

THE JURISPRUDENCE OF PUNISHMENT

Kyron Huigens

DRAFT – Please do not quote, cite, or reproduce

I. Introduction

Theories of punishment are often categorized as “deterrence theories,” “retributive theories,” “rehabilitative theories,” and so on. This is a mistake because deterrence, retribution, rehabilitation and the like are functions of punishment. There is no reason to treat the designation of one function or another as the defining feature of a theory of punishment. A better system refers to the principal schools of moral theory, so that we have consequentialist theories of punishment, Kantian theories of punishment, and so on. But this way of organizing the field is misleading. Current theorizing about legal punishment features a quick and uncritical traffic between moral categories and concerns and legal categories and concerns, and this has the potential to confuse. Conducting punishment theory more carefully and self-consciously as legal theory – as the jurisprudence of punishment – would dispel that threat. A jurisprudence of punishment would not eliminate moral categories and concerns from punishment theory. It would simply require us to direct the traffic between legality and morality along more precise routes.

The second part of this paper, following this introduction, will describe the place of moral theory in the theory of legal punishment as it is ordinarily conducted. The third part will explain why, under most schools of jurisprudence, including legal positivism, this is not problematic.

The fourth part of this paper, however, is more ambitious. Ordinary punishment theory portrays mens rea, culpability or (in the term preferred here) criminal fault in particular as a moral determination. This is implausible on its face. If one accepts the thesis that law and morality are separate normative systems, then the picture is one of the jury’s crossing over into a different normative system at a crucial moment in their deliberations, and crossing back carrying an essential component of legal punishment. The fourth part of this paper describes a version of legal positivism, exclusive or hard positivism, that would bar this scenario. The challenge then is to describe criminal fault in a way that is consistent with exclusive legal positivism. I describe a “specification” account of criminal fault that I have advanced before, but from the point of view of exclusive legal positivism. This description of fault avoids the implausible notion that juries move unhindered between fundamentally different normative systems, and has several other clarifying effects.

II. The Scapegoating Objection in Punishment Theory

It would be difficult to describe in a paper of any length how punishment theory is ordinarily conducted. Here, I will only put forward an instance of it, and proceed on the assumption that it is reasonably representative. The scapegoating objection is a staple of

criminal law classrooms, but upon closer examination one begins to wonder why. It has nothing in particular to do with law – which is just the feature that makes it representative.

The argument is aimed particularly at consequentialist theories of punishment. To quote one statement of the objection:

[T]he utilitarian must hold that we are justified in inflicting pain always and only to prevent worse pain or bring about greater happiness. This, then, is all we need to consider in so-called punishment, which must be purely preventive. But if some kind of very cruel crime becomes common, and none of the criminals can be caught, it might be highly expedient, as an example, to hang an innocent man, if a charge against him could be so framed that he were universally thought guilty; indeed this would only fail to be an ideal instance of utilitarian ‘punishment’ because the victim himself would not have been so likely as a real felon to commit such a crime in the future; in all other respects it would be perfectly deterrent and therefore felicitic.¹

It is never quite clear, however, what this objection means. This version of the objection is the one that John Rawls quoted in “Two Concepts of Rules,” and he understood the point to be this: “The question is whether utilitarian arguments may be found to justify institutions . . . such as one would find cruel and arbitrary.”² But why should we care whether utilitarian arguments would justify cruel and arbitrary punishments?

Rawls offered a plausible argument against the scapegoating objection.³ It may be that conditioning punishment on individual desert produces optimal social welfare, simply because most people believe, rationally or not, that this is how punishment should be done, and because they would withhold necessary political support for the legal system if it were done otherwise. If punishment according to desert is not cruel and arbitrary, then a consequentialist theory of punishment would not necessarily endorse cruel and arbitrary punishment in the pursuit of optimal social welfare.

For my present purposes, however, it is irrelevant whether or not the scapegoating objection fails and whether or not consequentialist punishment theory is right. I want to focus on why and how the scapegoating objection *purports* to show a consequentialist theory to be mistaken, whether it successfully does so or not. On this point, Rawls’s counterargument is unhelpful, because it merely presents the question again. Granted that a consequentialist theory of punishment does not necessarily endorse cruel and arbitrary punishment in the form of punishment regardless of desert, why does it follow that a consequentialist theory is not objectionable?

The answer to this question sheds some light on the kind of theory the theory of punishment usually is. Consider a different way in which the scapegoating objection

¹ E.F. Carrit, *ETHICAL AND POLITICAL THINKING* 65 (1949).

