THADDEUS METZ

CENSURE THEORY AND INTUITIONS ABOUT PUNISHMENT

ABSTRACT. Many philosophers and laypeople have the following two intuitions about legal punishment: the state has a pro tanto moral reason to punish all those guilty of breaking a just law and to do so in proportion to their guilt. Accepting that there can be overriding considerations not to punish all the guilty in proportion to their guilt, many philosophers still consider it a strike against any theory if it does not imply that there is always a supportive moral reason to do so. In this paper, I demonstrate that censure theory accounts for these intuitions much better than any other theory, including forms of retributivism such as desert theory and fairness theory, and explain why censure theory is able to do so.

KEY WORDS: punishment, retributivism, expressivism, censure, proportionality

I. INTRODUCTION

Many people have the following two intuitions about legal punishment: the state has a pro tanto moral reason to punish all the guilty and to do so in proportion to their guilt. The pro tanto qualification is important; any reasonable theory of legal punishment must not only acknowledge that actually punishing all the guilty would have intolerable costs, but also make room for mercy and forgiveness. Accepting that there are overriding considerations not to punish all the guilty in proportion to their guilt, many philosophers still consider it a strike against any theory if it does not imply that there is always a supportive moral reason to do so.

My thesis is that censure theory, suitably understood, accounts for these intuitions better than any other theory of punishment, retributivist or otherwise. Censure theory maintains that the political community has a pro tanto duty to censure injustice in proportion to the injustice done and that this must be done through proportional punishment. It is a relatively undeveloped theory of punishment,

having been substantively articulated and adopted only in the last 15 years.[1] Even those sympathetic to censure theory have failed to note its striking advantage of best cohering with firm intuitions about legal punishment.

I acknowledge that some utilitarians (and other non-retributive punishment theorists) may not find these intuitions as powerful as retributivists. It was Kant, after all, who grounded his theory on the judgment that a murderer should be punished even if the results of not punishing would be better.[2] Still, utilitarians typically do take such judgments seriously, and they should. They may not dismiss moral intuitions as conservative, cultural biases while simultaneously invoking intuitions about a maximizing practical rationality; considered judgments about morality are on par with those about rationality. It has been quite appropriate for utilitarians to work so hard over the last 30 years responding to the objection that their theory counterintuitively justifies punishment of the innocent. The following are, by the same token, relevant criticisms: utilitarianism incorrectly provides no pro tanto moral reason to punish all the guilty or to do so in proportion to their guilt. If some utilitarians continue to find appealing to these intuitions to beg the question, then they may read this essay as working primarily to develop the most promising form of retributivism.

---

[1] Central defenses of censure theory include: Joel Feinberg, "The Expressive Function of Punishment," repr. in *Doing and Deserving* (Princeton: Princeton University Press, 1970), pp. 95–118; Anthony Duff, *Trial and Punishments* (New York: Cambridge University Press, 1986), esp. ch. 9; Margaret Falls, "Retribution, Reciprocity, and Respect for Persons," *Law and Philosophy* 6 (1987): 25–51; Igor Primoratz, "Punishment as Language," *Philosophy* 64 (1989): 187–205; Jean Hampton, "The Retributive Idea," in Jean Hampton and Jeffrie Murphy, *Forgiveness and Mercy* (New York: Cambridge University Press, 1988), ch. 4, "An Expressive Theory of Retribution," in Wesley Cragg, ed., *Retributivism and Its Critics* (Stuttgart: F. Steiner Verlag, 1992), pp. 1–25, and "Correcting Harms Versus Righting Wrongs: The Goal of Retribution," *UCLA Law Review* 39 (1992): 1659–1702; Andrew von Hirsch, "Proportionality in the Philosophy of Punishment: From 'Why Punish?' to 'How Much?'" *Criminal Law Forum* 1 (1990): 259–290, and *Censure and Sanctions* (Oxford: Oxford University Press, 1993); John Kleinig, "Punishment and Moral Seriousness," *Israel Law Review* 25 (1991): 401–421.

[2] See Immanuel Kant, *The Metaphysics of Morals* Mary Gregor, trans. (New York: Cambridge University Press, 1991), pp. 142–143.

In the following, I will first describe censure theory and explain what I mean when I say that it best accounts for the intuitions that there is always some moral reason for the state to punish the guilty and to do so in proportion to their guilt (II). Then I will consider the ability of "forward-looking" theories, which justify legal punishment in terms of results to be achieved, to account for these two intuitions (III). I will explore several versions of forward-looking theory which are less familiar than utilitarianism but which promise to accommodate these judgments better than it does. However, I will end up contending that no version of forward-looking theory accommodates these judgments very well. Next, I will demonstrate that rival retributive or "backward-looking" theories such as fairness theory and desert theory also fail to do a good job of accounting for these intuitions (IV). In the following section, I will demonstrate that censure theory betters its competitors on this score, and I will provide the deep explanation of what enables it to do so (V). I will conclude by noting the major problems with censure theory which, in light of censure theory's promise, I believe warrant being addressed fully in another context (VI).

