

R. B. BRANDT

THE INSANITY DEFENSE AND THE THEORY OF MOTIVATION

The rule which makes insanity a defense against criminal culpability obviously needs some justification beyond just the intuitive belief that insane persons are to be excused, or that it is pointless and unjust to punish them in view of the (supposed) fact that they are unable to obey the law. The rule needs to be consistent with and supported by what we know of psychology, and with a general philosophical theory of the system of the criminal justice. What follows is an attempt in this direction.

What might a philosophical justification of the system of criminal law be like? There have been recent claims that a convicted criminal should be punished, as payment of a debt to society, or to make up for the unfair advantage he has taken over others who are lawabiding. But a widely accepted view is that the criminal law — and indeed every human institution — is justified if it can function so that it optimizes the general good. This view could be questioned, and if it is we could go on to defend it by examining the story on a deeper level, arguing that rational and informed persons would support, or want, for a society in which they expect to live a lifetime, only institutions which are benefit-maximizing.¹ (This view is not very different from that of Rawls² if we make some additions to his concept of rationality and drop his proposed strategy of maximin reasoning, along lines suggested by John Harsanyi.)³

¹ See R. B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979) chs. 10 and 11. What is there said about a social moral system can be transposed to hold of a legal institution.

² John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971) chs. 1–3.

³ John Harsanyi has argued that Rawls' basic conception leads to the idea that institutions including morality should be devised so as to maximize average utility. 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls'

If we take the benefit-maximizing line, we can go along with a thesis described by Plato:

In punishing wrongdoers, no one concentrates on the fact that a man has done wrong in the past, or punishing him on that account, unless taking vengeance like a beast. No, punishment is not inflicted by a rational man for the sake of the crime that has been committed — after all one cannot undo the past — but for the sake of the future, to prevent either the same man or, by the spectacle of his punishment, someone else, from doing wrong again.... Punishment is inflicted as a deterrent.⁴

If we are receptive to this general view about when institutions are acceptable, we hold that the aim of the legal system should be to maximize the general well-being, taking all the costs into account. More specifically, what the *criminal* law should aim to do is to maximize the general well-being, and to minimize the damage of crime and anxiety about the possibility of crime, at least human cost — that is, in part, without the infliction of pointless injury or suffering on anyone. It is widely, and rightly, supposed that this objective is best achieved by the public enunciation of general prohibitions (primarily against injury to others or their property), accompanied by threats of punishment for those who fail to conform, threats that are mostly made good. But since the liberty and welfare of the nonconformers are themselves goods which have an equal claim to respect, the system must be devised so that their rights to liberty and welfare are overridden only to the extent necessary for the general welfare. It is partly in order to protect these rights of the nonconformers that the system of criminal justice recognizes a set of justifications and excuses which protect the nonconformers when punishing them would serve no important social purpose. So much for general philosophical background.

The general idea, then, is that an optimal system of criminal justice will provide protection of the general welfare by a system of threats aimed at deterring the convicted from repeating, and deterring others

Theory', *American Political Science Review* 69, 37–63; 'Morality and the Theory of Rational Behavior', *Social Research* 44, 631–6; and 'Basic Moral Decisions and Alternative Concepts of Rationality', *Social Theory and Practice* 9, 231–44.

⁴ Put in the mouth of Protagoras, in *Protagoras* 324. But see *Republic* 614 ff. and *Laws* 862 ff.

from emulating them. But what is deterrence? Presumably, it is the impact on the motivation of possible offenders so that they are more inclined to obey the law than they otherwise would have been.

1. THE THEORY OF MOTIVATION

I now turn to the theory of human motivation. If this theory is accepted, the psychological assumptions on which the insanity defense is based must be substantially revised, but not necessarily the main thrust of the insanity rule itself, although I think that it too may be improved in certain details.

But, first let me note a general assumption of psychological work on the theory of motivation (and all other branches of psychology): that human behavior is an instance of causal laws.⁵ As the Statement of the American Psychiatric Association on the Insanity Defense says: "Psychiatry is a deterministic discipline that views all human behavior as, to a good extent, 'caused'".⁶ Various legal writers would reject this assumption, regarding it as inconsistent with any defensible theory of criminal justice, some writers saying that if everything a person does is caused then illness and badness are merged and the law should excuse everything. But there is, as we shall see, no reason to think that the law must excuse everything if causal determinism is accepted. And, suppose we adopted the alternative view that human behavior is not caused; then must we not regard it as random? Would this view make possible a defensible view of responsibility? Indeed, the system of criminal justice, in holding that legal sanctions may improve the criminal and deter him and others, seems to suppose that human behavior is caused. A few philosophers⁷ think there is a viable middle position, but I propose to ignore them.

⁵ Anyone who is doubtful whether anything is really known about "laws" of motivation or the brain-state background should consult a very recent review, Douglas C. Mook, *Motivation: The Organization of Action* (New York: W. W. Norton and Company, 1987) chs. 3, 9, and 10.

⁶ December 1982, p. 14.

