

A Role for Expression in Retributive Theories of Punishment

Clair Morrissey

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Philosophy.

Chapel Hill
2006

Approved by
Advisor: Thomas E. Hill Jr.
Reader: Michael Corrado
Reader: Gerald Gaus

© 2006
Clair Morrissey

ABSTRACT

Clair Morrissey: A Role for Expression in Retributive Theories of Punishment
(Under the direction of Thomas E. Hill Jr.)

In “Persons and Punishment” Herbert Morris defends a retributivist theory of the justification of punishment which revives the broadly Kantian tradition of arguing for the legitimacy of legal institutions that respect the members of a given moral community as persons. Morris’ theory uses the normative relationship between free and equal people as the source of constraints on just institutions of punishment. However, the benefits and burdens model he employs to make sense of how these institutions maintain and respect the normative status of persons is subject to question. I will argue that Joel Feinberg’s expressive function of punishment is compatible with the broadly Kantian justification of punishment and can be used to explain the nature of and constraints on just legal and penal institutions in place of Morris’ benefits and burdens model.

For the students at Vista Maria and C. A. Dillion Youth Development Center.

TABLE OF CONTENTS

	Page
Introduction	1
Morris' Retributive Justification of Punishment	6
<i>Benefits and Burdens</i>	6
<i>Personhood</i>	8
<i>Right to be Punished vs. Rehabilitation</i>	10
<i>Nature of the Right to be Treated as a Person</i>	13
<i>The Big Picture</i>	15
Wasserstrom's Objections to Morris	16
<i>Wasserstrom's First Objection</i>	16
<i>Analysis of Wasserstrom's First Objection</i>	17
<i>Wasserstrom's Reformulation of the First Objection</i>	19
<i>Analysis of the Reformulation</i>	20
<i>Wasserstrom's Second Objection</i>	21
<i>Analysis of Wasserstrom's Second Objection</i>	22
<i>The Social Equilibrium</i>	23
Feinberg's Expressive Theory of Punishment	26
<i>Penalties vs. Punishments</i>	26

Condemnation.....28

Vengeance.....30

Reactive Attitudes and the Community.....33

What Expression Is and What Expression Is Not.....37

*The Expressive Function’s Role in Legitimate Insitutions of
Punishment*.....40

Condemnation and Constraints on Institutions.....41

Conclusions and Future Directions.....43

Works Cited.....45

Introduction

There is an influential tradition in moral and political theory that takes as its starting point the reasonable agreement of a community of free and equal moral agents on the principles of social coordination. This approach is most commonly associated with the Kantian theory that derives principles for action from the Kingdom of Ends formulation of the Categorical Imperative. More recently this approach is associated with Rawls' Kantian conception of rational agents choosing the principles of justice from the Original Position.

A foundation of this framework is the belief that everyone is normatively equal. Every person has the basic right to freedom, in virtue of the fact that each person is capable of making choices according to rational (non-desire based) principles and capable of responding appropriately to reasons.

These free and equal people take themselves to be the authors of their actions and acknowledge that they are answerable to the greater community for those actions. They are responsible moral agents, related to other members of the community in a normative way. This relationship involves things such as rights, ethical duties above and beyond rights and respect for one another as persons. This respect for one another as free and equal members of the community includes preventing the degradation and humiliation of people by avoiding treating fellow people as animals or objects, incapable of making reasonable and rational decisions and available to be used as means to one's own ends.

Respect for persons is also manifested in conforming to principles and laws that could be derived from a point of view that treats everyone as free and equal. Principles derived

from this point of view will be those that treat people fairly, as they would need to be agreed upon (hypothetically) by all members of the community. Conforming to these kind of principles means that one treats everyone fairly, which is respectful of the status of each of the members of the community as normatively equal.

My aim in this paper is not to defend a particular interpretation of Kantian moral theory, but rather to use it as a framework for thinking about the justification of punishment. Many theorists working in this tradition take a version of retributivism to be the correct theory of the legitimacy of punishment. Retributivism is understood, in part, as a backward looking theory of desert. It is a complicated theory that includes, among other things, that punishment is justified only if the agent has done some legally blameworthy action.

Some versions of retributivism emphasize the value of respect for persons by stressing the conception of culpable lawbreakers as rational agents who choose to break the law. This respects the lawbreakers as free agents, capable of choice. The theory also seeks to prescribe requirements for just legal and penal institutions. Part of this prescription often includes the claim that in order to justifiably punish someone it must be established through a public and well-defined process that the person is culpable. Culpability is usually understood in these sorts of theories to mean that the lawbreaker acted deliberately, carelessly or recklessly, with adequate access or understanding of the situation at hand and the requirements of the law.