II. AN OUTLINE OF CENSURE THEORY

Censure theory, in the form considered here, holds that the government must punish in order to fulfill its duty to denounce injustice. More specifically, the political community has a pro tanto moral obligation to punish in proportion to the degree of injustice because only thereby can it discharge its obligation to denounce injustice in proportion to the degree of injustice. We of course may wonder whether it is true that the state must censure injustice or that censuring injustice requires punishment. However, in this essay I grant these claims to the censure theorist, in order to see how well her theory fares on other grounds. In other words, I concede to the censure theorist her account of *why* legal punishment is justified, in order to test her theory's ability to accommodate intuitions about *for whom* and *how much* legal punishment is justified.

"Injustice" includes at least the breaking of any just law, the content of which may be left open here. Censure theory is compatible with different notions of the content of justice (e.g., liberal,

conservative, libertarian, or egalitarian conceptions), and also with different spatial and temporal scopes of justice (domestic, international, intragenerational, or intergenerational). Similarly, "political community" or "government" may pick out any legal system, broadly construed, e.g., the leaders of a small town, the legislature of a nation-state, or the general assembly of the United Nations.

What does it mean for a political community to "censure" or "denounce"[3] an agent for having been unjust? I believe that both supporters and detractors of censure theory often characterize its central idea more narrowly than it needs to be. Specifically, censure is frequently (mis)construed as *essentially* a matter of *communicating something about the censured to the censured.* Gricean distinctions are often invoked to characterize censure's supposed informative nature. Hence, theorists often deem censure inherently to involve conditions in which (a) one agent transmits a symbol to another agent with the intention that she both understand a certain negative proposition about herself and recognize that the agent transmitting the symbol intends her to understand that proposition by means of the transmission, and (b) the other agent in fact understands that proposition by means of the transmitted symbol and recognizes that the transmitting agent intends her to understand that proposition by means of the transmission. On this model, agent X denounces agent Y (roughly) insofar as agent X intends to educate agent Y about X's reproachful view of Y and succeeds.

While censure can take this form, it may plausibly be understood to include a broader array of conditions, or at least the way I will use the term "censure" (and "denounce") is less restrictive. First off, I maintain that censure of a person need not involve communication *with the person censured*. It is possible to denounce someone without her understanding any proposition or even without transmitting a symbol to her with the intention of her understanding a proposition. For example, one can denounce the dead. Censure has probably been identified as a way to convey a message to

---

[3] Roger Wertheimer has impressed on me the need to distinguish between condemnation, which by definition involves imposing a harm, and censure, which does not. Therefore, I do not take the word "condemnation" to indicate the same idea as "censure." See his "Understanding Retribution," *Criminal Justice Ethics* 2 (1983): 19–38.

the censured because censure in the criminal justice system often involves such a relation. Theorists imagine a judge sentencing someone to a certain punishment in the course of telling him how seriously he has flouted the community's norms. I do not, however, suppose that the concept of censuring a person requires communicating with the censured.

At this point, one could still think that a necessary condition of censure is that someone other than the censuring agent, but not necessarily the person censured, understands a proposition by means of a symbol that has been sent to her with that intention. In order to denounce the dead, perhaps one must communicate something about her to someone alive. One might suppose, then, that censure is (roughly) a matter of an agent X intending to educate someone else (Y or Z) about X's reproachful view of Y and succeeding.

But I do not believe that censure should be viewed as necessarily informative. Suppose a victim's family gathers together once a year to spit on the grave of the criminal who killed a relative. This family can be viewed as denouncing the offender but not thereby communicating with anyone. The family members are not communicating with the offender, at least if there is no afterlife and they do not believe in one. Nor are they communicating with one another. Each member of the family already knows what the others think about the criminal; it would be pointless to spit on the criminal's grave with the intention of sending a message to other family members about one's view of him. Instead, we can understand this instance of censure as a matter of *expressing* disapproval but not *communicating* disapproval.

Another reason that I do not consider censure to be inherently communicative comes from reflection on the symbolic nature of communication. Communication is a matter of conveying a message by means of conventional representations. Censuring by means of the silent treatment, withdrawal of privileges, and imposition of harms is not obviously to use *symbols* that denote disapproval. Instead, these behaviors might be *non-symbolic expressions* of disapproval. Consider an analogy with expressions of disrespect. To kill a stranger for fun and to lacerate oneself out of self-hatred are two ways to express disrespect toward a person that are not communicative. In murdering another person, one expresses the judgment

that the other is worth less than oneself or less than one's pleasure, but without sending conventional representations to anyone with the intention of conveying a proposition. By injuring oneself behind closed doors, one likewise shows disrespect for oneself without using symbols to transmit any message. Similarly, I propose that making someone pay a fine or serve time in jail could be ways of expressing disapproval toward someone without using conventional representations.