² John Rawls, *Two Concepts of Rules*, *PHILOSOPHICAL REVIEW* (Jan. 1955)

³ The argument was formulated by several other philosophers as well. See, e.g., A.C. Ewing, *THE MORALITY OF PUNISHMENT* 66 (1929); J.D. Mabbot, *Punishment*, *MIND* (apr. 1939), reprinted in *THE PHILOSOPHY OF PUNISHMENT: A COLLECTION OF PAPERS* (H.B. Acton, ed. 1969) at 39, 50. The reply was also given contemporaneously to Rawls, in Anthony M. Quinton, *On Punishment*, *ANALYSIS* (June 1954), reprinted in Acton, *supra*, at 55, 63. H.L.A. Hart also drew the distinction, in terms of “general justifying aim” versus “distribution,” and it is possible to read Hart as defending a consequentialist theory of punishment as Rawls and Quinton did. See Kyron Huigens, *The Dead End of Deterrence, And Beyond*, 41 *WM. & MARY L. REV.* 943, 959. As argued below, however, Hart’s argument is different from theirs, integrated as it is into his general jurisprudence.

might fail. We might not care whether a scapegoat is punished. If this seems implausible, consider a closely analogous case instead. A number of states have abolished the insanity defense, and thus authorize punishment for those who have long been thought not to deserve punishment. The reasoning behind this abolition seems to be twofold: that it will preclude the acquittal of malingerers, even at the cost of punishing some genuinely undeserving defendants; and that it will send a strong deterrent message to potential wrongdoers. If the scapegoating objection turns on our revulsion at undeserved punishment's being inflicted for deterrence's sake, then these explanations give the punishment of the insane a strong odor of scapegoating.

If we aren't repelled by the infliction of undeserved punishment for the sake of deterrence, then the scapegoating objection doesn't work. The act-consequentialism that is the target of the scapegoating objection describes the abolition of the insanity defense perfectly well, and makes sense of its rationale. By bolstering deterrence, albeit at the cost of punishing otherwise legally irresponsible defendants, we enhance social welfare. Thus far, at least, a consequentialist theory of punishment is a perfectly sound theory, though we might find reasons to employ a more sophisticated consequentialism as we develop a more complete picture. The abolition of the insanity defense also obviates the counterargument against the scapegoating objection. It might be that to punish without regard to desert would cause a punishment system to lose critical political support, so that only punishment according to desert will optimize social welfare. Where the insanity defense has been abolished, however, it appears that to punish without regard to desert, at least in this respect, garners the punishment system political support. With no offset for dissatisfaction over punishment of the undeserving, punishing the insane will produce optimal welfare – without the need to introduce desert into the equation at all. In short, under present circumstances, the scapegoating objection gives us no reason to think that a consequentialist theory of punishment is mistaken. Quite the contrary: consequentialism, even in its most primitive form, makes sense of a recent development that at first seems anomalous, even outrageous. What else should we expect from a theory of punishment than an accurate description of our actual practices and a cogent explanation of the reasons behind them?

There may be no reason to expect something different from a theory of punishment, but we do. The scapegoating objection is not meant to evaluate theories of punishment. It is meant to evaluate theories of just punishment. Consequentialism as a moral theory is evaluated by running it up against common judgments about the good or right action. Consequentialism as a punishment theory is evaluated by running it up against common judgments about just punishment. Hence the scapegoating objection, and consequentialism's purported error in describing unjust punishment as just. To treat the scapegoating objection as something other than a way to evaluate theories of just punishment is to miss its point. It is no response to the scapegoating objection to say that a simple act consequentialism describes existing practices perfectly well, if those practices themselves are implausible as instances of just punishment.

A test for theories of just punishment, however, is not a test for theories of legal punishment. The scapegoating objection applies just as well to the punishment of children and the sanctions of social etiquette. It would be unjust to spank a child because I could not catch a different child who broke a window. It would be unjust to disinvite one friend to my party because I cannot disinvite the friend who insulted me, but whose

attendance is necessary to make the party a success. A punishment theory that authorized these acts would seem to be mistaken. But in this light, the scapegoating objection has nothing in particular to do with law.

The scapegoating objection gives us another reason to think that punishment theory ordinarily is not conducted as legal theory at all. The objection has no bearing on the dominant positivist tradition in jurisprudence. It is not, at least, the descriptive sociology that Hart thought his jurisprudence to be. Its subject is justice, not law. Hart failed in his aspiration to purely factual description in a number of ways, but none of the variations on Hart's project suggested by these failures has any obvious use for the scapegoating objection. To take the only remotely pertinent example, Stephen Perry has argued that isolating a concept of law for analysis requires normative choices as well as description, but the normative choices Perry is concerned with do not concern justice as such, as the scapegoating objection does. Its argument is not Perry's argument, any more than it is Hart's.