⁷ R. M. Chisholm, 'Freedom of Action', in K. Lehrer (ed.), *Freedom and Determinism* (New York: Random House, 1966), 28-44; and 'The Agent as Cause', in *Action*

Some writers who reject the view that human behavior is caused seem to take this view because they think a person can be justly punished only if and to the degree that his action was *morally blameworthy*. How, they ask, can a person be morally blameworthy for an action if its occurrence was written in the stars? But what does the term “morally blameworthy” mean? Need we think that the applicability of this term presupposes that behavior is uncaused? I suggest that a useful interpretation of “the act was morally blameworthy” is “the agent would not have committed that act but for an unacceptable level of character or moral/legal motivation”.⁸ (This is like saying that dropping a fly ball is an “error” only if so doing shows a substandard level of skill.) This proposal does not tell us how we are to define “unacceptable level of character or moral/legal motivation”, and I shall come to that. But it is clear that, given this explanation, an act can be morally blameworthy independent of its having been caused — and irrespective of how an agent got his character and motivation. An act can be morally blameworthy on this conception if it manifests an unacceptably substandard motivation *now*, even if its agent is as he is because of his genes, faulty upbringing, etc. One might ask: why define “morally blameworthy” in this way? One answer is that such a definition has implications acceptable to our intuitions. Another is that the definition makes it sensible for the agents of blameworthy acts to be negatively reinforced by punishment or social disapproval, thereby profitably improving the level of their motivation, which “blameworthy” implies is defective and needs improvement.⁹

Theory, eds. M. Brand and D. Walton (Dordrecht: Reidel, 1976), 199–211. Also Richard Taylor, *Action and Purpose* (Englewood Cliffs, N.J.: Prentice-Hall, 1966). These theorists can argue that the criminal law does not *cause* any modification for behavior, but simply makes clear to the prospective criminal that his *options* for choice are narrower or more unpleasant, than he might otherwise have thought.

⁸ Cf. R. B. Brandt ‘Blameworthiness and Obligation’ in *Essays in Moral Philosophy* ed. A. I. Melden (Seattle: University of Washington Press, 1958); and R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J.: Prentice-Hall, 1959) chap. 18; and ‘Traits of Character: A Conceptual Analysis’, *American Philosophical Quarterly* 7, 23–37.

⁹ It is worth noting that the frequent affirmation of the importance of “free will” in the literature on punishment in general and the insanity defense in particular may

With these background remarks behind us, let us consider what the theory of motivation affirms about human behavior.

In the first place, motivation theory, like the law, is concerned only with *acts*, bodily movements which are *voluntary* in a sense which excludes movements which are reflex, i.e., that occur during unconsciousness or sleep and during hypnosis. This agrees with the view of the Model Penal Code (hereafter MPC), which says that a culpable action must be a “product of the effort or determination of the actor, either conscious or habitual”.¹⁰

About acts of this type the theory of motivation puts forward some “laws”. These laws could stand more evidence, but I think they are the best we have at the present time.¹¹ The empirical evidence for the laws comes from such work as that on achievement motivation, experiments in industrial psychology, etc.¹² But the laws are also just an empirically supported elaboration of the common-sense view that, when making a decision, we line up the pros and cons for each action (the outcomes), estimate the probability of each, and ask ourselves how much we want each. Then we try to sum.

What, as I think, the currently best-attested psychological (non-neurological) theory asserts is that every voluntary action is a function of the following variables: (1) what the agent manages to think of (at the time of decision) as options for choice; (2) the agent’s beliefs, at the time of choice, about the possible consequences of these various

have been misunderstood by some; it may well be that at least many writers have no intention of using “free will” to imply lack of causal law. They may think that all free will really requires is that people make choices in the normal “rational” way: noting their options for action, their likely consequences, the desirability of these, and choosing accordingly. This conception does not imply lack of causation.

¹⁰ 1.13 (2) and 2.01 (2).

¹¹ Persons who have put forward these laws often, I think, do not assume that they are basic laws, but speculate that they are ultimately explainable by the physiology of the brain, in other words by principles of physics and chemistry, given the way the brain is structured. We can leave this open.

¹² For a review of some of the evidence with citations of data see R. B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979) chap. 3. For a much fuller account, see Mook, *op. cit.*, chs. 9 and 10.

options (counting the kind of act as itself a consequence), as well as how likely they are, given a particular act; (3) how *strongly* the agent wants (possibly unconsciously), or is averse to, the act and its expected consequences, at the time of choice; and (4) *how vividly* the agent represents these outcomes and their relationship to the act — what a person in some sense believes but is not salient in his thought may not influence his choice.

We might put this in other words as follows: the *strength* of an agent's *tendency* to choose one of the options open to him is a sum, of the intensity of his desire/aversion toward *each* anticipated *consequence* of that option, reduced by the believed improbability of the consequence occurring if that option is taken, and reduced again by any lack of salience of that consequence and its relation to taking that option in the awareness of the agent — summing over all the anticipated consequences of taking that option. The theory affirms that the choice the agent actually makes is for the option he has the strongest *tendency* to take — for which this sum is greatest. (One may wonder how in the world numbers are to be assigned to these, so that it is sensible to talk of the largest “sum”. I am going to assume that there is an intuitive understanding or at least an ordering of how much we want certain outcomes, how probable we think they are, and more dimly how nearly the representation of each approaches the vividness, say, of sensory perception). This “law” is only approximate and needs to be filled out in many details. For obviously the ordinary person has only vague notions of probability, has little idea of what it is to want something more or less strongly or how to form a product of this strength of want and probability, much less how to go about summing these products. There is a large literature about the contortions people go through when they are trying to decide which car to buy, or even which sandwich to purchase — within the framework of the sketched conception of laws applicable to human behavior.¹³

¹³ Experimental data show that individuals do not very reliably estimate the probabilities of outcomes, given their evidence. Moreover, we are not equipped with any summing device which enables us to know which sum of products comes out highest. There are all sorts of strategies individuals use to solve this problem, and all