In sum, against the narrow understanding of censure in terms of communication with the censured party, I suggest that it is possible to censure someone without communicating *with her* or even without *communicating* at all. I therefore construe censure (or denunciation) as follows: an expression of disapproval of someone for a wrong perceived to have been done by her. Now, I am not sure whether a normal speaker's sense of "censure" allows for this possibility, or whether I am stipulating to some degree. It makes no difference. The point is that, once one allows that the state has a duty to "censure" injustice and that it can do so in a non-communicative form (or once one agrees that the state has a duty to express disapproval of injustice with punishment), one has a theory which uniquely matches central intuitions about legal punishment. If the reader is firm in thinking that "censure" by definition involves communication, then she may instead call the theory I am sketching here "expressivism."

Some theorists hold that the *concept*, as opposed to the justification, of legal punishment includes the idea of censure.[4] That is, some maintain that "punishment" by definition involves expressing disapproval for a perceived wrong having been done. I do not commit myself to this analysis of legal punishment. For my purposes, it will suffice to adopt the following notion of legal punishment: the state punishes insofar as its officials intentionally impose hard treatment upon an individual consequent to an unjust act apparently having been done. As I understand censure theory, it maintains that legal punishment, so construed, is justified because needed for the state to discharge a duty to express disapproval of those who have in fact been unjust.

---

[4] See especially Feinberg, op. cit.

Censure theory is a retributive theory in the broad sense that the moral reason for punishing does not involve the promotion of any results. Instead of justifying punishment by "looking forward" to the production of virtue or happiness (or the reduction of vice and misery), censure theory "looks backward," justifying punishment as itself a morally right response to past behavior. One could also call this theory "intrinsic expressivism,"[5] to highlight the notion that the expression of disapproval is considered ethically warranted apart from any good consequences it might produce.

Although censure theory is not result oriented, most of its defenders do not consider the expression of disapproval *as such* to be an end-in-itself. That is, censure theory proponents typically claim that the duty to censure follows from other duties and that these latter duties are the ultimate ground for punishing. Specifically, the fundamental duties which have been thought to necessitate censure are the following: *standing up for justice, affirming the value of victims, and treating offenders as responsible.*

First, some censure theorists hold that the political community can discharge an obligation to disavow the violation of just laws only by denouncing such behavior. The idea is that the political community must distance itself from injustice and place itself on the side of justice. If the government did not censure injustice, then it would betray the value of justice. For example, after the Holocaust, it would have been wrong for nation-states not to express "Never again." By censuring Nazi leaders during the Nuremberg trials, many countries fulfilled their duty to stand up for international human rights standards.

A second duty thought to require censure is affirmation of the worth of the victim. The political community would degrade the victim of crime, if it did not denounce the wrong done to her. Not expressing disapproval of the unjust action would fail to treat the victim as worth taking seriously. To return to the Holocaust example, had the world community not censured the behavior of Nazi elites at Nuremberg, it would have been a slap in the face of Nazi victims.

A third obligation of the state which has been deemed to require censure is the duty to treat offenders as responsible for

---

[5] Primoratz, op. cit., uses this phrase.

their behavior. Most strongly, some contend that not to denounce offenders would be to respond to their actions as though they were mere events and so to treat the offenders as animals rather than agents. Without censure, Nazi headmen would not have been fully treated as persons capable of moral deliberation and action who chose not to exercise these capacities.

To sum up: if one thinks that the state has obligations to stand up for justice, to affirm the worth of victims, and to treat offenders as responsible, then one will be sympathetic to the claim that the state must censure people for breaking just laws. Of course, one might deny that the state has such obligations. One could also deny that these obligations entail an obligation to censure (or an obligation to censure by means of punishment). However, I am not concerned to provide a justification of the claim that the state must denounce injustice (with punishment); that is quite beyond the scope of this essay. I am merely pointing out what censure theory involves, in order to differentiate it from rival theories and ultimately to show that it accounts for commonsensical judgments about legal punishment better than any of them.

When I say censure theory "best accounts" for intuitions, I mean two things. First, better than any other theory of legal punishment, censure theory *shows that* the intuitions are true. Censure theory easily entails that the state has a pro tanto obligation to punish all those who have unjustifiably broken a just law and to do so in proportion to the seriousness of the violation. Second, better than its rivals, censure theory *explains why* these intuitions are true. Censure theory provides the most attractive rationale for the state's having some moral reason to punish all the guilty in proportion to their guilt. I will make these claims plausible in the next three sections.

III. FORWARD-LOOKING THEORIES

Forward-looking theories, which hold legal punishment to be permissible only if and because certain long-term results are forth-coming, are well known for being unable to show that there is some moral reason to punish all the guilty; for there can be situations in which punishing the guilty produces worse consequences than not punishing.