In spite of the irrelevance of the objection to positivism and its tangential relationship to legality, the scapegoating objection is a perennial feature of criminal law theorizing. *The* question in punishment theory is the justification of punishment, and it is the moral justification of punishment that is at issue. The interesting theoretical questions about criminal law doctrine also revolve around moral theory. An analysis of the defenses is an analysis of the blameworthiness of the accused. But to ask whether the accused is blameworthy is to ask moral questions. Does the accused meet the conditions of moral responsibility such as sanity and majority? Do his actions reflect, not only a nominal violation of a conduct rule, but also the added dimension of fault – a moral evaluation seemingly impossible to cabin in doctrinal terms – that makes punishment morally defensible? These are the principal questions of punishment theory, and it is hard to see them as questions of legal theory at all. They seem, like the scapegoating objection, to apply equally well to the punishment of children and punishment for violations of social etiquette. They seem to belong to moral philosophy.

III. Criminal Fault and Moral Fault

Perhaps the situation is not so bad. Is it true that the interesting theoretical questions about criminal law doctrine have been treated as moral philosophy? If so, is this really objectionable? The answer to both questions is no. Punishment theory's preoccupation with morality does not mean that it is being conducted as moral theory instead of legal theory. Some schools of jurisprudence do not draw a hard distinction between morality and law, or between moral theory and legal theory, but they are still reasonably identified as legal theory. Ordinary punishment theory can find a jurisprudential home here. Even if there is a clear disagreement between such normative jurisprudence and positivism on the separability thesis itself, particular analyses of law in relation to morality do not fall into two starkly divided camps. Ordinary punishment theory's preoccupation with morality can be accommodated by prominent versions of positivism.

Consider the issue from the point of view of criminal law theory, specifically on the question of legal desert for legal punishment. Criminal law addresses desert in two distinct kinds of defense, corresponding to two distinct uses of "desert." The accused might deny that he is a fully competent agent and, as such, a fair candidate for legal

punishment – as in the defenses of insanity, minority, or duress. Or the accused might contest an aspect of the alleged wrongdoing itself, the distinctive element in criminal offenses that is missing in cases of mistake of fact, for example. This latter defense, the absence of fault, will be the focus in this paper, for the most part.

Fair candidacy and fault are usually portrayed as requirements of moral desert. For example, Jeremy Horder has proposed the adoption of a “diminished capacity” defense that combines fair candidacy and fault considerations. From a passage such as this, which is not atypical, it seems that desert is a straightforward moral side constraint on legal punishment.

As when murder is reduced to ‘voluntary’ manslaughter, there is thus a formal role at the stage of conviction itself for the fact-finder to decide the fate of the offender on a broader range of moral criteria than is currently possible, whether or not the definitional elements of the offence in question involve some kind of moral appraisal (p 146).

At the same time, however, it seems that these conditions on the legal validity of judgments of punishment cannot be moral conditions, properly speaking. Fair candidacy and fault are paradigmatically determined by a jury. As a legal entity within a legal system charged with making legal decisions with legal consequences, a jury never decides moral questions. It can, of necessity, decide only legal questions.

Most theorists of punishment would describe these features of legal punishment as moral requirements that are incorporated into law. However, some legal theorists go beyond this incorporationist view to portray the requirements of valid legal punishment as moral requirements, full stop. Consider John Finnis’s description of criminal fault:

Sanctions are punishment because they are required in reason to avoid injustice, to maintain a rational order of proportionate equality, or fairness, as between all members of society. For when someone, who really could have chosen otherwise, manifests in action a preference (whether by intention, recklessness, or negligence) for his own interests, his own freedom of choice and action, as against the common interests and the legally defined common way of action, then in and by that very action he gains a certain sort of advantage over those who have restrained themselves, restricted their pursuit of their own interests, in order to abide by the law.

Notice that Finnis’s concern here is legal punishment in particular, not punishment in general. This is a description of fault that does not translate to the punishment of children or social sanctions on violations of etiquette, neither of which concerns “the legally defined common way of action.” This description of fault is, instead, fully integrated into Finnis’s jurisprudence.

A critical component of Hart’s account of law is his argument that law is not merely a system of commands backed by threats. Instead, from “the internal point of view” it is clear that law creates a sense of obligation under law and a basis for criticism of actions inconsistent with law. Finnis agrees with Hart that the right perspective for theorizing about law is the internal point of view – that of a participant in the legal system who feels obligations within the system. He argues, though, that the internal point of view can be and should be further specified. Participants in the system vary in their allegiances and attitudes to law, so there is a choice to be made about the participant whose view of law matters. The central or focal case of the internal point of view is that

of a person of sound morality or, more precisely, a person of practical reasonableness. 15 From this, Finnis concludes:

A sound theory of natural law is one that explicitly, with full awareness of the methodological situation just described, undertakes a critique of practical viewpoints, in order to distinguish the practically unreasonable from the practically reasonable, and thus to differentiate the really important from that which is unimportant or is important only by its opposition to or unreasonable exploitation of the really important. A theory of natural law claims to be able to identify conditions and principles of practical right mindedness, of good and proper order among men and in individual conduct. 18

Finnis's natural law description of criminal fault identifies "conditions and principles of practical right mindedness, of good and proper order among men and in individual conduct." Specifically, the defendant who is criminally at fault "manifests in action a preference (whether by intention, recklessness, or negligence) for his own interests, his own freedom of choice and action, as against the common interests and the legally defined common way of action." This is an account of criminal fault as moral fault, full stop.