A good many philosophers who do not question the above psychological causal account want this causal account filled out with a good many more details. What these details are is controversial. But something like the following is probably as near as we can get to an agreed view. First, a person's thought of a possible course of action being most strongly wanted, along the lines I have sketched, will result in the *formation of an intention* or disposition to follow a staged *plan* of action. (Some philosophers debate what an "intention" is, some holding it is just an everything-considered preference for or pro-attitude toward an action plan,¹⁴ others insisting that it is more than this, partly because intentions persist and mould future plans.¹⁵) This intended plan of action will presumably fix when it should be executed, and when that time is (believed to be) *now* the agent will (supported by the intention but also by the underlying desire for the outcome) begin to execute the plan by trying, or willing, to bring about the first stage of the plan, usually a certain bodily action — a "willing" the nature of which is itself debated (sometimes thought of as the focusing of attention on the projected bodily action, sometimes as involving an image of the sensation characteristic of the intended movement). This action will result in the setting of an appropriate next stage of the plan, or even in the wanted outcome. (If one wants more light in the room, the act of

sorts of proposals as to what these strategies are. See Mook, *op cit.*, and various review articles in the *Annual Review of Psychology*: J. R. Bettman, 'Consumer Psychology', 37, 257–89; G. F. Pitz and N. J. Sachs, 'Judgment and Decision' 35, 139–63; H. J. Einhorn and R. M. Hogarth, 'Behavioral Decision Theory: Processes of Judgment and Choice' 32, 53–88, especially 69–77; and P. Slovic, B. Fischhoff, and S. Lichtenstein, 'Behavior Decision Theory' 28, 1–39. Also see D. Kahneman and A. Tversky, 'Prospect Theory: An Analysis of Decision under Risk' *Econometrica* 47, 263–91.

¹⁴ Donald Davidson, *Essays on Action and Events* (Oxford: Clarendon Press, 1980) pp. 98–100. For more clarification of this view, see the following footnote on his reply to Bratman and pp. 220 ff.

¹⁵ See Michael Bratman, 'Davidson's Theory of Intention', in *Essays on Davidson: Actions and Events*, eds. B. Vermazen and M. Hintikka (Oxford: Clarendon Press, 1985). Also his forthcoming (Harvard) *Intentions, Plans and Practical Reasons*; and his 'Taking Plans Seriously', *Social Theory and Practice* 9, 271–87. Davidson replies lucidly in eds. B. Vermazen and M. Hintikka, *ibid.*

pulling the drapes aside will be enough.) The intention will, unless there is a change of mind, remain through the period necessary for reaching the desired outcome, monitoring the sequence of actions in view of the feedback resulting from earlier members of the sequence, and other information.¹⁶

We should note that all of this is consistent with holding that, directly or indirectly, acts are caused by a complex of the agent's beliefs and desires/aversions.

Among the desires/aversions which control behavior are some of especial interest to the law: (1) empathic/sympathetic concern for others, including an aversion to injuring others and to disasters like a nuclear explosion, and a desire to give assistance to others in distress; (2) aversions to theft, lies, rape, etc., learned in various ways, but probably mostly because of its having been made clear to an agent that they are normally harmful to others;¹⁷ (3) a desire to pay attention to laws, because they are intended for the welfare of all and enacted by a democratically elected body; (4) an aversion to doing what is considered morally wrong; (5) finally, motivation not to perform actions

¹⁶ See, for example, Bruce Aune, *Reason and Action* (Dordrecht: Reidel, 1977) chaps. 1 and 2, pp. 137–42; A. I. Goldman, *loc. cit.*, and 'The Volitional Theory Revisited' in *Action Theory*, eds. M. Brand and D. Walton, (Dordrecht: Reidel, 1976); Hugh McCann, 'Volition and Basic Action', *Philosophical Review* 83, 451–73; L. H. Davis, *Theory of Action* (Englewood Cliffs, N.J.: Prentice-Hall, 1979), 38 ff., 59–93; John Searle, *Intentionality* (Cambridge: Cambridge University Press, 1983), pp. 83–135; Wayne Davis, 'A Causal Theory of Intending', *American Philosophical Quarterly* 21, 43–54; G. A. Miller, E. Galanter and K. H. Pribram, *Plans and the Structure of Behavior* (New York: Holt, Rinehart and Winston, 1960), especially Ch. 4; William James, *The Principles of Psychology* (New York: Henry Holt and Company, 1913), II, 487–92; A. G. Greenwald, 'Sensory Feedback Mechanisms in Performance Control: With Special Reference to the Ideo-Motor Mechanism', *Psychological Review* 77, 73–101.

¹⁷ See Martin Hoffman, 'Developmental Synthesis of Affect and Cognition and Its Implications for Altruistic Motivations', *Developmental Psychology* 11, 607–22; 'Moral Development' in *Carmichael's Manual of Child Psychology*, ed. P. Mussen (New York: John Wiley and Sons, 1970), II, pp. 261–359, part on "the inductive method"; and 'The Contribution of Empathy to Justice and Moral Development', in *Empathy: A Developmental Perspective*, eds. N. Eisenberg and J. Strayer (New York: Cambridge University Press, 1987).

for which the law provides sanctions — to do so is risking unpleasant consequences. I shall call these “moral/legal motivations”.

Now, if the above model of motivation theory is correct, these motives have to compete with various other desires: for food, water, sex, human company, compliments, financial security. Moreover, there are emotion-based desires. People desire to act aggressively when angry, to escape when fearful, to restore a relationship when grieving. When emotions occur they markedly change the structure of motivations in the person at the time.

There is one feature of what I am calling “moral/legal motivations” which distinguishes them from many of the desires/aversions with which they compete for control of conduct. This is that they may be called “standing”; they are relatively fixed and permanent unlike desires for food or sex. (They are not the only desires with this status; a person’s motivation to achieve is also relatively unchanging.) An empathic person does not suddenly cease to be so on the following day. The same for aversion to injury of others, theft, lies, rape, etc. And the same for the aversion to being the target of the law’s criminal sanctions. Of course, these standing motives may not always control conduct: each has a certain level of strength, and in many situations this will not be enough to control behavior in the face of desires stronger at that time.