First off, there are theories such as utilitarianism and moral education theory, which make the justification of legal punishment contingent upon the (expected) realization of intrinsic value.[6] In cases where punishing fails to produce at least as much happiness or virtue as not punishing, these theories fail to show that there is pro tanto moral reason for the state to punish a guilty person. It seems quite possible that the costs of punishment could reduce its overall benefit to the point where some other response to an offender is optimific.

A natural first response is to appeal to rule-utilitarianism, in which acts of punishment are evaluated in terms of whether they conform to rules which maximally promote welfare when generally followed. However, this theory also faces the problem of counter-intuitively implying that there is sometimes no moral reason to punish the guilty. This is clearest in the case of a racist society. In such a context, the following rule could well maximize happiness: punish the guilty except when the guilty have harmed a person who belongs to a small and despised minority. Although following such a rule would greatly upset the minority, the happiness of the majority could substantially outweigh it.

As a second response, the forward-looking theorist might search for a way to discount the happiness of majorities. For example, she might specify virtue as the intrinsic value to be sought when considering what would happen if everyone performed an act. So, consider a rule version of moral education theory, according to which legal punishment of an individual is justified only if and because it is allowed by a rule which, if generally followed, would maximize moral reform. So far as I know, no one has advocated this theory, but it is worth considering it here. Since this theory does not take happiness into account, it appears to avoid the implication that a guilty member of a majority should not be punished in a racist society.

However, one can well imagine a situation in which the racism is so entrenched that little or no moral reform results from punishing

---

[6] For a recent defense of the utilitarian theory of punishment, see J. J. C. Smart, "Utilitarianism and Punishment," *Israel Law Review* 25 (1991): 360–375. For good statements of moral education theory, see Herbert Morris, "A Paternalistic Theory of Punishment," *American Philosophical Quarterly* 18 (1981): 263–271; and Jean Hampton, "The Moral Education Theory of Punishment," *Philosophy and Public Affairs* 13 (1984): 208–238.

members of the majority group. Suppose that, instead of reflecting on their behavior, majority members tend to backlash whenever one of their group is punished for harming a minority member. Imagine that majority members on balance respond much better to a judge's verbal instruction. In this case, the rule variant of moral education theory would provide no moral reason to punish the guilty.

Let us consider a different strategy on behalf of the forward-looking theorist. Specifically, "scalar consequentialism"[7] promises better to accommodate the intuition that there is always pro tanto reason for the state to punish a guilty individual. According to this theory, the more good consequences that an act has, the more moral reason there is to perform it. The difference between scalar consequentialism and the usual variety is that moral reasons to act are not tied exclusively to a certain (maximizing or satisficing) sum of good consequences. The scalar theory says that any time an act produces some good (or reduces some bad), there is a pro tanto moral reason to do it. A scalar consequentialist account of legal punishment would be this: the more good consequences that legal punishment has, the more moral reason there is to impose it. This theory does not make the rightness of punishment contingent on a particular aggregate of intrinsic value – indeed, it does not even invoke the concept of rightness. So long as punishment of the guilty will produce *some* degree of good results such as virtue or happiness, scalar consequentialism says that there is pro tanto moral reason to punish.

Scalar consequentialism, too, turns out to have counterintuitive implications about punishment of the guilty. One can imagine a case in which punishment of the guilty produces absolutely no good results (and does not reduce any bad). Suppose an innocent, friendless drifter wanders into a community in which everyone hates him for being different. A member of this community kills him, and the community keeps this fact to itself. Everyone is happy that this individual was killed, and so no one wants the killer to be punished. Here, punishing the killer will not make anyone any happier (or, let us suppose, any more virtuous). But surely the murderer should be

---

[7] Frances Howard-Snyder has done the most to develop this form of consequentialism. See, for example, "The Heart of Consequentialism," *Philosophical Studies* 76 (1994): 107–129.

punished. If so, then scalar consequentialism cannot show that there is pro tanto moral reason to punish all the guilty. Even though scalar consequentialism does not tie the permissibility of punishment to a specific amount of good results, there still must be some good results for punishment to be morally rational. This counterexample shows that invidious majoritarianism still plagues the scalar theory; sometimes punishment of the guilty has no good results, and in such a case the scalar theory implies that punishment is not morally rational.

Now, I must admit that there is one idiosyncratic form of scalar consequentialism which does entail that there is always pro tanto moral reason to punish the guilty, at least if the censure theory entails this. Such a scalar consequentialism would hold censure itself to be an intrinsic value. Specifically, the theory would say this: the more that a state censures guilty parties in proportion to their guilt, the more moral reason there is for a state to punish.[8] If censure theory is correct that proportionate censure of the guilty just is a matter of proportionate punishment, then deeming proportionate censure to be an intrinsic good that should be promoted will entail that there is always pro tanto moral reason for the state to punish the guilty.

