victoria University of Wellington New Zealand

Manuscript received April 1974.

2 Vide Abraham S. Goldstein, The Insanity Defense (New Haven and London: Yale University Press 1967), pp. 23ff. I would suggest that this book is perhaps the most useful single work in gaining a general balanced appreciation of the issues raised by the insanity defense.
3 The most important fact in the current polemics regarding psychiatry and criminal responsibility is the clash of elementary philosophical perspectives. . . . Psychoanalysis . . . purports to be rigorously scientific and therefore takes a determinist position. Its view of human nature is expressed in terms of drives and dispositions which, like mechanical forces, operate in accordance with universal laws of causation. On the other hand, criminal law . . . asserts the reality of a 'significant' degree of free choice, and that is incompatible with the thesis that the conduct of normal adults is merely a manifestation of imperious psychological necessity. [Jerome Hall, General Principles of Criminal Law (Indianapolis and New York: Bobbs-Merrill 1960), p. 455.]
6 The McNaghten Rules (10 Clark and Fennelly 200, 8 Eng. Rep. 718, 1843) are quoted in full in Biggs, op. cit., and in Sidney Gendin, 'Insanity and Criminal Responsibility', American Philosophical Quarterly, x: 2 (1973), pp. 99 ff. I have not quoted them in full here to save space, but the serious student should take the trouble to read them in their entirety, as this will in itself eliminate
some of the grosser misinterpretations that have become popular, to a great extent, I suspect, through their piecemeal quotation.

7 Vide Goldstein, op. cit., ch. 4.
8 Gendin, op. cit., p. 103: 'The quite mistaken interpretation that the rules provide both a necessary and sufficient condition for criminal responsibility is, I suspect, at the bottom of much of the traditional criticism that the Rules are "too narrow".
9 Vide Biggs, op. cit., ch. iv.
10 Brakel and Rock, op. cit., p. 380.
11 The controversy among psychiatrists on this point is outlined in Hall, op. cit., pp. 486 ff.
11a It is significant that the need for a ruling on control to supplement the McNaghten Rules was quite clearly recognized and that the issue was quite succinctly stated shortly after the publication of the McNaghten Rules. Hall cites two interesting quotations from commentators of the time.

1) Knaggs, Responsibility in Criminal Lunacy (1854), p. 69: '... this should be the test of irresponsibility - not whether the individual be conscious of right or wrong - not whether he had a knowledge of the consequences of his act - but whether he can properly control his action!

2) Maudsley, Responsibility in Mental Disease (1874), p. 133 et seq.: 'In the face of this example [after a case described by Maudsley] of uncontrollable morbid impulse, with clear intellect and keen moral sense, what becomes of the legal criterion of responsibility?'


12 Vide Goldstein, op. cit., ch.5, entitled 'The Misnamed 'Irresistible Impulse' Rule'.


21 ALI Model Penal Code, Proposed Official Draft 4 May (1962), Sec. 4.01, p. 66 (American Law Institute, Philadelphia).
22 The Canadian Royal Commission on the Insanity Defense (1956) adopts a broad interpretation of the McNaghten Rules as its recommendation, arguing that the Durham, ALI and Irresistible-Impulse Tests have greater weaknesses that would impair the uniformity of justice. According to the Commission, '... if an accused person has mental capacity to foresee and measure the consequences of the act he committed he should be held criminally responsible, unless by reason of disease of the mind he did not know that the act was morally wrong in the sense that it was something that would be condemned in the eyes of his fellow men' (Article 4, p. 33).
23 Braunner v. US (DC Cir. 1972).
24 For a fully developed view along similar lines, see H. L. Packer, The Limits of the Criminal Sanction (Stanford: Stanford University Press 1968).
26 Readers familiar with the philosophical literature on punishment will recognize the somewhat boldly presented view proposed herein as one purporting to resolve the utilitarian-retributive antinomy.
29 Since our criminal law is based on an assumption of free will, we should not punish men for what they cannot avoid. As an abstract proposition, this is unquestioned. 'All the several pleas and excuses,' said Blackstone, 'which protect the committer of a criminal act from the punishment which is otherwise annexed thereto, may be reduced to this single consideration, the want or defect of will' (Manfred S. Guttmacher and Henry Weihofen, Psychiatry and The Law [New York: Norton 1952], pp. 408ff.).
30 According to this view 'punishment' is not to be inflicted for its own sake, but only for the public good. Vide William Knaile, The Responsibility of Criminals (Oxford: Oxford University Press 1967), pp. 17ff.
32 I should like to acknowledge my general indebtedness to the writings of Professor H. L. A. Hart, who has probably contributed more profoundly in recent times than any other single person to our understanding of the concept of criminal responsibility.
It ought to be more widely recognized how extremely difficult it is to exclude one's values from the construction of his theory of penal law. . .” Jerome Hall, *General Principles of Criminal Law* (Indianapolis and New York: Bobbs-Merrill 1960), p. 2.