II. WHAT CAN THE SYSTEM OF CRIMINAL JUSTICE DO?

Let us now address ourselves to the question: What can the system of criminal justice do to insure that people conform to the law?

First, I have listed five kinds of desires/aversions which are forces inclining a person to conform his conduct to the law. Can the system of legal penalties strengthen any of these? Of course it can strengthen the fifth (motivation to avoid the penalties of the law) by making the penalty heavier or more certain; and it can add to the strength of the first and second (aversions to injuring and to specific offenses like theft), by conditioning, by negative reinforcement of the agent or vicariously by his observation of what happens to others. The law can also bring about, by conditioning, that we *notice* more options for

action, and attain a more *vivid awareness* of the consequences. (A hefty fine for running a red light tends to make one think that one had better begin to slow down when the yellow light appears. Lengthy incarceration may not have this effect.)

A different thing that might be done is to reduce the strength of desires/aversions which lead to breach of law: whether to get money, satisfy sexual desires, obtain a position of status in the community, and so on. Since such motives are formed mostly by the example of parents or other prestige figures, by the educational system or by fare provided by the mass media such as television, there is not much the law can do about them. What the legal system can mostly do is increase competing aversions, to theft, etc., in the way already suggested.

Now if the purpose of the system of criminal justice is to enhance general well-being by control of crime at least cost, we can see why it should and mostly does refrain from punishing various types of infringement of the statutes. For sometimes it is *better* that people *do* infringe the *statutes*, e.g., by actions to promote public goods, such as burning down a house when necessary in order to prevent a general conflagration. We *want* behavior of this sort. We call this sort of infringement of the statutes “justified” infringement of statutes — “lawful” in the broad sense. Violators of this sort are not held guilty of unlawful action. But there is a wider group of cases in which the law should not and generally does not punish (what we may call “excused” unlawful behavior), e.g., when the act would have been lawful if the situation had been what the agent thought or at least reasonably thought it was; when there was ignorance of the law (perhaps ignorance from which a reasonable man would suffer); action under “duress”; actions done as a result of involuntary intoxication by alcohol or drugs (perhaps because of addiction); “entrapment”; and (as mitigating) provocation. (Some of these “excuses” can be accommodated in the *definitions* of various offenses, requiring that they be done willfully, purposefully, recklessly, or negligently.)¹⁸

¹⁸ The MPC appears to compound confusion by putting these mental conditions into the *definition* of the crime. As Glanville Williams has pointed out, this move makes for difficulties in deciding the criminality of accessories. See his ‘The Theory of Excuses’, *Criminal Law Review* 1982, 734 ff.

Why not punish in these cases? In the case of justifications, the answer is obvious: we *want* people generally to do what the agents did, in their circumstances. The answer in the case of excused behavior is different. We must recall that the primary effect of punishment on the convicted is to affect motivations (and vividness of representations). But in the case of “excused” behavior punishment may not optimize the good, for change of motivations is not needed since as far as we know the agent’s motivations are already at an acceptable level. A person who acts illegally because of an honest mistake is not shown thereby to have a defect of motivation. There is no point in trying to improve the agent’s motivation, if he is, as far as we know, already where he ought to be. And as for the effects on others, there is not evidence that failure to punish in “excused” cases decreases deterrence, that is, the motivation of others not to commit crime because of their knowledge of punishment in the nonexcused cases. Much less would punishment, or the threat of it, affect the motivation of those to whose behavior such excuses would apply, since the threat of punishment can hardly operate to deter a person who commits an offense because he is ignorant of facts which make his act illegal. The system of excuses avoids inflicting pointless harm on persons who have not conformed with the law.

Some may be disturbed by the implication that a main aim of the criminal system should be to affect *motivations* in one way or another. It is true that the effect of the law may be somewhat educational cognitively: by emphasizing the standards of conduct as well as the probable consequences of a certain kind of action. But the main impact of the criminal system must surely be to affect motivation. This may seem to conflict with the idea that the criminal law does not concern itself with motivation. But this conception that the law is not concerned with motivation seems to ignore the fact that there are *two* kinds of motivation, the motives which *lead one* to break the law, and also motives (mostly *aversions*) about breaking the law, stealing, killing, etc. These latter motives surely *are* of interest to the law. Suppose we say Jane is guilty of a crime because she tampered with the brakes of her husband’s car, knowing it would bring about his death. The legal scholar may say Jane has *mens rea* because she *knowingly* acted to bring

about his death. But when we take motivation theory (and common sense) into account, it is clear that what is responsible for Jane's tampering with the brakes and thereby producing the death of her husband was not merely her wanting the money etc., and knowing that tampering with the brakes would bring about the things she wanted; what is also responsible is her relative *indifference* to bringing about her husband's death. So, but for that indifference, her action would not have occurred. Thus, while the MPC makes the normal condition of criminality that the agent do something forbidden, either having the act or its consequences as his "conscious object", or knowingly (= willfully)¹⁹ performing the act with its anticipated consequences, or recklessly (with awareness of the risk) doing a forbidden thing, it could as well say that a condition for conviction of a crime is *failure to be motivated to avoid* a foreseen forbidden consequence, or to be indifferent to a substantial risk that it occur. Thus talk about intention or foresight (which some have thought identical with *mens rea*) is misleading; it is true that an intentional act is in part a cognitive occurrence — there must be a plan, a cognitive map for future action, however simple — but it is an essential part of an intentional action that there should be *motivation* to get the end envisaged by the plan (or to avoid it), and what a person foresees but ignores is evidence about what he is indifferent to.²⁰

In view of the foregoing, it seems the law *could* say that a prohibited action, unjustified, is *culpable* only if it manifests (has as a *necessary* condition, in the circumstances) a defective level of *moral/legal motivation*. (The defective motivation is a "necessary condition" of an action if and

¹⁹ MPC 2.028.