There are two reasons why even this artificial form of consequentialism does not account well for the intuitions at hand. First, scalar theories in general do not provide sufficient normative guidance. A full-fledged theory of legal punishment should say not only why punishment is permissible but also when. However, scalar consequentialism does not even indicate a necessary condition for the permissibility of legal punishment. I am therefore inclined to think that it does not even count as a theory of legal punishment which could rival censure theory.

Even if scalar consequentialism in general counts as a genuine competitor to censure theory, a particular scalar consequentialism that deems censure to be an intrinsic good does worse than censure theory in explaining why punishment of all the guilty is justified. Censure theory says that censure is right and that punishment is

---

[8] Compare the "consequentialist retributivism" discussed in Michael Moore, *Placing Blame: A General Theory of Criminal Law* (New York: Oxford University Press, 1997).

therefore permissible, whereas the scalar consequentialism under consideration says that censure is intrinsically good and so should be promoted. Censure theory is more plausible in this respect since the scalar view oddly implies the following: it would be a good thing for people to break the law, since states would then be in a position to censure them and thereby promote intrinsic value. This scalar theory instructs people to commit crime, since only by that means can the state be in a position to realize the intrinsic worth of censuring the guilty. In short, even if a scalar consequentialism instructs us to promote censure and thereby shows that there is moral reason to punish all the guilty, it provides a deficient rationale for doing so.

There are forward-looking theories which differ from those considered so far in that they do not justify legal punishment in terms of results which have intrinsic value. The relevant results are instead those involved with the exercise of rights. For example, restitution theory says that punishment is allowed only when and because it will fulfill a victim's right to be compensated by, say, deterring crime and thereby reducing fear of attack.[9] And self-defense theory says that the innocent have the right to carry out threats of punishment to the guilty only when and because it will deter crime and thereby provide protection.[10]

These rights-based versions of forward-looking theory have the same problem as the intrinsic value versions. There will be circumstances in which punishment by the state simply will not compensate and will not protect, and in such cases these theories cannot entail that there is some moral reason for the state to punish someone who has unjustifiably violated a just law.

The factors making it difficult for forward-looking theories to show that every guilty person is a proper candidate for legal punishment also make it hard for them to demonstrate that the guilty should receive the proper degree of punishment. In particular, these theories have difficulty showing that all those who have committed serious crimes pro tanto warrant receiving a severe punishment.

---

[9] The best defense of restitution theory is Margaret Holmgren's "Punishment as Restitution: The Rights of the Community," *Criminal Justice Ethics* 2 (1983): 36–49.

[10] See, e.g., Daniel Farrell, "The Justification of Deterrent Violence," *Ethics* 100 (1990): 301–317; and Phillip Montague, *Punishment as Societal Defense* (Lanham, MD: Rowman and Littlefield, 1995).

This should be clear, given the previous argumentation. Utilitarianism cannot prescribe a substantial penalty for someone guilty of a serious crime, if better results are forthcoming with a lesser penalty. Similar remarks apply to moral education theory, restitution theory, and self-protection theory.

IV. BACKWARD-LOOKING THEORIES

The requirement that legal punishment generate certain results makes it hard for forward-looking theories to account for the idea that the state has some moral reason to punish all the guilty in proportion to their guilt. The most promising response to make on behalf of the forward-looking theories is to advocate "mixed" theories which combine forward- and backward-looking elements. A backward-looking theory does not conceive of the moral reasons for punishing in terms of any condition that will obtain in the long-run. Rather, a backward-looking theory holds that state punishment is justified because it is a right response to the fact of guilt. Fairness theory and desert theory are the most influential versions of backward-looking theory. One might think that either one of these theories, or a forward-looking theory combined with one of them, could account for the relevant intuitions.

However, neither fairness theory, desert theory, nor any forward-looking theory combined with them, can well account for the judgments that the state has some moral reason to punish all the guilty and to do so in proportion to their guilt. After demonstrating the inability of fairness theory and desert theory to accommodate these intuitions, I will bring out the reason for their failure in this respect. I will note that the problem is not with backward-looking theory as such, for censure theory is, I will argue in the next section, the backward-looking theory which can meet the challenge posed here.

Consider fairness theory, which holds that legal punishment is justified only when and because it will restore an equitable balance of burdens and benefits among citizens.[11] Fairness theory conceives

---

[11] Prominent defenders include George Sher, *Desert* (Princeton: Princeton University Press, 1987), ch. 5; and Michael Davis, *To Make the Punishment Fit the Crime* (Boulder: Westview Press, 1992).

of the maintenance of government as a cooperative scheme in which certain benefits go to society as a whole if and only if some subset of the total population undergoes certain burdens. The primary benefit of government is an orderly society, but one could also include benefits such as health care, education, and so forth. Obedience to law is one important burden that citizens must bear in order to keep a government functioning; imagine the chaos that would ensue if everyone routinely disregarded the law. Now, if a person receives the benefits of government without undertaking the burden of obedience necessary to generate these benefits, then she takes a "free ride" at the expense of law-abiding citizens. Punishment restricts the liberty of the guilty, so that they no longer get the benefits of government without the cost needed to produce them. The greater the liberty taken in breaking the law, the more punishment that is needed to correct the unfairness.