If one accepts both the broadly Kantian tradition and the retributivist justification for punishment, it appears that individual lawbreakers are rationally obligated to accept their own just punishments as legitimate. In virtue of their status as moral agents lawbreakers, in some sense, endorse or consent to their own punishment. In fact, as moral agents they seem

required to demand punishment as a constitutive condition of their own status as free and equal members of the community.

Kant, himself, recognizes that this is a somewhat puzzling consequence. Although he denies that specific lawbreakers will their own punishments, he writes in *The Metaphysics of Morals*:

No one suffers punishment because he has willed *it* but because he has willed a *punishable act*; for it is not punishment if what is done to someone is what he wills, and it is impossible *to will* to be punished.¹

The sense in which a lawbreaker must consent to or endorse (perhaps will) her own punishment cannot be at the level of particular individuals willing particular punishments. In fact, Kant believes if someone wills an action, it cannot be considered punishment.

However, he goes on to explain a sense in which he believes someone can intelligibly be thought of as willing her own punishment, writing:

Saying that I will to be punished if I murder someone is saying nothing more than that I subject myself together with everyone else to the laws, which will naturally also be penal laws if there are any criminals among the people.²

Expanding this point, he writes:

When I draw up a penal law against myself as a criminal, it is pure reason in me (*homo noumenon*), legislating with regard to rights, which subjects me, as someone capable of crime and so as another person (*homo phaenomenon*), to the penal law, together with all others in a civil union.³

Here, Kant is trying to preserve a sense in which a lawbreaker consents to her own punishment, but this sense is rather obscure.

¹ Immanuel Kant. *The Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge University Press. 1996. 6:335.

² Kant, 6:335.

³ Kant, 6:335.

Part of the difficulty with understanding these passages comes from difficulty with establishing the importance and nature of what Kant refers to as the will. In order to understand how Kant thinks lawbreakers can be committed to the legitimacy of their own just punishments, we must be able to give an account of the will and its relation to personhood. This is notoriously difficult to do and would not allow me to stay appropriately neutral to interpretations of Kantian moral theory. Rather than attempt to make sense of Kant's particular formulation of this puzzle, I will turn to Herbert Morris' contemporary account of Kantian inspired retributivism.

In "Persons and Punishment" Herbert Morris resurrects the question of what it would mean to be rationally committed to one's own punishment by exploring what it could mean to have the right to be punished.⁴ Morris answers this question in terms of the institutions that free and equal persons would consent to from a position antecedent to the particular instances of law-abiding and lawbreaking behavior. By appealing to the nature of a particular kind of community's institutions he has taken up Kant's suggestion that the way in which lawbreakers are committed to the legitimacy of their own just punishments is found in what it means for a group of rational individuals to be, together, under the law.

In this paper I am interested in exploring how the nature of the relationships of individuals in a community of free and equal agents under the law is related to and ultimately justifies institutions of punishment. Section I will present Herbert Morris' argument for a retributivist theory of punishment from "Persons and Punishment." In Section II, I analyze two important criticisms of this view given by Richard Wasserstrom in *Philosophy and Social Issues*. I will argue that Wasserstrom's objections show that a more robust understanding of the nature of

⁴ Herbert Morris. "Persons and Punishment," In *Punishment and Rehabilitation*. 3rd Edition. Edited by Jeffrey Murphy. Belmont, CA: Wadsworth Publishing Company. 1995. p. 74.

the relationship between members of a community under the law must be given. In Section III I will turn to Joel Feinberg's expressive function of punishment in order to show how it can be incorporated into a retributivist justification for punishment in light of Wasserstrom's objections.

Morris' Retributivist Justification of Punishment

Benefits and Burdens

Herbert Morris' highly influential paper "Persons and Punishment" revives the Kantian argument for the right to punishment. An important part of Morris' account is the benefits and burdens model that motivates the need for an institution of punishment. Morris asks us to imagine a group of people who are

constituted roughly as they now are, with a rough equivalence in strength and abilities, a capacity to be injured by each other and to make judgments that such injury is undesirable, a limited strength of will, and a capacity to reason and to conform conduct to rules.⁵

Further imagine that there are a set of primary rules that apply to the conduct of this community, which are similar to the core rules of our criminal justice system that prohibit violence and deception.