²⁰ There are parts of the MPC which take motivation in my broad sense (both desires and aversions) into account: in its statement about the excuse of duress, about the effect of renunciation of criminal purpose when there has been an attempt, about "extreme indifference to the value of human life" (210.2 (1b)), or about the aggravating effect of the purpose of a murder being "pecuniary gain", or the act's "manifesting exceptional depravity" (210.6 (3) g, h)).

only if the action would not have occurred but for the motivation.)²¹ This is a conclusion of considerable importance.

It might be objected that this proposal cannot be right, for it supposes a person can be convicted of a crime only when there is known to be a defect of character/motivation, when in many cases all we know about a person's mental state is this single action — and how then could we be justified in saying that the condition of defective motivation is met? (One might equally well ask whether the action was purposeful or done knowingly.) But this objection overlooks how we can justifiably make judgments about motivation on the strength of a single action. If a person is playing a friendly game of tennis, and if in the middle of it a child has a bad spill from his bicycle in the next court, and is screaming and covered with blood, and the person calmly continues with his tennis, we *do* draw an inference about his motivation, and it is a well-founded inference if we do, at least absent some explanation.²² Not that this inference comes out of the blue: we know how sympathetic or emphatic people normally behave in comparable circumstances. This man can hardly be either.

So the commission of an illegal unjustified act can be viewed as prima facie evidence (which is rebuttable by further explanation) that there is a defective level of character or moral/legal motivation. *What a legal "excuse" does* is *not* provide evidence that the agent's character/motivations are *perfect*, but merely shows that in the circumstances it is *unwarranted to infer* defective motivation from the particular unlawful act. And if defective motivation is a condition of culpability then an excuse is a shield against culpability.

This is as it should be since, as remarked above, one of the purposes

²¹ What I mean by saying that a certain motivational defect is a necessary condition of an event (illegal act) in the circumstances, is that there is a general natural law to the effect that whenever an act of the type in question occurs in circumstances like the present, there is always a defect of legal/moral motivation of some sort. This is to say that when the act occurs, the defect must have been present, or the act would not have occurred but for the defect.

²² See R. B. Brandt, 'Traits of Character: a Conceptual Analysis', *American Philosophical Quarterly* 7, 26.

of the criminal law is to improve in the agent (and others) any defective level of motivation relevant to lawabidingness, and if there is no ground in the agent's action for thinking his motivation defective, the punishment is uncalled for, serves no purpose — even that of deterrence of others.

III. THE INSANITY DEFENSE

The thesis I suggest about the insanity defense is that it is essentially the claim that the state of mind of the agent at the time of his unlawful act prevents conclusive inference from his act to a defective level of moral/legal motivation (as being its necessary condition), hence it provides a release from culpability. Some of the concepts in this proposal need explanation, which I propose to offer. This proposal differs from the widespread view that the justification for excusing acts done in an insane frame of mind is basically that insanity involves an *incapacity* to obey the law, either cognitive (can't know what the law requires) or volitional (knows what the law requires, but unable to conform conduct to it). And it is thought to be senseless and unjust to punish in face of this incapacity: senseless because penal sanctions could not reach him (or others in his situation by the threat of them), and about the kind of incapacity there is to obey the law. Hence we shall see that there is reason to be unhappy with the legal rule about have a "fair chance" to obey the law.) We shall see, however, that there is obscurity about both the sense in which there is a cognitive defect, and about the kind of incapacity there is to obey the law. Hence we shall see that there is reason to be unhappy with the legal rule about the insanity defense as it now stands, and with this standard rationale.

Let us look both into this obscurity and at the same time into problems of the law about insanity as it now stands.

Let us look at the statement in MPC, the most widely accepted rule at present. It says²³ that a person "is not responsible for unlawful conduct if at the time... as a *result of mental disease or defect* he lacks sub-

²³ 4.01 (1).

stantial capacity either to *appreciate* the *criminality* [wrongfulness] of his conduct or to conform his conduct to the requirements of law" [my italics]. Further, "mental disease or defect" does not include any abnormality manifested only by repeated criminal or otherwise anti-social conduct.²⁴ The statement of the law does not say anything about a rationale.

There are some things about the MPC rule which are not satisfactory. First, a minor point. We might ask what is a "disease" or "defect".²⁵ Is this to include a mild neurosis, or any one of the problems listed in the Diagnostic and Statistical Manual (of the Psychiatric Association) edition III, or an I.Q. below 80? The implication of the MPC is that *any* level passes if it is serious enough to cause a "substantial inability" to identify illegal conduct or to avoid conduct which is illegal.²⁶ Very well. "Substantial inability" is rather vague, but this is probably not a defect since the concept probably could not be spelled out in detail, and there are some things which can be left to the good sense of the jury and judge.

But, and more serious in its vagueness, what is it to "appreciate" the immorality/criminality of one's act? Evidently something more than just readiness to *say* that an act is wrong. What else more? Some writers talk about the "emotional meaning" of the act (or of "immorality" or "illegality"). Does this mean ability to represent the whole situation (and the wrongness of the act?) to one's self *vividly* (a factor which we saw above is one of those of which action is a function), with corresponding emotional repercussions? Or is it awareness of generally accepted social standards for behavior of the kind in question? Or is it to have at least *some reasonably strong aversion* to perform-

²⁴ 4.01 (2).

²⁵ Thomas Maeder offers a definition in *Crime and Madness* (New York: Harper and Row, 1985), p. 85.

