ABSTRACT. This paper argues that the general practice of punishment cannot be successfully defended by appeals to social contract arguments based upon the work of John Rawls. Several attempts to present such justifications are discussed, including those by Murphy, Morris, Sterba, and Hoekema. It is argued that social contractors would not choose a practice of punishment because such a practice is a symbolic expression of society's disapproval of offenses against the law. Social contractors would instead choose a practice which might have some deterrent effect upon crime but would lack the feature of social contract theory.

Is it rational to agree to be punished? Would a rational person "contract" either for an institution which administers punishment or for principles which would act as side-constraints on such institutions? Unsatisfied that utilitarian or retributivist theories alone can provide answers to these questions, several writers have tried to show that a social contract theory based upon the work of John Rawls provides a moral justification either for practice of legal punishment or for particular principles governing the use of punishment. The purpose of this paper is to assess these arguments and address the question of whether a social contract theory will provide a moral justification for the existence or the practices of an institution of legal punishment.

These arguments are important in that they address the major dilemma in the attempts to find firm foundations in moral theory for the practice of punishment and, thus, these arguments address the fundamental moral principles of the criminal justice system. Since John Rawls' and H.L.A. Hart's work, the problem of moral justification of punishment has been divided into two parts: the justification of the general practice of punishment and the justification of particular punishments for particular persons. The chief theories which have been offered to provide these justifications have been utilitarianism, with deterrence as a chief justificatory purpose of punishment, and retributivism with the desert of the offender as the condition which justifies punishment. Neither of these theories by itself has fared well in the literature either as a justification of the general practice of punishment or as a justification for particular side-constraints on punishment. A common criticism of utilitarianism has been that the principle of utility would permit not only the punishment of offenders but the punishment of the innocent as well, when such punishment would maximize the social benefits of punishment. This criticism is well known in the literature and does not need to be developed here.
If utilitarian justifications have fared poorly, retributivism has not won the day by default. First, there has been wide disagreement over just what constitutes a retributivist theory of punishment. Second, there are well known, serious problems in determining the severity of punishment which is "deserved" by an offender, and, third, it is not clear that the general practice of punishment by an institution is justified even if there is a defensible theory of personal desert. Is it really obvious that society has a moral right to punish a person even if the person deserves to be punished? When retributivism is considered as a justification for particular punishments, the theory is subject to criticism concerning the alleged correlation in severity of offenses and punishments. Furthermore, retributive theories have not usually been sufficiently complex to consider the wide variety of kinds of treatments a person might be said to deserve under different conditions.

Given the difficulties which have faced utilitarian and retributive theories, it is not surprising that their proponents have looked for a strategy which would ground one or the other of these theories in a wider, more powerful theory from which they could deduce the basic principles of either a deterrence or a retributive theory of punishment. Forms of the social contract theory formulated by Rawls have seemed initially to be well suited for this task. If one could show that rational persons choosing principles or institutions which would govern them would choose to be punished for violations of law or would choose to have punishment administered by the principles of either a deterrence or a retributive theory, one would seem to have provided a moral justification for punishment as a practice or for the principles of punishment one wishes to defend. After all, one could then claim that this practice or these principles are, in effect, ones which persons have chosen or would choose in making rational choices about how they are to be governed.

An example of this strategy is found in the work of Jeffrie G. Murphy. In explaining Kant's solution to the problem of justifying the punishment of criminals, Murphy claims that the Rawlsian method yields as a result that people would choose principles of punishment and that we may interpret their having chosen these principles as acts in which criminals have, in a sense, willed or chosen their own punishment.

The criminal himself has no complaint (about his being punished), because he has rationally willed or consented to his own punishment. That is, those very rules which he has broken work, when they are obeyed by others, to his advantage as citizen. He would have chosen such rules for himself in an antecedent position of choice--what John Rawls calls "the original position".