It might be worth pausing to compare fairness theory and censure theory. While neither of these theories makes the legitimacy of state punishment contingent upon results obtaining, they differ with regard to legal punishment's function. According to fairness theory, the point of state punishment is to correct exploitation on the part of the offender. One who violates the law has obtained a benefit at the expense of law-abiding citizens, and state punishment takes the benefit away, so that removal of the unfairness is concomitant with the imposition of punishment. In contrast, for censure theory, the fundamental ends which are concurrently realized with legal punishment are expression of support for justice and for the victim as well as treatment of offenders as responsible for their behavior.

Censure theory can accept fairness theory's contention that breaking the law often involves a form of exploitation; hence, it may be that one way of expressing support for justice is to remove an advantage which offenders have unfairly obtained. However, censure theory need not view the ends of punishment solely in terms of the rectification of exploitation. For instance, what makes murder warrant a stringent penalty, according to fairness theory, is just that the offender has taken a great liberty at the expense of *those who have not murdered*, a liberty that the state must remove only with a severe penalty. But censure theory can maintain that stringent

punishment is required in order to affirm the worth of *the murdered party*, or to treat *the offender* as a morally responsible agent.

These differences between fairness theory and censure theory entail the further differential ability of them to account for the notion that all the unjust are proper candidates for legal punishment. There are at least two situations in which punishing the guilty will not restore a fair distribution of benefits and burdens; in these situations the guilty party has not received any unfair advantage from breaking the law. First, imagine that a private citizen from another country (or, more fantastically, from another planet) intentionally drops a bomb on New York City from an aircraft. Since this alien has not received the benefits of American government and, hence, has not exploited fellow citizens in breaking the law, fairness theory fails to justify her punishment. But surely punishment is justified for such a breach of the law against murder.

Here is a second scenario in which it is possible to break a just law without taking unfair advantage of law-abiding citizens. Suppose a person is convicted of crime X when he is in fact innocent of having committed any crime. Imagine that he receives Y years of jail, which is whatever amount of punishment fairness theory prescribes for crime X. Finally, suppose that, after serving Y years, the person then commits crime X. Fairness theory cannot prescribe state punishment of this person for actually having committed crime X, since the person already unfairly lost Y liberty which fits crime X. Of course, sympathy and compensation are due to the wrongly convicted person. However, we surely do not accept that the state has no moral reason to punish him for committing the crime, which fairness theory entails.

The justification of legal punishment solely in terms of rectifying unfair advantage means that fairness theory must forbid punishment of someone who breaks a just law when he receives no benefit at the expense of others or when he has unfairly received great burdens in the past. I conclude that fairness theory fails to show that the state has some moral reason to punish all the guilty. Let us address the other major version of backward-looking theory, namely, desert theory.

For an agent to deserve something is for her to warrant, because of some personal feature that she manifested in the past, some-

thing of the same kind and in the same proportion as that feature.[12] The desert theory of punishment, then, maintains that the state may punish a person only if and because he warrants hard treatment due to, and in proportion to, his past unjust behavior. So construed, desert theory and censure theory differ in that censure theory essentially prescribes an expression of disapproval on behalf of the political community, whereas it would appear that the state's giving someone the hard treatment she deserves is not necessarily a matter of expressing anything.

One might wonder whether the two theories collapse if one conceives of punishment as inherently constituting a form of censure (see section II); then it would the case that one deserves censorious punishment. However, there would still be an important difference between desert theory and censure theory, when "punishment" is defined as a form of censure. According to censure theory, the fundamental moral reasons for expressing disapproval through punishment are that doing so is a matter of the political community's fulfilling requirements to stand up for justice, affirm the value of victims, and treat offenders as responsible. According to desert theory, in contrast, the basic rationale for punishing (which, by hypothesis, is a form of censure) is that the state is in the best position to give people the negative response proportionate to what they have earned.

Desert theory is amenable of different interpretations, depending on the way one understands "negative response" and "proportionality." Consider desert theory$_1$: legal punishment is justified only if and because the guilty deserve to *suffer* in proportion to the injustice done.

This straightforward version of desert theory cannot entail that there is some moral reason for the state to punish every person who has broken a just law. For example, desert theory$_1$ will not prescribe legal punishment when a guilty person has previously suffered a great harm that was undeserved; legal punishment in

---

[12] With this statement I summarize Don Scheid's thorough discussion of desert in his "Constructing a Theory of Punishment, Desert, and the Distribution of Punishments," *The Canadian Journal of Law and Jurisprudence* 10 (1997): esp. 457–460.

such a case would impose more suffering than is deserved.[13] For instance, suppose that a serial rapist gets hit by a bus, injured but not killed. It is quite natural for desert theorists to think, "He got what he deserved."