Jeremy Horder's account of criminal fault, implicit in the passage quoted above, is of the same kind. Ronald Dworkin, like Finnis, describes legal validity in terms of moral validity, and Horder's account is an explicitly Dworkinian view. Horder's recent book on the excuses (a British term that translates into the absence of fault in American terms) offers three proposed defenses, "suggesting three ways in which the law should be developed to become *distinctively* liberal in its excusatory outlook" (197). According to Horder, those defendants who are 'shortcomers' fail at least sometimes because of their mental or emotional make-up (p 163). Horder's diminished capacity excuse would allow a defendant 'to combine evidence of lost self-control following something rather less than the gravest of provocations, with evidence of a mental deficiency falling short of insanity or of some other serious mental disorder' (p 150). This defence is meant to serve the liberal value of reciprocity between society and the individual. This reciprocity, Horder argues, entails an opportunity for the defendant to raise moral arguments against criminal liability. Horder writes: "Fact-finders should be provided with a more general, formal means of expressing an opinion about the moral appropriateness of conviction." (p 144).

These proposed defenses might be read as liberal law reform, but Horder's invocation of Dworkin indicates a different ambition. Horder describes his proposed defences, including his "diminished capacity" defense as serving the Dworkinian legal values of 'equal concern and respect' (p 4). This is not mere liberal advocacy. It is instead a way of describing law in terms of moral validity. According to Dworkin, law is valid when put in the best moral light in the course of its interpretation; and law is put in its best moral light when it demonstrates equal concern and respect. Because they rely on an interpretive jurisprudence that incorporates the morality of political liberalism, Horder's proposed defenses, and his defense of them, constitute an account of criminal fault as moral fault, full stop.

While positivists would reject such a view of criminal fault as inconsistent with the separability thesis, most do not reject the notion that moral validity can be incorporated into accounts of legal validity, and would not reject an account of criminal fault as incorporating moral fault. Hart rejected the simplistic view that crime is "the intentional or reckless doing of a morally wrong act," and carefully distinguished between the principle that it is morally wrong to punish those whose wrongdoing is not voluntary (or intentional) and the principle that only those who have voluntarily (or

intentionally) committed a moral wrong may be punished. Nevertheless, there are strong suggestions in Hart's essays on punishment that a correct description of criminal fault can incorporate a description of moral fault.

Consider his treatment of the issue of desert as it arises in the scapegoating objection. In *The Prolegomenon to the Principles of Punishment*, Hart considered Rawls's answer, to the effect that optimal welfare might require that society not be unsettled by regular scapegoating. Then Hart went on:

But official resort to this kind of fraud on a particular occasion in breach of the rules and the subsequent indemnification of the officials responsible might save many lives and so be thought to yield a clear surplus of value. . . In extreme cases many might still think it right to resort to these expedients *but we should do so with the sense of sacrificing an important principle. We should be conscious of choosing the lesser of two evils*, and this would be inexplicable if the principle sacrificed to utility were itself only a requirement of utility.⁴

The difference between Hart's and Rawls's analyses is that Rawls takes the external point of view and Hart takes the internal point of view. The requirement of desert as a condition of punishment can be an obligation felt by a participant in the legal system, and so can be an obligation imposed by that legal system.

Hart might be referring here only to legal desert, with no suggestion that the desert required is moral desert as well. But he left us no account of desert in exclusively legal terms, and his other descriptions of criminal fault strongly suggest that the moral requirement of desert can be incorporated into law. In a later essay, Hart wrote:

Human society is a society of persons; and persons do not view themselves or each other merely as so many bodies moving in ways which are sometimes harmful and have to be prevented or altered. Instead, persons interpret each other's movements as manifestations of intentions and choices. . . The bearing of this fundamental fact on the law is this. If as our legal moralists maintain it is important for the law to reflect common judgments of morality, it is surely even more important that it should in general reflect in its judgments on human conduct distinctions which not only underly morality, but perforce the whole of our social life. 182-83

The phrasing here suggests that the question is prescriptive, concerning how the law ought to be written. It is important to note, however, that the passage occurs in a discussion of whether desert has any necessary connection to retributivist (or deontological) accounts of criminal fault, to the exclusion of consequentialism. That is, the discussion occurs at the same level of controversy occupied by the scapegoating objection – that concerning the choice of theory. Hart's argument is offered in support of the thesis that our choice of theory should not be affected by our views on moral fault, because moral fault can be a feature of criminal fault on any theoretical description of legal punishment. This argument assumes that morality can be implicit in criminal fault as a descriptive matter.