Similarly, James Sterba has argued that from behind a veil of ignorance a person would adopt principles which provide "a morally defensible account of retributivism that is not based on vengeance, yet generally conflicts with the requirements of utility". Furthermore, other social contract arguments have been offered to establish limits on the severity of punishment and to defend procedures which assure that no punishment would preclude the compensation of a person who had been unjustly punished.
There is a glaring omission in all of the above appeals to contract arguments in support of various principles of punishment. A Rawlsian social contract argument requires that we be given in advance a description of the state of nature or the original position from which choices are to be made. If it is assumed in advance that there will be practices of punishment roughly resembling our own, then one might appeal to the social contract arguments in support of particular side-constraints on the use of punishment. However, those appeals will not then yield any justification for the practice of punishment, because the most fundamental moral questions concerning punishment will be begged, namely, whether the practice of punishment is morally justifiable at all. On the other hand, if one does not assume in the original position that there is a practice of punishment of which the hypothetical contractor has knowledge, then some additional argument must be provided to show that the contractor would even choose to have the practice of punishment exist at all. Given this problem we have to divide the problem of assessing the social contract methodology into two parts. First, we must consider whether a contractor would choose to have such a practice at all. Second, given that there is such a practice, we must assess what, if any, side-constraints can be justified by the method of the social contract argument.

David Hoekema divides the problem of the justification of punishment in this way in his use of a social contract argument. Hoekema describes the social contractor as knowing, as part of the general knowledge possessed in the original position, that persons have rights and that having any first-order right, such as a right to one's property, also confers on one a second-order right to defend that first-order right. Hoekema then argues that social contractors would choose to have an institution of legal punishment in order to deter potential lawbreakers from violations of rights. According to Hoekema, the rational contractor would transfer the second-order right to society by agreeing to have an institution of punishment in order to achieve deterrence. Hoekema's justification of the practice of punishment is that the practice would be chosen. Anyone who is punished for violations of the rights of others has, in a sense, chosen to be punished by having chosen to have an institution which must punish offenders in order to protect first-order rights. This differs from the traditional deterrence theory in that it is the choice to transfer second-order rights and not the belief in deterrence which is said to provide the justification of the practice of punishment. The choice itself is said to be rational given the belief that punishment will deter violations of rights.

Hoekema's assumption that the contractor has a desire to deter others from violating the contractor's own rights is important here. This assumption seems to forestall one type of criticism which has been made of social contract theories of punishment. George Sher has argued that even if the contractor would agree to practices which are needed to maintain a just society, this hypothetical agreement would still not justify the practice of punishment. According to Sher, what is required for such a justification is the actual consent of an individual which "... must stem from actual desires and intentions." Since the hypothetical consent does not stem from actual desires and intentions, it does not offer a basis for the claim that an offender has agreed to be punished for offenses committed.
Sher's claim that the contractor's consent must stem from actual desires and intentions could mean one of two things. On the one hand, Sher might be claiming that only actual persons can give their consent to social contracts since only actual persons have actual desires and intentions. If this is Sher's claim, then the claim is not so much a criticism of Rawls' method as a refusal to allow the method at all, because Rawls' social contractors are not actual persons. On the other hand, Sher might be claiming that Rawls' method requires that some actual, fairly specific, desires and intentions be ascribed to the hypothetical contractors. These actual desires and intentions would be required to show the rationality of a choice of any practice or principle which would maintain justice in face of noncompliance. If the latter is Sher's criticism, then the criticism can be met by ascribing to the social contractor in the original position some specific desire such as a desire to deter others from violating the contractor's rights. In this way, Hoekema's assumption that the contractor has a desire to deter others from violating its rights provides an answer to Sher's criticism of social contract theories of punishment.

Suppose we grant, as Hoekema does, that the contractor has an actual desire to deter others from violating its (the contractor's) rights. Is this now an acceptable justification of the practice of punishment? Is it clear that a social contractor would choose on grounds of deterrence to have a system of punishment? Since it is assumed by Rawls that the contractor has no interest in the well-being of others but only in the contractor's own well-being, the contractor would make a choice of some systematic way to protect its own rights. But the institution to be chosen would be one which would maximize the contractor's own opportunities, rights, and goods no matter what role the contractor might turn out to be playing in the society. Suppose we limit consideration to those violations of rights except for the taking of life. That is, suppose that the contractor is not at risk to lose its own life as the result of being deprived of a right. This supposition is important for what follows. Whether a contractor's reasoning changes when life itself is at stake will be considered later.

It may now be asked whether this contractor would choose a system of legal punishment. Would such a contractor necessarily have an interest in general deterrence of potential offenders? It is not clear that general deterrence would be the main concern of the contractor, since the contractor may turn out to be one who is an offender rather than one who is a victim, and the contractor would choose to maximize its opportunities for whatever social role it turned out to play. Whatever the contractor would choose must satisfy the condition that the practice would be in the contractor's self-interest in whatever role the contractor fills. In particular, the Rawlsian contractor will choose from among the alternative principles and practices that one (or the ones) whose worst position is better than the worst position provided by the other alternatives.