Let us modify desert theory so that it corrects this defect. We need a version that will screen out suffering which the guilty have experienced prior to state conviction. So, perhaps the guilty do not deserve just any kind of suffering, but rather legal punishment in particular. Maybe those who have broken a just law deserve suffering to be imposed by a government which intends to make them suffer because they were unjust. Consider, then, desert theory$_2$: legal punishment is justified only if and because the guilty deserve *legal punishment* proportionate to their injustice.[14]

Desert theory$_2$ also fails to solve the problem. Desert theory$_2$ is vulnerable to the wrongful conviction example discussed in the context of fairness theory. If the state accidentally punished an innocent person of a crime which he then committed after his sentence, desert theory$_2$ could not prescribe legal punishment for him. He would have already received the legal punishment proportionate to his injustice, and it would give him more legal punishment than he deserves for the state to punish him after the actual commission of the crime.

Let us revise desert theory one last time. Consider desert theory$_3$: legal punishment is justified only if and because the guilty deserve to receive *legal punishment consequent to their injustice* that is proportionate to this injustice. This, finally, solves the problem. Desert theory$_3$ does entail that there is pro tanto moral reason for the state to punish all the guilty.

However, to account fully for the intuition, desert theory$_3$ must not only entail it but also explain it. *Why* on grounds of desert should the suffering of an offender prior to his sentence not influence whether legal punishment is deserved? And *why* on grounds of desert should the legal punishment of an offender prior to his

---

[13] This objection is clearly developed by Gertrude Ezorsky in the introduction to her edited volume, *Philosophical Perspectives on Punishment* (Albany, NY: State University of New York Press, 1972), pp. xxii–xxvii.

[14] For an influential recent example, see Michael Moore, "The Moral Worth of Retribution," repr. in Joel Feinberg and Hyman Gross, eds., *Philosophy of Law*, 5th edn. (Belmont, CA: Wadsworth, 1995), pp. 632–654.

offense not influence whether legal punishment is deserved after the offense?

As far as I can tell, the defender of desert theory$_3$ can answer these questions only by invoking an institutional theory of desert according to which desert claims are entirely a function of conventions. For an example, consider grades. A grade is plausibly a school-dependent positive response deserved for writing an essay. No grade of "A" can be deserved by someone not enrolled in school, for both what is deserved and the degree of what is deserved for writing an essay seem to depend utterly on the institutional context. The same might go for punishment. Perhaps both what is deserved for doing an evil deed and the degree of what is deserved are government-dependent negative responses. On this view, only a particular kind of harm, namely, legal punishment consequent to the offense, is deserved by those who are unjust, since that is simply what our convention dictates. Any harm suffered prior to the commission of a crime, or prior to sentencing for a crime committed, is not germane to what an offender deserves, since our institutional norms dictate otherwise.

There are two serious problems with this explanation of why the state has pro tanto moral reason to punish all those who have broken just law. First, since conventions differ among societies, the institutional desert theorist turns out to be unable to say that *every* state has some ethical justification for punishing all the guilty.

Second, it seems wrong to think that desert claims about punishment are entirely a function of institutional practice. The institutional desert theory implies that if our convention were to impose a light penalty for murder, there would be no coherent way to hold that this offense actually deserves a greater penalty. But this seems quite counterintuitive. Furthermore, the institutional desert theory implies that when a rapist still at large gets hit by a bus, he does not thereby receive any of what he deserves. But getting hit by a bus *does* seem to be a way for someone who has seriously harmed others to get some of what he deserves. Likewise, getting accidentally punished by the state before an offense *does* seem relevant to fixing a proportion between the degree of one's offense and the amount of legal punishment one deserves. Our considered judgments about desert claims suggest that legal punish-

ment consequent to the offense is not the only negative response which can be deserved for injustice; suffering after an offense but prior to sentencing, as well as punishment prior to an offense, can apparently figure into desert proportional to an offense. I conclude that, even if desert theory$_3$ entails that there is pro tanto moral reason to punish all the guilty, it cannot provide an adequate explanation of why this is so. And an acceptable account of the intuition must both entail it and explain it well.

As with the forward-looking theories, desert and fairness theory have the further difficulty of not entailing that all the guilty should pro tanto be punished in proportion to the gravity of their offenses. Fairness theory must enjoin a small penalty for a serious harm, if the offender was wrongly punished to a large degree by the state in the past or if the benefits received from the state were not substantial. And desert theory, in its natural formulation, recommends a light penalty for a heinous crime, if, for example, the guilty person had previously suffered greatly from natural causes or unjust legal punishment.