This reading is consistent with Hart's "soft" or incorporationist legal positivism. Hart argued that in some hard cases law runs out, and judges have discretion to make law. They might do this by making the best moral decision they can. Dworkin argued that controversial moral claims could not be incorporated into law in this way without fatally detracting from the certainty supplied by secondary rules, as the positivist account of law would have it. Hart replied that the tolerable level of uncertainty in a legal system varies, and so there is no reason to expect the incorporation of moral principles to fatally

⁴ prolegomenon @ 11-12 (emphasis added)

undermine it. Dworkin also argued that legal positivism is motivated by doubts about moral realism and a resulting positivist rejection of moral criteria in law. Hart denied that any such anti-realist motivation can be found in legal positivism, because legal positivism is simply agnostic on the question of moral realism. “For whatever the answer is to this philosophical question, the judge’s duty will be the same: namely, to make the best moral judgment he can on any moral issues he may have to decide.” 354

The criticism that theorists of punishment ought to avoid descriptions of moral desert in their descriptions of legal desert is, it seems, unwarranted. Punishment theorists who equate moral fault and legal fault are well supported in that assumption by eminent philosophers, notably Finnis and Dworkin. Those punishment theorists who incorporate moral fault into their descriptions of legal fault are supported in this by Hart, as well as by many other legal positivists, including Jules Coleman and Stephen Perry. This paper doesn’t attempt in any way to adjudicate between these schools of jurisprudence, but only to note that particular theories of legal punishment are potentially consistent with at least one of them. My only recommendation to the punishment theorist is to consider which kind of jurisprudence is congenial to her take on punishment. How does one conceive of the relationship between legal and moral validity – with respect to fault, to fair candidacy, or to any other issue in criminal law that seems to involve moral categories? If some attention is not paid to this question, then the confusion inherent in conducting a theory of legal punishment as moral theory is a danger.

So far, however, there is no reason to fear that ordinary theories of legal punishment are mistaken in a fundamental way, by not describing law at all. The theory of punishment as it is ordinarily conducted is not a subdivision of moral theory; or, if it is, then that status can be explained in Finnisian or Dworkinian terms that bring out the legal features of this subdivision. However, it is still possible that theorists of legal punishment *have* made the fundamental mistake of not describing law at all. There is a school of legal positivism that rejects both the notion of law as continuous with morality and the notion of law as incorporating morality. If this way of thinking is correct, then punishment theory as it is ordinarily conducted would have to change, in order to avoid describing legal punishment in moral terms.

IV. Criminal Fault as Legal Fault

Joseph Raz has advanced an exclusive legal positivism; that is, a legal positivism that excludes descriptions of moral validity from descriptions of legal validity. One of the hallmarks of a positivist theory of law is the sources thesis. That is, law is identifiable by its institutional source, as opposed to its moral validity. To describe law as law requires one to trace the law back to this source; to describe its pedigree. The difference between inclusive and exclusive legal positivism has to do with the pedigree of moral norms. For the inclusive legal positivist, a moral norm is a legal norm if it is authorized by a pedigreed legal norm. For the exclusive legal positivist, a moral norm is a legal norm only if it has its own pedigree.

Raz’s exclusive legal positivism relies heavily on a “service” conception of legal authority. The service provided by law is the exclusion of reasons that might mislead one into ill-considered action. This conception of authority is analogous to Raz’s conception of the role of rules in practical reasoning. We have reasons for action, and practical rationality consists largely in our acting on the balance of reasons. However, we also

have second order reasons – reasons about reasons – that sometimes dictate that we not act on the balance of reasons. Rules are second order, exclusionary reasons of this kind. A rational rule points us toward action according to the balance of reasons in a range of similar situations of choice. If we follow a rule consistently, however, we will inevitably find ourselves acting against the balance of (first-order) reasons on some occasions. But this is still rational, because the rule has enabled us to act in accordance with the balance of reasons in the vast majority of cases. The rule has, moreover, enabled us to do this both consistently and without further deliberation – deliberation that is costly in itself, and that might have led us to erroneous conclusions and irrational actions.