²⁶ The American Psychiatric Association, in its 1982 report on 'The Insanity Defense', proposed that exculpating mental disorders "must be *serious*. Such disorders should usually be of the severity — if not always of the quality — of conditions that psychiatrists diagnose as psychoses". This view conflicts with the MPC and is too restrictive, as will become clear below.

ing acts of that kind for reasons other than self-interest — a state which many contemporary philosophers would say is part of what it is for a person to think a certain action is morally wrong? Moreover, the MPC does not make up its mind whether responsibility is contingent on a person's believing that his act is "criminal" or "(morally) wrongful". If the accused's belief that his action was immoral is relevant for responsibility, then someone needs to straighten out just what "thinks the action was immoral" actually means. Such an important concept in the statement of the law ought to be reasonably precise in meaning. Just what is the cognitive condition for culpability supposed to be?²⁷

Finally, and most important, the vagueness in the concept of lack of "substantial capacity ... to ... conform his conduct to the requirements of the law". There are, of course, some things a person physically cannot do — for instance, move his car in a no-stopping zone when there is a traffic jam. The law does not require him to do these things. From the determinist's point of view a person in a sense cannot do anything other than what he in fact does, being the kind of person he has become. So presumably "lack of capacity" must refer to something else. Usually there is no question but that the accused was physically able to refrain from doing what he did, e.g., fire a revolver. Is it an inability to *control* behavior?²⁸ But if our account of motivational psychology is correct, this concept hardly belongs in a scientific psychology and what is called "self-control" boils down to some fact about the relative

²⁷ See the opinion by Judge Cardozo in *People v. Schmidt*, 216 N.Y. 324, 110 N.E. 945 (1915).

²⁸ Wechsler seemed to explain "will" and "self-control" in terms of "amenability to influence by the law". Correspondence between M.S. Guttmacher and Herbert Wechsler in *Model Penal Code, Tentative Draft # 4*, (Philadelphia: American Law Institute, 1955), printed in American Law Institute, *Model Penal Code, Tentative Drafts Nos. 1, 2, 3 and 4*, American Law Institute, 1956, p. 192.

Sir James Stephen remarked in 1883 that the "power of self-control must mean a power to attend to distant motives and general principles of conduct, and to connect them rationally with the particular act under consideration, and a disease of the brain which so weakens the sufferer's powers as to prevent him from attending or referring to such considerations, or from connecting the general theory with the particular fact, deprives him of the power of self-control". *A History of the Criminal Law of England*, II (1883), p. 170.

strength of *aversions* to a certain sort of behavior. (“Self-control” is the fact that the act/consequences are so aversive that they stand out in the agent’s thinking and he is motivated to avoid them.)

I suggest, then, that, in line with our psychological¹ knowledge, we might explain this “lack of capacity” as follows: to say that a person *lacks capacity* to conform is to say that, even if his legal/moral motivations were at an acceptable level, he might still not obey the law on account of some (other) mental/brain state he is in.²⁹ (In the case of such a person a defective state of motivations is *not* a necessary condition of his failure to obey.) This proposal would seem to allow us to excuse the effect of the sudden impulses of the kleptomaniac, or of deep brooding, or of the inability to represent the relation of consequences to action, or of some breakdown of the whole system of motivational control, perhaps as a result of brain-damage — the concept of “other mental/brain state” is open-ended.

I suggest then that what the law expects of a person, if he is to be held nonculpable for some offense, is that the strength of the agent’s standing motivations mentioned as support for obedience to law — empathy/sympathy, aversion to the act-type in question, respect for law, fear of the penal sanctions of the law — be in sum not less than an acceptable (not necessarily average) level. And the function of argument for nonculpability in the law, and of the insanity defense, is to make it reasonable to believe that what the agent did is causally (not merely logically) compatible with his (moral, etc.) desires/aversions not being below that acceptable level, although *prima facie* unlawful behavior indicates a defect of moral/legal motivation.

The basic idea of the insanity defense, then, I suggest, should be that *some* actual mental/brain state of the agent (not necessarily a “mental disease or defect” in some specified sense, like those identified as of a type listed in DSM III) other than the “standard” grounds for excuse could have prevented him from abiding by the law despite the presence of an acceptable level of moral/legal motivation. We might put this more formally as follows: “A person is culpable (responsible) only if

²⁹ Guttmacher speaks of the “overwhelming force of the unconscious in many seriously disordered patients”. *Loc. cit.*, p. 172.

his act — an act within the meaning of the law (as explained earlier) — would not have occurred *in the presence of a sum-total acceptable level of standing aversion to breach of law, and to immorality, and to injury to others, and to the specific type of action in question, and to the sanctions of the law*". Alternatively we can say that an unlawful act is *nonculpable* when an agent's mental/brain condition is such that his act (might have) occurred even if he otherwise possessed an acceptable level of moral/legal motivations.³⁰ (Of course, we need some explanation of "an acceptable level" of moral/legal motivations.) My suggestion is that the statement of the MPC should be interpreted in this way.

This proposal is not really different from the MPC in its implications for the obvious cases. It is the conceptual framework that is different, and there could be difference on border-line cases, depending on how the MPC is interpreted. (I say nothing about the final clause relating to sociopathic cases since a defense of insanity would hardly be mounted on the basis merely of a criminal record.) The MPC *could* be construed to say, if it takes motivational theory into account, that no defect of motivation is proven on account of mental disease or defect; whereas my proposal is that no defect of motivation is proven on account of the presence of *some* other state of mind, including ones not identified as some recognized form of insanity. On either view, there is no point for law to aim at improving the legal/moral motivations of the accused for they may not need improvement. Nor can punishment be justified on the ground of deterring others (not merely those mentally like the accused) in view of the infrequency of the insanity defense and the even smaller frequency of its success. My proposal enables the insanity defense to be part of a general theory of excuses. My proposal is also compatible with essentially a "product" theory of the insanity defense: the excused offense might not have occurred but for some mental/brain state of the agent different from his moral/legal motivations.