It ought to be more widely recognized how extremely difficult it is to exclude one's values from the construction of his theory of penal law. . .” Jerome Hall, *General Principles of Criminal Law* (Indianapolis and New York: Bobbs-Merrill 1960), p. 2.

Brawner, p. 23.

Brawner, p. 35, also pp. 102ff.


See Myles Brand (ed.), *The Nature of Human Action* (Glenview, Ill.: Scott Foresman 1970).

A very clear statement of a 'degree-of-control (influence) - sufficient-to-negate-responsibility' type of test is proposed by Wechsler, the operative criterion being 'whether the capacity of the accused to control his conduct in accordance with the law was impaired so greatly that he cannot justly be held criminally responsible' (Herbert Wescher, 'The Criteria of Criminal Responsibility', *University of Chicago Law Review*, xxii [1955], p. 372). According to this formulation the jury may hold impairment less than total to suffice, unlike T(1), which appears to demand a 'complete' or 'total' lack of control. Thus we might allow a variant of T(1)* to allow for the possibility that responsibility may terminate when control diminishes to a certain degree, say k, a critical point to be decided by the jury. Thus according to T(1)*, degree of loss of control (or influence) of behavior may be greater than the critical point, k, where responsibility is negated.

T(1)*: If x controls A up to a certain degree, k, then x is responsible for A, where the decision concerning the establishing of degree k in a particular case, as well as the decision of whether that degree is met, is up to the jury.

In other words, if we call the point at which control (totally) terminates t, then according to T(1), it is always the case that k = t, whereas according to T(1)* it may be the case that k ≠ t.


Brawner, pp. 128ff.


Vide Goldstein, op. cit., ch. 8.

Brawner, p. 114.

Ibid.

There is considerable evidence that in the old New Hampshire rulings, from which the Durham Rule was drawn, the term 'product' was interpreted in the very strong sense of implying loss of control of the agent when the action is a product of mental disease. Hence there is an important difference between these rulings and the Durham-Carter Rule. Judge Doe is quoted in Herbert Wechsler, 'The Criteria of Criminal Responsibility', *University of Chicago Law Review*, xxii (1955), pp. 369ff, as uttering an interpretation surprisingly similar in expression to Aristotle's statement (see *Nicomachean Ethics*, iii, ch. 1, 1110a): ‘When the disease is the propelling, uncontrollable power, the man is as innocent as the weapon - the mental and moral elements are as guiltless as the material’ (*State v. Pike*, 49 NH 399, 441 [1870]). Judge Ladd's equation of product and control is even more explicit: 'If the defendant had an insane impulse to kill his wife, which he could not control, then mental disease produced the act. If he could have controlled it, then his will must have assented to the act, and it was not caused by disuse, but by the concurrence of his will, and was therefore crime' (*State v. Jones*, 50 NH pp. 369, 399 [1871]).

Weschler's discussion of productivity is the most clear-headed outline of the conceptual alternatives on this issue that I am familiar with. He also traces various interpretations of the Product Rule through a history of court rulings. Weschler notably concludes: ‘... the logic of the situation points ... to the development of causal concepts less relaxed than but-for cause but less demanding than complete exclusion of volition... That this necessity is not acknowledged in the court's opinion and that no attempt is made at the elucidation of the causal standard for the jury seems to me quite indefensible’ (Herbert Wescher, 'The Criteria of Criminal Responsibility', *University of Chicago Law Review*, XXII [1955], p. 371).

Ability (such as the ability to tie a knot or to play the violin) may be distinguished from control of an individual action. If I perform an action, they define I control that action, whereas I need not have demonstrated an ability to do that type of action generally, as it might have been a fluke.

What we regard as the cause of any phenomenon depends on our viewpoint or the purpose we have in mind. From the viewpoint of the ordinary medical practitioner, the cause of any pathological manifestation is that which he must deal with if he is to cure his patient. To a medical research scientist, however, the cause is that which must be dealt with if the disease is to be eradicated, or others prevented from succumbing to it. To a penologist or criminologist, the cause is likely to be something
wholly different, some social phenomenon that his discipline is competent to work on. [Henry Weihofen, 'The Flowering of New Hampshire'. *University of Chicago Law Review*, XXII (1955), p. 360].

57 Barbara Wootton, *Crime*, p. 51.