These rules provide a benefit to each member of the community, as the rules provide room for each individual to pursue what he or she values, free from violence and interference by others. This mutual benefit is created by each person assuming a kind of burden. The burden is the practice of self-restraint in situations where individuals are inclined to act in ways that interfere with the freedom of others. When someone fails to follow the rules they are relieving themselves, unfairly, of a burden that the other members of the community bear. In this way, the rule breaker gains an unfair benefit.

⁵Morris, p. 75.

Morris gives three (primary) reasons a system of punishment should be connected to a violation of the primary rules of this society. The first is that there should be a way to provide assurance to each member of the community that they will not be assuming an unfair burden. If there is no such assurance, individuals will be less likely to voluntarily assume these burdens, and the mutual benefit gained from each individual doing so will be lost. That is, being a rule-abiding citizens would be unduly risky.

The second reason is that “fairness dictates that a system in which benefits and burdens are equally distributed have a mechanism designed to prevent a maldistribution in the benefits and burdens.”⁶ When sanctions are attached to rule-breaking there is less of a chance that an unfair distribution of benefits and burdens will be created. As the members of the community are for all relevant purposes equal, fairness in the distribution of benefits and burdens requires that they each take on the same share. In order to maintain the fairness of the distribution, there must be a way of rectifying unfair distributions.

The third reason is “it is just to punish those who have violated the rules and caused an unfair distribution of benefits and burdens.”⁷ The rule-breaker has taken an unfair benefit at the expense of the rest of the community. He owes something to the community, because he took something that does not belong to him by rights. Justice, the punishment of the individual for breaking the rule, would restore a fair distribution of benefits and burdens.⁸ It is, in this case, a way of exacting a debt from the lawbreaker.

⁶ Morris, p. 76.

⁷ Morris, p. 76.

⁸ Morris notes that punishment is not the only way to reestablish the fair distribution of benefits and burdens. Forgiveness and pardon can also do this, and thus, this conception of punishment makes room for these important features of interpersonal interaction.

These reasons give rise to a host of institutional provisions, as it must be established (for the punishment to be legitimate) that the lawbreaker really does owe the community something. Exceptions will need to be made for those who it is unreasonable to think could have restrained themselves, or if it unreasonable to think the lawbreaker could have behaved otherwise. For example, someone with diminished mental capacity cannot reasonably be held accountable for failing to abide by the law.

To justifiably punish an individual the community would have to establish that the lawbreaker has actually gained the relevant sort of unfair advantage. This gives rise to a large burden of proof as well as requirements of due process. As the community's rules are such that they establish a social equilibrium of benefits and burdens based on each individual having a sphere of action free from interference, it is unjust to punish those who could not have acted otherwise, who do not gain the kind of advantage at issue or who are innocent. These illegitimate sanctions would be unjust because they would only serve to create the kind of unfair distribution that the institution of punishment is in place to correct.

Personhood

Morris argues that the right to be punished should be understood at the level of institutions, not particular judgments. We imagine a community of free and equal people who are deciding on the laws and institutions that will serve to regulate their community. From this position, fairness dictates that there be a mechanism for maintaining an equal distribution of benefits and burdens among all members of the group. Morris identifies one part of this mechanism as the institution of punishment. The institution of punishment is justified by appeal to the nature of the community. It is with the legitimacy of this sort of mechanism that

we move from claims about the personhood of individuals to a theory that seeks to give an account of legitimate institutions for the community.

For Morris, the right to be punished is the right to have institutions of punishment. Lawbreakers have the right to be punished (or would in some sense endorse their own punishment) because they are free and equal members of the community. In this way, claiming a right to be punished is to demand the right to be treated like a person.

We treat someone as a person when we 1) permit the person to make choices about a range of possible actions that are consistent with the sort of respect for persons that underpins the community of which they are a member and 2) our responses to the person recognize that person as having made a choice. We respect individuals as capable of making choices, as being able to commit to an action and take responsibility for both the negative and positive consequences of that action. Moreover, we react to individuals in terms of the choices they make. We praise good actions and blame bad actions. We attribute to the individual responsibility for actions and the ability to choose.

Steven Darwall's distinction in "Two Kinds of Respect," can be helpful in explaining the kind of respect at issue here.⁹ Members of the community extend to one another respect by taking one another into appropriate consideration in deliberation. This is what Darwall refers to as recognition respect, which we owe to all persons as such.