³⁰ The U.S. Court of Appeals, D.C. District Circuit (312 F. 2d at 851) held that a "mental disease or defect includes *any abnormal* condition of the mind which substantially affects mental or emotional processes and substantially impairs behavior controls" [my italics]. This leaves open the question of how to determine that a state of the mind is "abnormal".

My proposal, then, is that an agent should be held culpable or non-culpable in the circumstances cited. But it is a different question under what conditions we can properly *infer* (or the court or jury can infer) that one of these conclusions fits the evidence, that it is defective motivations or possibly “some other” mental/brain state. The answer to this can seemingly only be that the truth of one assessment or the other can properly be inferred from its being the best and simplest explanation of the unlawful act when the *total evidence* is taken into account.

It is reasonable to expect that it is the job of the *defense* to bring forward and support arguments in favor of such a claim of nonculpability, and to expect to succeed by a preponderance of the evidence, not beyond a reasonable doubt.

It is in presenting and interpreting the “total evidence” that the psychiatrist (or a character witness) will enter the picture. If he is testifying for the defense presumably he will explain the evidence for thinking that the accused’s motivation, emotional state, defect of cognitive processing, or total breakdown of the machinery (e.g., extreme schizophrenia) was of a certain sort as well as why his evidence calls for a certain sort of interpretation. It is for the jury to decide whether such an explanation offered by the defense psychiatrist, or a defective level of moral/legal motivation as asserted presumably by the psychiatrist for the prosecution, is the best and simplest explanation of all the facts.

What sort of evidence might support a conclusion that the accused should be excused as “not guilty by reason of insanity” (as distinct from being excused on some such ground as mistake of fact or duress)? One would be a record of emotional instability in the past, e.g., if the accused had been under treatment by a psychiatrist for that reason. Another would be the record of a recent head injury of a kind often followed by bizarre behavior. Another would be just reports of recent strange behavior or the reports of examination by a psychiatrist. In contrast to this would be “character”-witnesses: evidently the past manifestation of a high level of moral/legal motivation (say when there was strong self-interested motivation to do something incompatible with this) is pertinent — and we think that such motivation is relatively permanent. What the jury has to do is to make a common-sense reconstruction of the accused’s motivation. The very absence of a

motive for the crime is itself a relevant fact. The jury must try to identify the motive (if any) which led to the crime and whether this motive should have been weaker than acceptable contrary moral/legal motivation. Or it may find that the accused's motives had no rational connection with what he did, thereby pointing to an infirmity of the reasoning process rather than to defective motivation as the source of the offense.

This proposal may at first sight seem less precise and convincing than the formulation in the MPC. But provided we can make a couple of conceptions clearer, it seems more simple. Moreover, it has the virtue of placing in center stage the concept of adequate moral/legal motivations, thus making clear why an insanity defense is sensible. The law excuses when all that its sanctions can affect is, as far as we know, already in good condition, when punishment would inflict pointless harm.

I have suggested that the proposal contains some terms which call for explanation. These are: "the best or simplest explanation", and "an acceptable level of moral/legal motivation". We must say something about them.

The first does not present any real difficulty. Philosophers have argued at length about the proper meaning to be assigned to "best and simplest explanation". But the general idea is familiar enough for a jury to work with. They are presumably familiar with discussions of what is "the best explanation" of the Challenger disaster, or the collapse of a bridge, or the failure of a car to start. No more than such examples, I think, is needed for understanding the concept involved.

The second notion, that of an "acceptable level of moral/legal motivation" is another story and calls for a fuller account.

The first thing to note is that apparently the ordinary person's conception — hence the community standard — of an "acceptable level of moral/legal motivation" is roughly close to the law, as indicated by the relative severity of authorized sentences. People regard a willingness to kill another adult human being except in defense of self and others as wholly unacceptable, although there are circumstances in which a person's killing is less than normally objectionable, i.e., when there is

provocation, or extreme need. Of course, a man's foresight into consequences and the representation of them may be dimmed by angry desire which makes his behavior less objectionable (and is a mitigating factor) but still not acceptable — his anger would not have controlled his behavior if his aversion to killing had been up to par. What a person does recklessly, indifferent to a risk, say of killing someone, is also unacceptable but *less* objectionable than killing someone either as an end in itself, or as a means to the agent's ends, or as an expected consequence of an act. All this conforms with legal conceptions of merited punishment. How about assault or battery? It is clear that a person is falling short of community standards if he strikes (except in games, etc.) or even touches another in an unwelcome way. But, as in the law, this is viewed as far less objectionable than a willingness to kill. And so on down the line. For the most part, community standards agree that a man's lack of aversion to an act-type can be *ordered in degree* of acceptability, in roughly the same order as is indicated by the range of sentences authorized by the law. But only in general. Laws which impose a penalty on behavior which does not injure anyone in person or property conform less well with community standards. The same applies for the felony-murder rule. Moreover, there is so much disagreement on some matters that one can hardly claim there is a "community" standard at all, e.g., euthanasia and various kinds of sexual behavior. But despite these reservations the extension of the concept "an acceptable level of legal/moral motivation" seems not objectionably unclear. Of course, the jury will have to draw the line about when motivation is "acceptable". Members of a jury may disagree among themselves, but the hope is that their conclusion about what motivation is acceptable, after discussion, will be representative of "community standards". Presumably members of the jury will be reasonably familiar with community standards on such points, and indeed are themselves, we hope, representative of those standards, so that mostly they need only look within.