Under systems of legal punishment, the most plausible candidates for 'worst' position in the practice are: the victim, the offender, and the innocent victim of punishment who has been falsely accused and convicted. It need not be decided here which of these is the worst position. The worst position might vary depending upon the type of crime committed, whether the offender was apprehended, and other conditions. Whichever of these is the worst position, a social contract argument
must show that a person in that position is better off given a system of legal punishment than he or she would be under any alternative practice for dealing with offenders.

Punishment, as a practice, does not seem to satisfy this condition. The practice hardly seems to maximize the opportunities of offenders. If one has chosen to violate the rights of others, then one would not choose practices which would, in general, deter violations of rights. At most, one would choose some practice which would deter violations of one's own rights consistent with leaving oneself undeterred to pursue one's violations of other's rights.

One can imagine such an institution. Suppose there were an institution which would provide restoration to the victim of what is lost (or compensation where restoration is not possible) and, henceforth, provide some special protection of the victim so that the same type of loss would not be suffered again. For example, if a person's home were burglarized, the victim would be compensated for any loss and, thereafter, a special surveillance system would assure that no further burglary of the victim's home would occur. The institution would provide no mechanism to deter violations of rights generally, but only in special cases of persons who had already been victims. If one has no interest in the well-being of others, then one's choice is to protect one's own well-being by assurance of restitution and prevention of a repeated loss.

This type of practice would leave any of the three candidates for worst position at least as well off as he or she would be under a system of legal punishment, and it leaves two of them better off. In the case of the victim of an offender, he or she would receive compensation or restitution which would leave the victim as well or better off than he or she would be under a system of punishment. If it is assumed that the violations are fully compensable, then the victim will have been restored to his or her status prior to the violation.

The other two candidates are better off under the imagined practice than under a system of legal punishment. The innocent person who is falsely accused and convicted suffers a loss by having to make restitution for another's offense, but he or she will not suffer any social stigma expressible by punishment, since by hypothesis there is none. Under a system of punishment, the falsely accused loses moral standing and carries the stigma of a criminal record with all of the disabilities that go with it.

Finally, this type of practice also maximizes the opportunity of the contractor in the case where the contractor is also a violator of rights. All that the offender loses in that case is the ill-gotten gain, through restitution, and a minimal loss of opportunity, through the inability to choose the same victim again since that person will now be under special protection.

The practice which has been described is not a practice of punishment in that it lacks the important, even essential, feature of institutionalized punishment of symbolically expressing public outrage or disapproval of offenses. It is for good reason that the practice described above would lack this expressive feature, for it is the expressive feature which would not be shown by the hypothetical contractors. The
contractor is assumed not to have an interest in the well-being of others and, so, not have an interest in a shared or common good, the violation of which is the occasion for, and even the justification of, the moral outrage symbolically expressed through acts of punishment. The expressive function of punishment condemns, on behalf of all persons, the offense and the character of the offender. It is this act of condemnation that would not be chosen by the contractor in either of the possible roles as chooser. As one who obeys law, the contractor lacks interest in others, and, as one who chooses to violate law, the contractor would not choose to reduce further the chances of successful violations. 

The scheme which the contractor would choose is a kind of regulatory practice, and, as such, it might well reduce criminal activity. This would not make the practice one of punishment, however, since regulatory practices lack the symbolic feature of punishment. Thus, one should not think that a justification of the practice of punishment has been discovered in the social contract argument even if a justification of a near relative of punishment has been found.

The above result also has bearing on those proposals in the literature to analyze punishment as a type of redistribution scheme. It has been argued that punishment serves to redistribute the burdens and benefits of adherence to law in such a way as to offset the alleged disproportion of burden and benefit brought about by offenders. According to the argument, this redistribution would be justified because the scheme of redistribution would be chosen by all, even possible offenders, in making a social contract. One highly problematic part of these arguments, the notions of burden, benefit, and their distribution, has recently been criticized by Richard Burgh. If my argument in this paper is correct, then these attempts are flawed in their very foundations. The foundations for distribution of benefits and burdens is in the models of regulatory and redistribution practices and not in punishment. Thus, the attempts to ground either retributive or utilitarian theories of punishment in a social contract theory fail. What such attempts would justify is not in the proper category of social practice.