Authority is like rules in this respect. The directives of an authority should be based on reasons that apply to individuals making their own decisions. The justification of authority lies in the fact that an individual has a better chance of acting consistently according to right reasons if he follows the directives of authority. In order to provide this service, however, an authority has to preempt the reasons on which an individual might have acted. This is part of the nature of authority. Without its preemptory force, the reasons that an authority gives are not authoritative at all. “What distinguishes authoritative directives is their special preemptory status. One is tempted to say they are marked by their authoritativeness.” 212

The exclusive nature of Raz’s legal positivism is dictated by this conception of authority. Moral reasons can be reasons on which law is based, because they are reasons for which individuals act. But moral rules as such cannot have legal authority; they are instead the kind of reasons that legal authority preempts. Without its preemptory force – law’s power to exclude non-legal reasons – legal authority is not authoritative. Moral reasons as such cannot be incorporated into law without depriving law of its authority. Since law must be the kind of institution that at least purports to exercise authority, law cannot incorporate moral reasons in the way that inclusive legal positivism purports to do. If a pedigreed legal norm merely designates or recognizes a moral norm, the moral norm itself still lies outside the line of authority. To apply it reverts to reasons the agent has independently of law, and thereby detracts from legal authority’s necessary preemptive power. Only a moral norm that has its own pedigree, that lies wholly within the line of legal authority, will not detract from the necessary preemptive power of legal authority.

Applied to the description of criminal fault, this exclusive legal positivism undercuts the notion that criminal fault can consist of moral fault (as Finnis and Horder would have it) or incorporate moral fault (as Coleman or Perry would have it). To see this, start two procedural steps forward from the jury’s finding of criminal fault. In the paradigmatic criminal case, a jury’s finding of criminal fault is embodied in a verdict. A jury verdict is an authoritative legal directive to another legal authority. The trial judge is bound by the verdict to enter a particular judgment concerning punishment. The trial judge cannot adduce moral reasons and enter a judgment that contradicts or alters the verdict. If this were possible, if the verdict were treated as just another reason supporting the judgment, then the verdict would lack preemptory force. Without that preemptory force, it would not be authoritative and could not be described as law. For all we know so far, a verdict might consist of or incorporate moral reasons. But given that a verdict is law, preemptorily binding on a judge entering judgment, a legal judgment of punishment

does not consist of or incorporate moral reasons; the verdict on which it is based excludes them.

Taking one procedural step back, we can see that a verdict in a criminal case cannot consist of or incorporate moral reasons, particularly with respect to a finding of criminal fault. A jury's finding on fault is an authoritative legal directive to another legal authority. The jury at that later stage of its deliberations at which it issues its verdict is bound to issue a verdict reflecting its precursor finding on criminal fault. The jury cannot later adduce moral reasons and enter a verdict that contradicts or alters its precursor finding on fault (as happens in cases of jury nullification). If this were possible, if the finding on fault were treated as just another reason supporting the verdict, then the finding on fault would lack peremptory force. Without that peremptory force, it would not be authoritative, and could not be described as law. For all we know so far, a finding of criminal fault might consist of or incorporate moral reasons. But given that a finding of fault is law, peremptorily binding the jury when it later issues its verdict, a verdict does not consist of or incorporate moral reasons; the finding of fault on which it is based excludes them.

Taking one last step backward, we can see that a finding on criminal fault cannot consist of or incorporate moral reasons. The law of the jury's instructions is obviously authoritative for the jury. A deliberating jury is bound to make a finding on fault that accords with its instructions. The jury cannot adduce moral reasons and make a finding on fault that contradicts or alters its instructions. If this were possible, if the instructions were treated as just another reason supporting the finding on fault, then the law of the instructions would lack peremptory force. Without that peremptory force, this law would not be authoritative, and could not be described as law. For all we know so far, the law of the instructions might consist of or incorporate moral reasons. But given that the law of the instructions is law, peremptorily binding the jury when it makes its finding on fault, a finding on fault does not consist of or incorporate moral reasons; the law of the instructions on which it is based excludes them.

If Raz's account of legal authority is correct, then punishment theory is in need of a description of criminal fault that has this kind of relationship to the law of the jury's instructions – that is, to the prohibitions of the criminal law. In a number of previous articles, I have advanced a conception of criminal fault that is part of a larger project: a theory of legal punishment built along the lines of Aristotelian virtue ethics. One would naturally assume that this theory describes criminal fault as moral fault, as Finnis does. However, an Aristotelian punishment theory need not feature such a conception of criminal fault. Practical reasonableness does not have to play the role with respect to fault that Finnis describes. A theory of legal punishment might be recognizably Aristotelian because it draws on other features of virtue ethics instead, such as its rich conception of deliberation on ends and its account of resulting responsibility for ends. In this fashion, the conception of fault in my version of an Aristotelian punishment theory draws on the moral particularism of virtue ethics, but is otherwise unlike Finnis's conception of criminal fault.

Let me lay out this conception of criminal fault as briefly as I can, and then argue for its compatibility with exclusive legal positivism. I will conclude with some comments on why this compatibility matters.