To say that other persons as such are entitled to respect is to say that they are entitled to have other persons take seriously and weigh appropriately the fact that they are persons in deliberating about what to do.¹⁰

An important feature of this kind of respect is that it requires me to reflect on what I am required to do, in virtue of the fact that I owe the other person respect.¹¹ Further, recognition

⁹ Steven Darwall, "Two Kinds of Respect," *Ethics* Volume 8, Number 1. (Oct. 1977): 36-49.

¹⁰ Darwall, p. 38.

respect is about having a certain kind of status. The status of personhood. Note that this is not respect for each individual as a morally good person. It is respect for an individual as the sort of thing that should be taken into my considerations in particular ways.

Darwall differentiates recognition respect from the sort of respect that reflects my positive appraisal of another person or their character. He refers to this second sort of respect as appraisal respect.

Unlike recognition respect, one may have appraisal respect for someone without having any particular conception of just what behavior *from oneself* would be required or made appropriate by that person's having the features meriting such respect.¹²

Appraisal respect can be something like a positive feeling of admiration, whereas recognition respect requires that one act in a particular manner toward the object meriting respect.

The kind of respect that seems important for creating legal institutions that are interested in leaving as much freedom as possible for each person to pursue her life as she sees fit is recognition respect. Often times we seem to illegitimately import appraisal respect into our legal institutions and policies. Unfortunately I do not have the space in this paper to explain how this can happen and why it is illegitimate. So, I will put that issue aside and assume for the purposes of argument that what is relevant for institutions of punishment is respect for the status of others, not the moral worth of others.

Right to be Punished vs. Rehabilitation

To say that we have the right to be punished is to claim that we have the right to institutions that mirror considerations of the status of personhood. These institutions are

¹¹ This does not limit respect for the status of persons to just my personal interactions with others. One important way in which we respect the status of personhood is by creating institutions that respect personhood, and conforming our conduct to the dictates of these institutions.

¹² Darwall, p. 39.

starkly different from a set of institutions that do not advocate punishment, but rather advocate replacing the system of punishment with a system of rehabilitation or therapy.¹³

Some theorists argue for a system of rehabilitation in place of a system of punishment. These sorts of rehabilitation theories view lawbreaking as symptomatic of psychological failings, disorders or problems. As Morris describes the view: “when an individual harms another his conduct is to be regarded as a symptom of some pathological condition in the way a running nose is a symptom of a cold.”¹⁴ Those who hold this sort of view deny that the lawbreakers deserve to be punished because the lawbreakers cannot be thought of as legally culpable for their actions. They advocate institutions that treat mental illnesses exclusively, in place of institutions that seek to reestablish some kind of fair balance between members of the community.

Morris objects to this system of social control because it treats all lawbreakers as mentally ill or sick, writing, “the logic of sickness implies the logic of therapy.”¹⁵ Taking the fact that someone has failed to conform her actions to the laws of the community as evidence of sickness or pathology is questionable, if not wrong. However, this inference is central to a theory that advocates replacing a system of punishment with a system of rehabilitation or therapy. For the most part, it seems that lawbreakers are not ill in ways that are incompatible with being culpable for their actions. Thus, in some cases replacing institutions of punishment with institutions of rehabilitation can be disrespectful to the personhood of the lawbreakers.

¹³ Karl A. Menninger presents such an argument in *The Crime of Punishment*, New York: Viking Press, 1969.

¹⁴ Morris, p. 78.

¹⁵ Morris, p. 79.

Morris' motivation for accepting a system of punishment over a system of therapy is that a system of punishment treats the lawbreaker (and each member of the community) as a person, and the system of therapy that views all lawbreakers as suffering from some sort of pathology does not. Punishment recognizes lawbreakers as choice makers, and responds to them as such. The sort of rehabilitation under discussion recognizes lawbreakers as ill, and responds by treating them with the idea that lawbreakers can (in best case scenarios) re-enter society as fully functioning members. Institutions of punishment already hold lawbreakers to have at least the same minimal level of rational capabilities and civic responsibilities. It is, in part, this fact about the members of the community that justifies a system of punishment in the first place.