Finally, the argument here has assumed that the social contractor is not at risk to lose its life as a result of noncompliant acts. Does the critique of the social contract argument change if one assumes that the contractor knows that life, as well as goods, are at risk due to noncompliance? It definitely does. It is in the case of murder that the model of redistribution benefits and burdens seems totally out of place. The problem in murder is not to properly distribute the benefits and burdens of murder but to prevent it altogether. The regulatory practices are not of the proper type to consider here. A social contractor who is averse to risk of life would choose some practice which would prevent murder from occurring. Whether such a practice would be punishment with the symbolic condemnation of murder is not clear. It has not been shown that it would be. It is not my purpose in this paper to try to imagine to what sort of practice a social contractor might agree to reduce or even eliminate the risk of loss of life due to noncompliance. Even if it turns out that in this special case the practice chosen would be one of punishment, the justification of punishment would only be for one special type of noncompliance and not for the general practice of punishment.
This does not preclude the possibility of using social contract arguments to support various principles which limit the practice of legal punishment, given that the practice exists to begin with. The claim here is that the practice itself receives no moral underpinning from the Rawlsian social contract methodology. The further outcome of this paper is that the search for general moral justification of punishment has been futile. Neither utilitarian nor retributive theories have provided it, and now the attempt to provide it in a social contract theory fails also. The only remaining view appears to be that of philosophical anarchism with respect to punishment. Pincoffs has written:

The short answer to Tolstoy's question, "By what right do some people punish others?" is that, needing a practice, we do not know any better one than legal punishment. 19

Given the current state of the practice of punishment and its effectiveness, this outcome should present a challenge to look hard for alternative practices.

ENDNOTES

1 The author wishes to thank Jeffrie G. Murphy and the participants of the NEH Summer Seminar on the Philosophy of Crime and Punishment for their helpful discussions of an earlier version of this paper. This project has been supported by the National Endowment for the Humanities as part of a large summer seminar project and by a Faculty Research Grant from Sam Houston State University.


5 It is not the purpose of this paper to rehearse the various positions on the definition of retributivism. For our purposes a retributivist position will be one which holds that it is morally justifiable to punish a person convicted of an offense provided the offender deserves the punishment. The desert is usually acquired by being convicted of an offense. Further, the retributivist view will claim that the severity of the punishment assigned must correspond in some systematic way to the severity of the offense. See John Kleinig, Punishment and Desert, (The Hague: Martinus Nijhoff, 1973); Honderich, op. cit.; Sidney Gendin, "A Plausible Theory of Retribution" The Journal of Value Inquiry 5 (1971), 1-18; Jeffrie G. Murphy, Retribution, Justice, and Therapy, (Dordrecht: D. Reidel, 1979); Edmund Pincoffs, The Erationale of Legal Punishment, (New York: Humanities Press, 1966); Hugo A. Bedau, "Concessions to Retribution in Punishment" in Cederblom and Glizek, 51-73, and "The New Retributivism" Journal of Philosophy 75 (1978), 601-20; Richard W. Burgh, "Do the Guilty Deserve to be Punished" Journal of Philosophy 79 (1982),
THOMAS W. SATRE


6 Murphy, Retribution, Justice, and Therapy, 83.


14 Joel Feinberg, "The Expressive Function of Punishment", in Doing and Deserving, op. cit.

15 Hoekema is aware of the problem of taking into account the possibility that the social contractor may turn out to be an offender (criminal). Hoekema says in defense of his including the offenders among the choosers, "Will this lead to complete rejection of punishment? It will not, for, in the first place, general compliance with the law which is brought about by institutions of punishment benefits criminals also. More important, under the conditions of ignorance criminals do not know that they are criminals". ("The Right to Punish and the Right to be Punished", 251.) This claim is not accurate however. The choice of punishment benefits the chooser in the chooser's role as possible victim but not in the chooser's possible role as criminal. The chooser is, at the same time, choosing to maximize his well-being in both of these roles.


17 Jeffre G. Murphy, Retribution, Justice, and Therapy, op. cit., 83-4; Herbert Morris, "Persons and Punishment" The Monist 52 (1968), 476-78.

18 "Do the Guilty Deserve to be Punished?", op. cit.

19 The Rationale of Legal Punishment, 136.