Criminal fault is an inference, drawn in the course of the adjudication of wrongdoing, to the effect that the practical reasoning of the defendant is deficient. This deficiency is relevant in one way or another to most of the reasons usually cited as functions of punishment: deterrence, retribution, incapacitation, education in public norms, social catharsis, and so on. Practical reasoning is deficient if it pays no heed to public norms; if the reasoning agent is not deterred by the prospect of punishment, and so on. This assessment of the quality of the defendant's practical reasoning is not limited to his reasoning in connection with the offense – even if “in connection with” is given a very broad construction. It extends, in addition, to an assessment of the defendant's set of standing motivations, or ends – to their acquisition, development, maintenance, and ultimate issuance in the alleged offense. From an opposite perspective, criminal fault is an aspect of criminal wrongdoing. That is, the manner, circumstances, and specifics of the individual instance of wrongdoing alleged against the defendant are the subject matter of the adjudicative assessment of the quality of his practical reasoning that I have just described.

This adjudicative assessment of fault is related to the positive, legislated criminal prohibition in two ways. First, the adjudication of fault is complementary to the general prohibition of the statute. From the point of view of an assessment of the quality of practical reasoning, the legislation of a criminal prohibition is a generalization about how people reasoned well or poorly in a large set of specific, more or less similar, cases. A jury determining criminal liability applies the prohibition to a particular set of facts, and in the course of determining the relevance of one to the other, the prohibition is made more specific. In other words, a finding of fault requires the jury to return the generalized prohibition to the level of specificity at which it originated. At this specific level, the actual course that the defendant followed, and the quality of decisions he made in that regard, can be meaningfully assessed. With respect to the justification of punishment, this complementary specification of the prohibition brings whatever justifying force the prohibition has down into the individual case.

Second, because legislation is necessarily general, the description of criminal fault in criminal prohibitions leaves gaps regarding specific cases – gaps that are filled in the adjudication of fault, both at trial and in sentencing, by the process of specification. For example, we might use “premeditation” to define the requisite fault in the prohibition of homicide, but relative to the background justification of the prohibition this might prove to be a hugely over- and under-inclusive rule. A grieving wife who pulls the plug on her terminally ill and gravely suffering husband, premeditates just as much as a contract killer; but with respect to fault these two defendants certainly seem to differ. The difference in fault is “interstitial fault,” that is, fault in the interstices of legislated, positive fault. Interstitial fault is mostly relevant to sentencing, but even in the trial on the merits it will have an effect on the jury, at least at the margins. For example, one would expect it to make an acquittal in the mercy killing slightly more likely and an acquittal in the contract killing slightly less likely (quite aside from the possibility of jury nullification).⁵

⁵ Darryl Brown has drawn on empirical studies of jury decision-making to show how juries do justice within the limits of their instructions. Darryl Brown, *Jury Nullification Within the Rule of Law*, 81 *Minn. L. Rev.* 1149 (1997); Darryl Brown, *Plain Meaning, Practical Reasoning, and Culpability: Toward a Theory of Jury Interpretation of Criminal Statutes*, 96 *Mich. L. Rev.* 1199 (1998).

This analysis brings to the fore a value that constantly competes, in criminal law, with the rule of law value of formality. Both features of this conception of fault reflect a concern for fine-grainedness in criminal law. Fine-grainedness can be defined as a low level of under- and over-inclusiveness in our rules of criminal liability, relative to a background of moral desert; or defined, alternatively, as a high degree of congruence between our legal judgments of desert and our moral judgments of desert. This attention to similar moral judgments creates an inevitable tension in the criminal law that helps to explain a number of its features and anomalies. The pursuit of fine-grainedness requires a relaxation of formality and detracts from our pursuit of traditional rule of law values such as notice of the prohibition and *ex ante* legislation. The pursuit of formality, on the other hand, entails a loss in fine-grainedness and at least the risk of a loss in the credibility and prestige of the criminal justice system. Both adjudicative findings on fault that complement legislated prohibitions and the regulation of interstitial versus positive fault in legislation, adjudication, and sentencing, serve the purpose of balancing formality and fine-grainedness.

This conception of criminal fault is, I believe, consistent with Raz's exclusive legal positivism. Notice, first, that it is the necessity of fine-grainedness in criminal law that has led ordinary punishment theory to rely on or incorporate descriptions of moral fault into its accounts of criminal fault. The law's aspiration to a low level of under- and over-inclusion relative to moral desert makes legal desert appear to be, ideally, identical with moral desert. If a high degree of congruence between judgments of legal desert and judgments of moral desert is the goal, then it is natural to conclude that legal desert is, in the end, moral desert. The jurisprudential question is the nature of this identification of judgments of legal desert and judgments of moral desert.

