

John Hanson
February 28, 2008
Philosophy Defense
Prof. Tamler Sommers

Persons, Prisons, and Retribution:
Human Limits and Morris's Legal Framework

In his classic article *Persons and Punishment*, Herbert Morris argues against what he calls the therapeutic approach to punishment, suggesting that systems of criminal justice based on consequentialist principles will inevitably violate basic human rights founded in personal autonomy. Being driven solely by a concern with results and based on a view of persons as malleable, endlessly improvable objects, such systems will be inescapably inclined to attempt the kind of personality and behavioral alterations implied by their theory, as well as to restrict fundamental rights in their efforts to reform the convicted. These violations, he suggests, reflect the therapeutic approach's lack of respect for the very idea of justice, morality, and the autonomous person, and arise from the theory's failure to distinguish between reasonable choice and coerced behavior. Against this, he proposes a system in which rights based in the freedom of rational, autonomous agents serve as the foundation for the penal justice system. Rather than reform and treatment, such a system would accord guilt, retribution, and personal autonomy central roles in how punishments are measured and dispensed, hence avoiding the Orwellian extremes implied in consequentialism.

While previous critics of Morris have focused on the moral and technical aspects of his theory, I will argue in this paper that the psychological realities of imprisonment do not affirm

Morris's vision of justice, but instead violate the foundational rights upon which the legitimacy of his system is implicitly based. Because retributivism, in any form remotely resembling our present penal system, must engage in the isolation and containment of human persons, the safety of our selves cannot be ensured against the psychological strains implied in such modes of punishment. The mental pressures endangering those rights come in many forms; for this paper, I will focus upon the impact of prison stress, and on the long term ways in which the experience of prison can affect the personalities and psychological constitutions of the imprisoned.

Ultimately, I will argue, the evidence suggests that the right to a stable, morally autonomous self is violated by the kinds of stresses imposed by imprisonment, seemingly undermining the very thing Morris claims retributivism is built to preserve, namely, the autonomy of the rational person. Thus these data pose a serious, though purely negative, challenge to the legitimacy of Morris's retributive framework as it applies to prisons, and, as I shall sketch out briefly in the counter-arguments section, potentially all forms of retributive punishment.

The paper will proceed in three parts. In the first, I will outline Morris's theory, and the vision of autonomy that he claims retributivism is bound to respect. In the second, I will outline my criticism of this theory and discuss the data that suggests autonomy is not respected by imprisonment. The third section will be concerned to address two objections: first, whether the data I am using is truly representative of what a purely retributivist prison would be like, and second, whether a different system of punishment might not respect the rights that imprisonment does not, thus salvaging retributivism if not imprisonment. I will conclude that neither of these

objections is sufficient to salvage Morris's project, though they raise interesting questions for modes of punishment other than imprisonment which deserve further inquiry

I.) Morris's argument and the notion of autonomy that underlies it.

Persons and Punishment is concerned to defend a vision of criminal justice in which retribution serves as the only morally legitimate ground for punishment, and in such a way that the rights underlying a just system include the right to be punished retributively. In order to establish the desirability of a system with such a seemingly strange feature, Morris lays out two opposing systems of justice and argues that only the one based on retributive principles is adequate to the task.

In the first system, agents rationally enter into systems of mutual benefit and burden by which they can insure their long-term success. This system would be codified in a set of laws, and would define criminality as the unfair renunciation of the burdens that the participants have freely agreed upon. Rectifying this imbalance would be the sole task of punishment, and no more; rehabilitative efforts are strictly out of the question provided the criminal is fully sane, as the only way to commit a crime in this system is to take an unfair action.

Furthermore, this system's legitimacy is grounded in the fact that it respects the rights of those under its power. These rights, which Morris thinks include the right to be punished under a just system, validate the moral claims of the system, and can never be given up, transferred or

disregarded.