Morris' objection to a system of rehabilitation does not preclude the incorporation of certain sorts of therapy or rehabilitation into penal institutions. Rehabilitation programs can be offered as part of legitimate penal institutions. However, these programs would be subject to the sorts of constraints that respect for persons dictates. For the most part respect for persons constrains the content and the approach of rehabilitation programs. Programs would have to respect the participants as rational agents who need support, resources or assistance; rather than treating them as if they are ill and need to be, in some sense, fixed.

Moreover, Morris' argument does not rule out the possibility of a separate set of institutions designed for those lawbreakers who are not deemed culpable for their actions. We may well have state sponsored mental health institutions, drug rehabilitation centers and the like for those lawbreakers who are mentally ill, drug addicts or otherwise not culpable for

their actions but who the state, nevertheless, has a legitimate interest in detaining.¹⁶ Morris' argument from respect for persons is levied against those who argue that the entire institution of punishment should be replaced by systematic rehabilitation and therapy. It is not an argument that rehabilitation and therapy are never justified.

Nature of the Right to be Treated as a Person

Morris argues that the right to be treated as a person is an inalienable right that individuals have in virtue of being human.¹⁷ He claims that there is no intelligible sense in which I can transfer my right to be treated as a person to another individual. What right, exactly, would I be transferring? The right to have my choices taken seriously as my choices? It does not seem that I can transfer this right in the way that I transfer say, my property right to my car. Transferring one's right to be treated as a person seems just as unintelligible as transferring one's right to life.¹⁸

It seems equally unintelligible to waive one's right to be treated as a person. How could one do this? You could volunteer to act as an instrument for some other person or group (for example you could work as a member of the president's secret service), but you volunteer to do so on your own volition. I could even forgive someone for treating me as an object or instrument, but the act of forgiveness presupposes that I have the right to be treated as a person in the first place.¹⁹

¹⁶ This could happen in cases where lawbreakers who are not criminally culpable are dangerous to members of the community. The community could have a legitimate interest in the safety of its members, and could thus incapacitate or detain non-culpable yet dangerous lawbreakers.

¹⁷ Morris, p. 88.

¹⁸ Morris, p. 90.

¹⁹ Morris, p. 90.

Given the way Morris construes the right to be treated as a person, it is also unclear how one could forfeit this right. The right itself comes from each individual's capacity to make decisions. There are some people who may lack this right in virtue of lacking the capacity, but it is unclear how one could lose it through a freely chosen action. Moreover, the right is derived from the capacity not the quality of the decisions people make. You could not forfeit your right to be treated as a person by simply making a lot of bad choices, as the right is tied to the capacity of choice-making; not of good choice-making.

One could object that it seems perfectly intelligible to transfer the right to be treated as a person under a contract. In contracts, we can exercise our decision-making capacity to transfer or waive our right to be treated as a person. Morris claims that this line of argument will not ultimately succeed in showing that we can waive our right to be treated as a person, writing: "any agreement to being treated as an animal or an instrument does not provide others with the moral permission to so treat us."²⁰ This consideration against the ability to contractually agree to waive one's right to be treated as a person is not about the possession of this right by an individual. Rather, Morris employs the nature of the relationships between members of the community to constrain the kind of legal institutions we have. His suggestion seems to be that the unintelligibility of waiving one's right to be treated as a person comes from how the members of the community are to treat one another, not directly from the constraints placed on how you, in virtue of having a capacity for choice, can act.

I find this way of using the nature of the community to place constraints on legitimate social institutions quite useful and suggestive. In Section II I will use this kind of approach in response to criticisms of Morris' benefits and burdens model.

²⁰ Morris, p. 90.

The Big Picture

Stepping back we can see the big picture provided by Morris. In virtue of being human, we have the basic right to be treated as persons.²¹ We have the right to be respected in our choices when those choices are consistent with the respect that underpins the communities of which we are part. This right is not something that can be transferred or alienated by agreement, nor can it be waived. In virtue of this right, we have the right to institutions that reflect our status as persons.

The institutions of criminal justice and punishment respect our status as persons, and institutions that replace punishment with the kind of therapy that views lawbreakers as necessarily pathological do not. The equal distribution of benefits and burdens (a kind of social equilibrium) also reflects this respect for personhood. We agree to a set of primary rules that makes it possible for us all to live our lives free of the unjust interference, as long as we respect this right of others. If we violate these rules, we have taken advantage of the other people, and are subject to sanction. Thus, a system of punishment that respects both the legitimate choices made by people and the value of each individual being able to live out their conception of the good life without interference or under an undue burden is a justified system of community coercion.