The "trackback" conception of criminal fault from an exclusive legal positivist's point of view, given earlier, accords with the "specification" conception of criminal fault – fault as the specification of a legal prohibition – given just above. The "trackback" account of criminal fault is nothing other than a description of the pedigree of a jury's finding on criminal fault, proceeding from judgment to adjudication to legislation. The "specification" account is also an account of this pedigree; it merely proceeds in the opposite direction from the "trackback" account, from legislation to adjudication to judgment. If a jury's finding on fault is a moral norm, then it is legally binding on the entry of the verdict and the ultimate judgment of punishment because it has its own pedigree – just as the verdict and the judgment do – tracing back to the legislature's creation of a criminal prohibition. A jury's decision on a criminal law defense is not merely a formal opportunity "for the fact-finder to decide the fate of the offender on a . . . range of moral criteria," as Horder's Dworkinian account would have it. Neither does a jury "reflect in its judgments on human conduct distinctions which . . . underly morality," as Hart would have it. The law does not merely recognize and authorize moral determinations on fault and fair candidacy; those determinations have their own legal pedigree, and are law on that account.

One might object to this argument in the following way. The specification of the prohibition in adjudication is said to return the prohibition to the level of specificity at which people make the judgments that are then generalized in the form of the prohibition. This is inconsistent with Raz's account of authority, because it is a return to the individual reasons that precede the dictates of authority. In other words, specification

deprives authority of its peremptory force. If the determination of fault involves specification, then it is not a legal determination, but a moral one.

Similarly, one might object to the specification account of fault on the matter of pedigree. If specification is a return to the moral judgments of individuals, then a determination of fault, like any moral judgment of an individual, lacks any legal institutional pedigree. At best, fault on this account is a moral determination that the law recognizes for purposes of criminal adjudication. Because this account of fault is inconsistent with the sources thesis, according to the exclusive legal positivist, fault on this account is a moral determination, not a legal one.

Both of these objections misunderstand the nature of specification in the adjudicative determination of fault. The process of specification is not a revival or a return to the reasons and judgments of the accused, and only a revival of or a return to those reasons could support the objections made above. The specification of the prohibition in the determination of fault is a consideration of facts; that is, of the circumstances and manner of commission of the alleged offense under adjudication. These facts are compared to the similarly specific facts pertaining to the choices that were generalized into the prohibition. These facts, though they were reasons for the individuals who acted on them pre-prohibition, are, at the point of adjudication, neither the reasons of the accused and nor the reasons of the adjudicating body. The adjudicative process in which specification occurs does not adduce the reasons on which the accused relied in deciding on the action that is alleged to constitute an offense. But only a return to or a revival of these reasons would undermine authority, and only these reasons lack a legal pedigree.

V. Conclusion

The argument of this paper is agnostic on the question of which school of jurisprudence is correct. The argument is that punishment theory should not be mistaken for a branch of moral theory. Instead, it should be recognized and conducted as a branch of legal theory, of jurisprudence. That it is not so regarded is understandable given how punishment theory is usually conducted. The basic theoretical value of clarity seems to me to be a sufficient reason to approach punishment theory differently; to abjure the pervasive practice of moving uncritically between moral categories and concerns and legal categories and concerns. For various reasons indicated in the course of this paper, conducting punishment theory as jurisprudence would not eliminate moral categories and concerns, or the concepts and methods of moral theory, from punishment theory. It would simply require us to employ morality more precisely.

However, while this paper does not attempt to adjudicate the disputes of jurisprudence, I myself am persuaded, for now, of the truth of exclusive legal positivism. As I indicated in the Introduction, I have always found implausible the notion that the jury makes a moral determination in the process of adjudicating guilt – especially if the supposedly moral determination is a singular and absolute prerequisite for criminal liability. Inclusive legal positivism could, in principle, explain such a process, but I am, at present, unpersuaded that this actually could be done. It seems to me, instead, that the implausibility of the accepted view of criminal fault is a standing refutation of any incorporationist account of fault.

Finally, one important effect of conducting the theory of legal punishment as jurisprudence harks back to my opening paragraph. If we conduct punishment theory as legal theory, then we can drop the baggage of moral theory. Moral theories may be no more contentious than jurisprudential theories, but there is no reason for punishment theory to carry the weight of both sets of controversy. For example, there is much in virtue ethics that is enormously valuable in punishment theory, including its account of deliberation on ends and a related responsibility for the state of one's ends. Virtue ethics' well developed doctrines of moral particularism seem especially helpful in properly understanding criminal fault. On the other hand, virtue ethics often appeals to ideas of human nature and the necessary features of human society that are both more contested and less useful to punishment theory. If we conduct punishment theory as jurisprudence, it is easier to justify a cafeteria approach to the insights of moral theory. One side effect of this – perhaps superficial and perhaps not – would be to override the usual categorization of punishment theories. The categories of deontological and consequentialist theories of punishment could be seen as no more precise than their own misnomers, the retributive and deterrence theories. There is no reason to think that moral theory as a whole cannot be selectively sampled for insights into legal punishment. The first step, however, is to see that legal punishment should be studied as law.