Backward-looking theories initially appear to be more promising than forward-looking theories, since they do not make legal punishment's permissibility depend on results and since they intrinsically include accounts of proportionality. However, the problem common to fairness theory and desert theory is that they make the justification of legal punishment *depend on its being proportional to something beyond guilt which obtained in the past*. Fairness theory justifies punishment only if it corrects unfair advantage, where unfair advantage is a function of benefits and harms one has received in the past. Similarly, on the straightforward understanding of desert theory, punishment is justified only if it matches the amount of suffering (or punishment) one has coming when taking into account past suffering (or punishment). We need a backward-looking theory that makes the justification of punishment depend merely on its being proportional to the degree of guilt and not on any past benefits or burdens that an offender has had. Censure theory does exactly this.

V. CENSURE THEORY'S ADVANTAGE

Censure theory best entails that and explains why there is pro tanto moral reason for the state to punish all of the guilty in proportion to their guilt. Censure theory says that the political community has a moral reason to punish all those who have broken a just law in proportion to their injustice since it has a pro tanto obligation to express disapproval in proportion to the degree of injustice. Granting the assumption that proportionate disapproval is a matter of proportionate punishment (as I do in this paper), censure theory naturally accommodates the intuitions about legal punishment addressed here.

Censure theory does not justify legal punishment in terms of results to be produced, as the forward-looking theories do. Hence, it does not hold punishment of the guilty hostage to good fortune, which may not be forthcoming. Censure theory also does not tie the permissibility of punishment to the degree to which the offender has been benefited/burdened in the past, as both fairness theory and desert theory (on its natural interpretation) do. The fact that legal punishment will have bad results, or that the state has wrongfully convicted the offender in the past, does not vitiate the state's pro tanto duty to express disapproval (by means of punishment) of those who break just laws.

I have indicated in sections III and IV why the forward-looking and rival backward-looking theories have respectively failed. I will now indicate the fundamental reason why none of these theories can account for the idea that the state has some moral reason to punish all the guilty in proportion to their guilt. The basic problem with the forward-looking theories and rival backward-looking theories is this: they make the moral justification of legal punishment *contingent upon achieving a certain state of affairs which is metaphysically distinct from the mere imposition of punishment proportionate to the seriousness of the injustice*. The forward-looking theories require legal punishment *to promote benefits in the future*, whereas the rival backward-looking theories require punishment *to realize a balance between hard treatment and benefits/burdens the offender had in the past*.

Censure theory, in contrast, does not make the justification punishment contingent upon the realization of any property meta-

physically distinct from a proportionate punishment of the guilty. Censure theory requires legal punishment merely to express disapproval in proportion to the seriousness of the offense. Since all the guilty are proper candidates for proportionate censure from the political community, which, by hypothesis, must take the form of punishment, all the guilty are proper candidates for proportionate legal punishment. Censure theory holds that proportionate punishment *just is* the right way for the political community to express proportionate disapproval of the guilty, where this expression *just is* the right way for the state to distance itself from injustice, to stand up for victims, and to treat offenders as responsible. Again, disavowing injustice, treating victims as valuable, and responding to offenders as moral agents are constituted by the state censuring injustice proportionately, which, in turn, is constituted by proportionate legal punishment. These property identities are what enable censure theory best to entail and explain the intuitions that the state has a pro tanto obligation to punish all the guilty and to do so in proportion to their guilt.

VI. CONCLUSION: CENSURE THEORY'S WEAKNESS

Throughout this discussion, I have been granting the censure theorist three important claims: (1) the state has duties to disavow injustice, to stand up for victims, and to treat offenders as responsible; (2) fulfilling these duties is a matter of censuring offenders proportionately to their offense; and (3) proportionate censure is a matter of proportionate punishment. If these claims are true, then, I have argued, censure theory best accounts for the idea that the state has some moral reason to punish all the guilty in proportion to their guilt. However, we cannot assume that these claims are true, and in fact many question their truth.[15] Proponents of censure theory have yet to defend these three claims thoroughly or convincingly.

If the arguments made here are sound, then censure theory can be said to answer well the questions of *whom to punish* and *how*

---

[15] The *locus classicus* for doubt about censure theory's basic claims is H. L. A. Hart, *Law, Liberty, and Morality* (Stanford: Stanford University Press, 1963), pp. 65–66. Many others have since echoed Hart's criticisms, to which I respond in an unpublished manuscript, "Why the State Must Censure with Punishment."

much to punish. However, it has yet to answer adequately the question of *why punish*. It has not been clearly demonstrated that the political community has some moral reason to express disapproval of injustice by means of hard treatment. That one can best accommodate considered judgments about whom to punish and how much only by supposing there is such moral reason is some evidence that there in fact is. However, much more argument needs to be made to establish that the state has a pro tanto obligation to censure injustice with punishment. Censure theory's advantages with respect to the issues of whom is liable for legal punishment and to what degree make it worthwhile exploring elsewhere whether there is such an obligation.[16]

Department of Philosophy
University of Missouri – St. Louis
8001 Natural Bridge RD
St. Louis, MO 63121
USA