²¹ Morris argues that all humans have the right to be treated persons just in virtue of being human (p. 87-93). As I am presenting his argument I will follow him in this. However, it should be noted that a weaker version of this argument can be given. It could be presented in terms of individuals who are capable of making choices having the right to be treated as persons, in virtue of the fact that they are capable of making decisions.

Wasserstrom's Objections to Morris

In *Philosophy and Social Issues* Richard Wasserstrom presents two objections to Morris' benefits and burdens model. In this section I will present and analyze these two objections. My analysis will reveal that Wasserstrom's objections do not challenge the framework on which Morris constructs his benefits and burdens model, but do show serious problems with the benefits and burdens model itself. These challenges to Morris' account require someone working within this tradition to give an alternate (perhaps simply a more robust) account of how a kind of balance between members of a community should be maintained through social institutions. Section III will lay the groundwork for such an account.

Wasserstrom's First Objection

Wasserstrom's first objection to Morris' benefits and burdens model is that "it is not always plausible to think of criminal and law-abiding behavior in terms of benefits and burdens."²² Wasserstrom offers the case of rape to show the force of this objection. It seems implausible to claim that each member of the community has to suppress an inclination to rape someone else. If most people in any given community do not have a motivation to perpetrate rape, it is misleading to say that they have incurred a burden in having to restrain themselves.

Moreover, the construal of rape as the taking of an unfair benefit seems strange if not wrong. One would more naturally characterize this crime as a physical and psychological

²² Richard Wasserstrom. "Punishment." In *Philosophy and Social Issues*. Notre Dame: Notre Dame University Press. 1980. p. 143.

violation of the victim. In fact, Wasserstrom argues,

rape, torture, murder – many of the worst things one person can possibly do to another – do not neatly or obviously fit the model of the misallocation of benefits and burdens described by Morris as constituting the background justification for the punishment of the guilty.²³

We need a model that can account for the justification of punishment that also makes sense of why we think the crime is wrong in the first place. The benefits and burdens model fails to justify the institutions of punishment in a way that makes sense of why these violent or heinous crimes are, in fact, crimes.

Analysis of Wasserstrom's First Objection

Putting aside whether it is appropriate to describe heinous crimes as the misallocation of benefits and burdens, Wasserstrom's first objection appears to misunderstand how the benefits and burdens model is meant to operate in practice. He seems to take the model as a guide for how to distribute particular punishments for particular offences perpetrated by particular lawbreakers. When giving his argument for the first objection he does so in terms of the features of a hypothetical rape case, writing:

because [those who lack the inclination to rape] are not, therefore, burdened at all by the criminalization of rape, it is difficult to understand in what respect they have been unfairly burdened; nor is it easy to see the manner in which the rapist has unfairly benefited himself, *as against those who abstain from rape.*²⁴

The reasoning behind the objection focuses on what particular inclinations individual hypothetical persons might or might not have and how these inclinations create (or fail to create) burdens on these hypothetical persons.

This misunderstands the level of theorizing at which the model is meant to operate. Instead of a guide to particular laws and sentences Morris intends the benefits and burdens model as

²³ Wasserstrom, p. 144.

²⁴ Wasserstrom, p. 144.

a guide to constructing just institutions of punishment. He explains the comparison between his theory of retribution and rehabilitation as that between “two complex types of institutions,”²⁵ and in regard to the institution of punishment he sees the benefits and burdens model giving rise to “operative principles” that promote fair distributions of benefits and burdens.²⁶

Rather than assessing the ability of the model to address particular inclinations of particular people, we should assess its ability to model just penal and legal institutions. From a point of view antecedent to the token instances of lawbreaking the benefits and burdens model can make sense of the types of actions we count as relevant lawbreaking, and why we deem it appropriate to punish the offender. Without information about our own actual inclinations, motivations, desires, circumstances, we ask ourselves: what kinds of laws/punishments should we have? What kinds of things should we require people to restrain themselves from?

Life is endlessly complicated. In any particular situation there will be countervailing reasons to do one thing rather than another. The benefits and burdens model gives us an ideal that generates presumptions for conducting the particular cases in one way rather than another. The ideal may not hold in every single case. This is to be expected. The value of the ideal comes from how it can be used to guide the kinds of reasons and considerations we take up in the particular instances. The reasons that can and should come into play, and those that are irrelevant.

²⁵ Morris, p. 75.

²⁶ Morris, p. 76.