In contrast to this rights-based approach, Morris offers a second system whose central ideal is not punishment but therapy. Under this kind of system, the criminal's choice to be criminal would be seen as evidence of a character defect, and hence his treatment would endeavor to change his behavior, rather than punishing him for acting as he did. Nowhere is his willing participation in society taken into account, nor are the absolute rights of the first system even given consideration. Instead, he is seen by the therapist-jailer as a collection of largely predetermined impulses, desires, and deliberations capable of extreme modifications in the name of utility. His actions are demoted to mere events, and the full force of the state's power is brought to his reformation.

Morris suggests that in this fully therapeutic system, "the logic of cure will push us towards forms of therapy that inevitably involve changes in the person made against his will" (42-43). Such violations would "change him and his judgments," (43) in ways that tread on the rights which must underlie any truly just system; they show a lack of respect for our capacity to make our own decisions, degrade the meaning of our achievements, and slowly erode the line separating moral sanity and amoral madness. To neglect these rights and the human traits that accompany them is, in Morris's mind, tantamount to an invalidation of the system itself, since "the [just] system rests on the assumption that spheres of legitimate and illegitimate conduct are to be delimited with regard to the choices made by persons." (500) Absent those boundaries, the choices of those persons lose any legal significance, and the system becomes morally bankrupt.

Implicit in both Morris's criticisms of therapy and his positive account of retributive justice are a number of assumptions that prove to be problematic when applied to retributive systems as they exist in the real world. In particular, his vision of human beings as autonomous, rational persons deserves further discussion, since it is a complex issue on a number of levels, and like many retributivists Morris is exceptionally vague in defining exactly what being an autonomous, rational being means. Drawing from different parts of his paper, two elements of his account stand out as relevant for our discussion, though there are many more which must be ignored for the sake of brevity.

The first of these is that

(1) Autonomous agents are able to reason about morality without outside constraints.

It seems to be Morris's expectation that the agent's moral judgment must not be impinged upon by factors external to the agent, and that their ability to conclude what is right and wrong must in some sense be their own. Concerns about losing this capacity underlie much of Morris's discussion of the therapeutic state. As he writes about the status of moral reasoning under such a system, "not just a person's acts are taken from him, but also his attempt at a reasoned justification for the acts," while under retributivism, "we make him pay the price and we respect his right to retain the judgment he has made" (43). To possess moral judgment in this sense is to be able to reason about one's actions in a way meaningful to their performance. Most of us

presumably have this capacity normally, though a person who has been brainwashed a la Patti Hearst would not, since the process of being traumatically indoctrinated robs one of the capacity to make moral judgments of one's own. To fail to respect this element of autonomy would thus be to fail to respect the rights of the agent, since it undermines so much of what those rights are built to protect.

(2) Autonomous agents are free to shape their actions and their characters as they choose.

The second element of autonomy concerns the agent's capacity to control their actions and the character resulting from them. As Morris states explicitly while comparing the status of animals with (presumably autonomous) humans, "we have thought ourselves capable where animals are not of making, of creating, among other things, ourselves" (42). Failure to recognize this capacity is to cease treating others as autonomous persons: "When we treat a human being merely as an animal or some inanimate object, our responses to the human being are determined, not by his choices, but by ours in disregard or in indifference to his" (46). A therapeutic system would presumably treat criminals in precisely this way, seeking to reform them regardless of what they might want for their lives in much the way we train dogs for subservience regardless of their wishes.¹

¹ A point of clarification on the exact sense of autonomy Morris intends may be in order. It can be asked whether this second condition of autonomy entails that we are free to act however we wish, since our moral origination of our actions would seem to imply that the system can never control us for fear of encroaching on our status as originator. Morris doesn't seem to think this is an issue, apparently assuming there is a distinction between being

6

These two conditions provide the background for Morris's discussion theory of criminal justice, though there are certainly more implicit in his argument. However, even this limited list, as I shall argue in the next section, proves deeply problematic for retributivism once the realities of imprisonment are taken into account.