Some courts have not overlooked the importance of community standards as something which a jury should bear in mind in its decision whether an act shows defective motivation (= morally blame-

worthy). The Supreme Court of Rhode Island³¹ said that the jury is to “evaluate the defendant’s blameworthiness in light of prevailing community standards”.

There is, of course, no black and white line here which makes it easy for the jury to decide which is the “best explanation”. One can feel some sympathy with George Will’s remark,³² cited by Joseph E. diGenova and Victoria Toensing,³³ “The most morally indefensible crimes are becoming the most legally defensible... the more odious the crime — premeditated or spontaneous — the more reasonable doubt there is about the person’s sanity at the time”. This objection, however, will hardly be convincing to readers who reflect on a recent case of Henriette Cornier, a woman who asked to be allowed to take a neighbor’s 19-month-old child for a walk, took the child up to her room, severed its head, threw the head out of the window, and calmly waited for the police.³⁴ Is this a “morally indefensible crime?” The act was morally wrong all right, but this is not to say it was morally blameworthy. Was there possibly something very wrong with the accused’s mind/brain state, different from her standing moral/legal aversions, which produced such a hideous action? The decision by the jury in some such cases may not be easy. Mr. Will is right that when something horrible is done which seems wholly unmotivated, which members of a jury cannot imagine a normal person doing in those circumstances, there *is* some presumption of “some other” mental problem, unrelated to standing moral/legal motivations. This is where the record of previous behavior (and while in detention) is relevant. The jury may make mistakes where there is no sharp line but the job of the jury is to do the best it can.³⁵ (It is not clear that the decision is

³¹ State v. Johnson 399 A. 2d 469, 1979. Cited by D. H. J. Hermann, *The Insanity Defense* (1983: Charles Thomas, Springfield, Il.), p. 55 f.

³² *Baltimore Sun*, June 24, 1982, A19.

³³ ‘The Federal Insanity Defense: a Time for Change in the Post-Hinckley Era’, *South Texas Law Journal*, 1984, 728.

³⁴ Cited by Maeder, *op. cit.*, p. 42.

³⁵ The British Royal Commission on Capital Punishment proposed (1953) an equally indeterminate job for the jury, affirming that a person is not responsible for his unlawful act if “at the time of the act the accused was suffering from disease of

any harder than that required by MPC: whether the accused was suffering from such mental defect or illness that he was substantially unable to conform to, or appreciate the legality [morality] of his action.) The severity of the problem is somewhat mitigated, as we shall see in a moment, in that the disposition of convicted criminals and those judged not guilty by reason of insanity will not be very different in a rational system of penal and mental institutions. Incidentally, the proposal comes out exactly where the MPC does on the matter of sociopathic individuals who seem precisely to lack adequate moral/legal motivation and are therefore guilty.

V. THE DISPOSITION OF THE NOT GUILTY BY REASON OF INSANITY

What should happen to persons who are judged not guilty by reason of insanity or "guilty but insane"? Obviously they should be committed to an institution for observation and treatment. Release could come fairly rapidly (the court having decided they were "insane" at the time of the offense but possibly not now). The use of drugs may remove the symptoms (say, of schizophrenia) quite rapidly and they may remain suppressed for some length of time, perhaps permanently. Psychological therapy can have the same effect. If the crime was a crime of violence, of course, procedures should be more strict for it is already known that the person is somewhat dangerous to society. In such cases gradual discharge, probation, frequent reports to the therapist for a time, and general monitoring of progress are in order and should be mandatory.³⁶ Some of these may be beyond treatment, and hence their free circulation in society would be a dangerous thing. The prospect of early release will not encourage average persons with standard legal/

the mind (or mental deficiency) to such a degree that he ought not to be held responsible". Cited by S. H. Kadish, S. J. Schulhofer, and M. G. Paulsen, *Criminal Law and its Processes* (Boston: Little, Brown and Company, 1983), p. 834.

³⁶ As proposed by the court in *Benham v. Edwards*, U.S. District Court ND Georgia 501 Suppl 1050, 1980.

moral motivations to break the law. Deterrence will be maximized for the average person by the mere prospect of a criminal trial. The stigma of being judged mentally defective will itself be a substantial discouragement to crime.

How will this differ from treatment of persons who are convicted, in view of their unlawful action, having manifested a defective level of their moral/legal motivations? A defect of moral/legal motivation is itself a kind of mind/brain defect. One hopes that many cases of this can be treated successfully by therapy, probation with regular reports to a probation officer, placement in a job, and so on. In the case of "sociopaths" or "hardened criminals" this may not be possible. It may be that their defect was a result of experiences during the early years of life and cannot be remedied by anything done later. Then they should be kept out of circulation for the sake of protection of society, except in so far as determinate sentences are required by the values of a liberal democratic society.

There is another point to be remembered: if one is a determinist, one will think that defect of legal/moral motivation resulting in crime is no less a matter of genes, upbringing, etc., than is the case with the mind/brain problem of those adjudged insane. Some persons have had the advantage of good genes (intelligence, health, energy) and family upbringing and location in society which not only produce good moral/legal motivation, but also assure that they will not be put in a position where they are strongly motivated to disobey the law. With others the opposite is the case. There is an unjustified inequality in the lottery of life, which bestows good things on some and bad things on others. We all have a moral obligation to work toward the amelioration or removal of these inequalities. But that is not the job of the criminal law. The system of criminal justice is an uneasy compromise, attempting to accommodate both the need to protect society from harm and to avoid imposing pointless suffering on those who have broken the law.

Department of Philosophy
University of Michigan
Ann Arbor, MI 48109
U.S.A.