II.) Rights and the Effects of Prison

My particular concern centers around Morris's claim that systems of therapeutic punishment must inevitably seek to violate the autonomy of those under their supervision. Morris seems to be of the opinion that this is an issue limited to the therapeutic methods he opposes, and much of his argument against consequentialism hinges on the assumption that a system based on culpability and desert rather than reform will not suffer from the same defects. This is, I believe, a profoundly flawed view; both systems must seek to punish those who commit crimes, and hence must resort to specific methods to accomplish this. Yet those methods often advocated by retributivists, if not Morris himself, can result in the violations of autonomy that Morris so fears, and thus violate the rights on which his system is based. There is a substantial, if controversial,

constrained in a way consistent with being the moral originator, versus having one's status as moral originator taken away. We could imagine, for example, being imprisoned but still have one's acts still be one's own, though in new constraints; it would be entirely different to be hypnotized, in which case one's acts could be legitimately attributed to an outside force. Though the bulk of this paper is devoted to showing that being in prison can have such an effect given long enough, for the sake of the argument, I am assuming that Morris intends a distinction between being confined and ceasing to be autonomous along these lines holds.

literature on the effects of prison life on prisoners, much of which points to alterations of the self that, while not Orwellian in their intent and result, seem just as disturbing as those that Morris claims undermine our rights to autonomy under consequentialism.

Given the assumptions about the meaning of autonomy outlined above, we have a fairly specific set of standards by which we can judge the evidence. We will not be looking for mere discomforts or displeasures, since producing these is one of the principle aims of punishment and one that, I think, is fairly reasonable. Rather, I will try to bring forward some examples of effects induced on prisoners by their environments that truly violate the notion of the autonomous self that so much of Morris's theory hinges on, focusing in particular on violations of (1) autonomous moral reasoning, and (2) the ability to freely create our actions and character.

One of the best investigations of this subject can be found in the work of Hans Toch, particularly his 1975 book, *Mosaic of Despair: Mental Breakdowns in Prison*, which explored the psychic trauma experienced by many prisoners during the course of their incarceration. Built primarily around a study done on inmates at New York prison in 1972, Toch's book found that 7.7 percent of young offenders had reported experiencing mental breakdowns, while around 6.5 percent of all adults had had similar experiences when unreported incidents were taken into account. Amongst women, the rate was considerably higher, topping out around 10.8 percent. Such breakdowns typically manifested themselves in the form of self-mutilation by razor blade, hanging, drug use, or fire, and showed varying rates of recurrence (Toch 160-176). The justifications offered by prisoners also varied; for some, the cause was a loss of value in life, for

others an attempt to radically separate themselves from their past failures or to establish their control over their imprisonment. (165-170). While the rates for such breakdowns at lower security prisons were considerably lower, they remain far above those of the general populace, and point to what could be a problem endemic to the system of prison punishment.

This thesis is supported by the fact that in most of the cases Toch investigated, the decisive factor in the breakdown was the loss of autonomy associated with imprisonment. Many inmates pointed to the cumulative stresses of sudden separation from the outside world, the loss of meaningful relationships, and a precipitous decline in self-worth in inciting their episodes of self-mutilation. Speaking of the spiral of doubt that precedes many breakdowns, Toch writes, "His past documents his failures and negates his hopes he deals with others not as an equal but as an inferior. His peers are judges, sponsors, enemies, threats. Love is not for him but for the worthy; as substitutes, he buys loyalty, pity, or at best, restraint." (Toch 50) It is because of his placement in isolation to others and the world, as well as the reinforcing environment of the prison, that the prisoner's concept of self is so endangered.

Having established that such experiences exist in prison systems, it can now be asked whether these effects satisfy the criteria set out above for violations of autonomy. On the one hand, it seems intuitively right that the convicted should be subject to some level of pain and discomfort, and that developing a proper sense of guilt would be an appropriate response to being punished, one not inconsistent being the author of one's character. Yet it also seems that taking such emotions to the point of self-mutilation crosses a line into difficult territory, where

the standards for respecting autonomy are in real danger of being violated. In particular, it seems to me that the second criteria we set out above is challenged in this situation. If Toch's data is correct, the environment of prison has such a profound influence on the patterns of behavior exhibited by inmates that they would do things that they would never do otherwise, such as engaging in self-mutilation. And if autonomy requires that one be morally responsible of one's actions and character, then how can we make sense of behavior falling so far outside the boundaries of normalcy? The blame must be put on either the convict or his environment. It seems unlikely that all of subjects of Toch's study were self mutilators outside of prison, or that that choice was one they might have made regardless of the constraints of their environment, and so the change in behavior must be attributed not to the will of the agent but to the prison environment. This is a clear violation of the right to shape one's actions and character as one chooses.

Although the cases discussed above are extreme (if common), and in many cases rarer than in the maximum detention facilities in Toch's studies, there are any number of other ways in which prisons subtly influence the personalities of those who inhabit them. For example, one of the most consistently observed changes in the outlook of long-term prisoners is the development of "philosophy of minimum expectation," in which prisoners come to expect no positive developments in their situation. (Flanagan 123). Others have pointed to the willful ignorance of long-term considerations that prisoners develop, and the focus on the present moment that most decisions come to reflect (Flanagan 122-123).

These changes, though less shocking than those discussed above, reverberate in the same key, and raise problems for both of the criteria of autonomy. In both cases we have an outside force, namely the prison environment, establishing a deep-seated system of preferences in the minds of the inmates regardless of their consent. Are agents who have lived under the influence of this environment really free to reason on their own after years of being unintentionally trained to think along specific lines? And by extension, do their choices and character come from themselves and not from their conditioning? It seems the answer in both cases must be no: the ability to freely gauge benefits and consequences in the long term, and to organize our time as we see fit are some of the most basic ways in which we express both the moral freedom to choose our actions and to reason morally. Introducing constraints on that behavior, such as were detailed above, is a clear violation of those rights. The long term effects of prison, extending sometimes for years after release, have also been shown to have unpleasant consequences. Conditions such as post-traumatic stress syndrome have been observed in many long-term prisoners, manifesting themselves in high levels of anxiety, chronic depression, insomnia, and, perhaps most importantly, a feeling of intense difference from previous selves. (Maruna, Liebling 15). Even more extreme alterations have been observed in the habits of wrongly convicted inmates after their release. Adrian Grounds, after conducting a series of interviews with the subjects and their families, observed that many of the released had difficulty readjusting to the outside world and continued to utilize strategies learned in prison to deal with their work and family lives, including emotional and physical withdrawal, self-isolation and non-

communication. Many were unable to resume their former relationships and eventually came to live alone. (Grounds 36-37)

Unlike the more abstract changes in how people consider short-term and long-term gains, behavioral modifications of this kind constitute a clear violation of the right to freely decide one's actions and character in exactly the sense outlined above. The convicted in these cases have had changes foisted upon them that are far from what they wanted, and hence can no longer correctly identify as the free creators of the person they have become, breaking our second criteria. This is precisely the kind of change that Morris feared from the therapeutic approach, yet it is, in at least some cases, emerging from the experience of undergoing retributive punishment. Of course, since I have described only a very limited and admittedly selective cross-section of the extensive literature on this subject, a word of caution is definitely in order. Questions about the effects of prison, especially long-term effects, are a subject of considerable debate. Many of the long-term changes discussed above are hard to quantify, and as a result have been criticized for lack of rigor. Those attempts to gauge the impact of prisons on inmates in strictly quantitative terms have yielded mixed results, and in a few cases results contrary to what I have argued. An excellent summary of these studies can be found in James Bonta and Paul Gendreau's paper, *Reexamining the Cruel and Unusual Punishment of Prison Life*. In the studies they investigated, a much more complex picture of the prison environment emerges, in which prisoners experience varying responses to confinement, in some cases maintaining pre-imprisonment levels of wellness and in some becoming vastly more successful. However, they

admit that more phenomenological approaches to the subject might indeed yield different results, and take a fairly positive view towards the research into prison stress conducted by Toch and his associates.

Does the fact that prison can have a wide range of possible impacts on its prisoners present a challenge to the idea that they are engaged in personality alterations? I believe the answer is no. Regardless of the complexities of these issues, the fact that the alterations caused by prison stress have been observed at all, regardless of questions about methodology or degree, should still raise major questions for any proponent of retributivism as a morally acceptable approach to punishment. That some inmates do not develop convoluted personalities does not do away with fact that a substantial number do; we can imagine that a similar situation would apply even in the case of the most sinister system of criminal therapy, but that that does not minimize its violation of the right of personhood in those cases where it succeeds. If agents convicted under a retributivist prison system cannot ensure that their autonomy, here understood as the capacity to engage in moral judgment and to be the moral originator of one's actions and character, then that system is violating its most basic principles, period.

If it will be granted, as I have argued, that there are major issues about personal autonomy raised by the practice of imprisonment, and that this fact poses a problem for Morris's theory, then it will be incumbent on a defender of Morris to show how can imprisonment can properly be part of a just system without simultaneously violating the rights on which such a system is founded.

III.) Counterarguments and Conclusion

Though there are any number of counter-arguments available to the defender of Morris, I will focus on two in particular, the first concerning the way I have used these data and the second concerning its impact for retributivism as a theory. I will argue that both of these concerns are unfounded, though the second point is important in that it illustrates the broader problem of which this data is only a single instance.

The first of these objections questions whether I am justified in inferring the psychological impact of imprisonment under pure retributivism from psychological data drawn from our presently existing system, since that system is not itself fully retributivist. The violations of autonomy that I have singled out from these data, a retributivist might argue, is endemic in our prison system because our system is not properly retributivist. Indeed, the data I have presented is from a system equally indebted to consequentialist designs, and therefore it cannot be validly applied to the theoretical concerns of retributivism.

This is a very solid point. Obviously, arguments of this kind can only be made with the data we have, and the data we have are from a system which has elements of both retributive and consequentialist methods. Thus, any point we can make with it will inevitably be tainted by the other side's techniques.

That said, it seems to me that the issues at hand stem more from the nature of prisons in

general than from any particular moral commitments they might have. If we return to the data, the factors most important in inducing these non-punitive effects are not distinctively retributivist or consequentialist. They are, rather, the most basic elements of the imprisonment experience: confinement in a small space, isolation from the outside world, and large blocks of unoccupied time in which to consider one's predicament. Taken together, the impact of these forces can be surprising. Though there is no question in my mind that different methods of managing this system can have an effect on how powerful that impact can be, at root, the problems I have singled out are proper of prisons *qua* prisons, and are not at base, a question of the philosophical approach we take to them as a set of institutions.

If this interpretation is correct, then my central claim has been confirmed: imprisonment-based retributivism violates the rights it is supposed to protect. However, to the defender of retributivism, this might not be the end of the line, since my thesis is concerned solely with imprisonment. They might agree that imprisonment cannot be justified under retributivism, but object that this does not prove a problem for retributivism more broadly, because imprisonment is only one of a great many methods of punishment, all of which could conceivably be made to fit Morris's system if we understand it charitably enough. In other words, there are problems with imprisonment, but retributivism is not necessarily tied to imprisonment and hence this is not a necessary problem.

This objection is solid, not because it is right *per se*, but because it highlights a more general difficulty for a theory such as Morris's. This difficulty goes beyond the central concern

of my thesis, but it illuminates some aspects of the data as it relates to the broader question of retributive punishment.

Suppose that we are searching for the ideal method by for punishing the guilty. We know it must respect the *lex talonis* principle foundational to so much of retributivism; the punishment must be proportional to the crime. We also know that it cannot harm our ability to reason morally, or our ability to determine our destinies in the long run, or our recognition of our choices as our own. What must the resulting punishment look like?

The difficulty arises from trying to reconcile these two demands in whatever approach we settle on. If we raise the penalty for a truly heinous crime high enough to pay back for that act, we begin to push at the limits of what humans can psychologically handle. For example, suppose we were to replace imprisonment with a system of fines. Parking violations would likely stand at the very bottom of this ladder, being for a minimal crime and accordingly being a for a fairly small amount of money. Assault and theft would stand on the next step, followed by involuntary manslaughter and with rape and murder at the top. Assuming that each step up the ladder implies a higher and higher fine, when we reach the truly heinous crimes, the fines would have to be high enough to virtually eliminate all hope of having any real success later in life, and indeed, in some respects, could amount to a sort of *de facto* enslavement to the demands of paying for the crime that was committed. The psychological impacts of such a situation are easy enough to imagine; people living in dire poverty with no hope of recovery and the overwhelming threat of not being able to make the next payment hanging over their heads will likely begin to suffer ailments not

unlike those afflicting prisoners, in spirit if not in form. The same problem would presumably apply to whatever methods might be suggested; no matter how we go about it, hard punishment will be hard, often in psychologically distressing ways, and as long as we hold to Morris's framework, we will always be bound to reach this kind of tipping point in our pursuit of justice. Thus these problems are not just contingencies that retributivism has accidentally attached itself to; they are the deep-seated problems inherent to retributivism.

Taken from this angle then, the data I have offered about prison stress are actually examples of a larger conflict which definitely deserves further investigation, both empirically and philosophically. They illustrate the fact that a retributivist approach that respects autonomy will always be caught between two problematic positions; seeking to punish adequately for the crimes committed, and seeking to respect the autonomy of those being punished. Though the argument I have offered for most of this paper applies solely to imprisonment, it is thus an important fact to keep in mind as these data apply to retributivism more broadly, and suggests serious questions must be raised about retributivism as a system as it pertains to human limits.

Regardless of whether these difficulties are limited to imprisonment or extend more broadly, however, Morris's system has lost much of what made it so appealing in the first place. In its original formulation, it demanded absolute respect for the persons living under it, embodied in the rights that form and validate the legal superstructure constructed over it. These rights, he argued, included the right to be punished, and to have one's autonomy as a moral agent respected. To violate them would invalidate the moral legitimacy of the system as a whole.

However, the criticisms Morris levels against consequentialism are just as true of retributivism, at least so far as it requires imprisonment which will violate the rights to autonomy. As is apparent from research on prison stress, long term behavioral patterns of prisoners, and post-traumatic stress amongst the wrongly convicted, the effects of imprisonment can often undermine autonomy in quite a dramatic fashion, denying the imprisoned the moral status Morris claims they deserve as the authors and judges of their actions. If this data is in fact an accurate representation of the essential realities of imprisonment, then Morris's insistence on the primacy of rights must also be his undoing, since it demands both that he punish and, in the process, violate the autonomy he had originally sworn to protect.

Works Cited

Bonta, James, and Paul Gendreau "Reexamining the Cruel and Unusual Punishment of Prison Life." Law and Human Behavior 49 (1990): 347-372.

Haney, Chris, C. Banks, and P.G. Zimbardo, "Interpersonal dynamics in a simulated prison." International Journal of Criminology and Penology 1 (1973): 69-97

Jamieson, Ruth, and Adrian Grounds. "Release and adjustment: perspectives from studies of wrongly convicted and politically motivated prisoners." Liebling 33-65

Liebling, Ali, and Shadd Maruna, ed. The Effects of Imprisonment. Portland: Willian, 2005.

Flanagan, Timothy J. "Lifers and Long-Termers." The Pains of Imprisonment. Ed. Robert Johnson. Beverly Hills: Sage, 1982.

Morris, Herbert. "Persons and Punishment." On Guilt and Innocence. University of California, Berkley. London: 1979. 31-58.

Murphy, Jeffrie "Marxism and Retribution" Philosophy and Public Affairs 2 (Spring 1997): 217-243.

Toch, Hans. Mosaic of Despair: Human Breakdowns in Prison. Washington: American Psychological Association, 1975.