Wasserstrom's Reformulation of the First Objection

Wasserstrom anticipates a move that abstracts away from talking about particular instances of punishment and lawbreaking. Rather than discussing particular inclinations, the inclination at issue in this second order model “is [that] in each individual to do some of the things prohibited by the criminal law.”²⁷ Instead of providing an analysis of rape in particular, the second order model provides an analysis of lawbreaking in general. Wasserstrom describes the benefits and burdens in the general model as:

I restrain myself from doing those wrong things I am inclined to do and it is only fair that you restrain yourself from doing those wrong things you are inclined to do. You benefit from my law abidingness and I benefit from yours.

The relevant burden in this formulation of the model is just that incurred by restraining some inclination or other that would violate the freedom of others, not the burden of suppressing inclinations to violate each of the laws.

Wasserstrom argues that the move to a more abstract analysis of benefits and burdens loses the force of Morris' original claim. Explaining this point, he writes:

What began as a powerful appeal to a direct and obvious sense in which committing a particular crime burdened unfairly those who did not commit that crime and benefited unfairly those who did has been altered to become a far more abstract, more controversial appeal to the general benefits and burdens of law abidingness.²⁸

For Wasserstrom, the initial attraction of Morris' account is the ability of the model to make sense of why we have institutions of punishment in the first place. We punish because this person was burdened (unfairly) by that person. That person must now give up their unfair benefit.

²⁷ Wasserstrom, p. 145.

²⁸ Wasserstrom, p. 145.

Analysis of the Reformulation

Wasserstrom rightly anticipates a move to a more abstract understanding of the benefits and burdens model. The second order model more fully captures what Morris has in mind, as it captures, in part, the sense in which Morris is interested in justifying the institutions of punishment by appeal to a situation antecedent to any particular instances of lawbreaking or law-abiding behavior.

Wasserstrom's critique of this move rests on whether one is similarly attracted to the way in which the first order model can be applied directly to particular situations. I agree with Wasserstrom that in appealing to general benefits and burdens Morris loses some of the force of his argument. However, I do not think that it loses force simply in virtue of being a more abstract account.

The difficulty with this second order model comes from the difficulty of articulating what it could mean to have a general second order benefit or burden and what these benefits and burdens have to do with particular (especially heinous or violent) violations of persons. Although Wasserstrom only argues that the first order model has a problem accounting for how some horrible crimes can be explained in terms of benefits and burdens, I think this objection holds for the second order model as well.

The more abstract conception of benefits and burdens can help us to make sense of a move to institutions from particular instances, but there is still a problem describing all lawbreaking as simply a taking of an unfair benefit (whether it be particular or general). It is a problem with employing the concepts of benefits and burdens in the area of lawbreaking behavior. It seems that when especially serious or violent crimes are committed the unfair burden of suppressing inclinations when someone else has unfairly benefited by failing to

suppress some inclination is not a full account of the harm done to the victim or the community.

I think that a great deal of Morris' approach can be saved by taking a step back from the benefits and burdens model itself. It is important to remember that Morris does not start his justification of punishment with benefits and burdens directly. He begins with a community of free and equal agents. The dictates of fairness and the need to acknowledge the status of these agents as free and equal gives rise to the benefits and burdens model.

Focusing on the nature of the community that gives rise to the model in the first place may be able to help us retain some of the important parts of Morris' theory, while avoiding criticisms that Wasserstrom rightly levies against the particular model Morris describes. I will return to this suggestion in addressing the second of Wasserstrom's objections.

Wasserstrom's Second Objection

The second objection offered by Wasserstrom is that "it remains to be seen how it is that *punishing* the wrongdoer constitutes a taking of the wrongfully appropriated benefits away from him or her."²⁹ If the justification for punishment focuses on the allocation of benefits and burdens, it is most natural to turn to a system of restitution rather than retribution. After all, restitution or compensation of the victim appears to be a much more direct and natural way of setting everything right.

If we take the benefits and burdens model seriously, then it seems like the most important part of punishment is payment of a debt. Exchange of money (or other sorts of compensation) more straightforwardly pertains to the payment of debts than incarceration or other forms of harsh treatment. The monetary system is set up to calibrate prices, so it would be more

²⁹ Wasserstrom, p. 145.

practical to use it to settle the debts lawbreakers' incur in taking unfair benefits from the community.

Analysis of Wasserstrom's Second Objection

Wasserstrom's second objection illuminates a deep problem with providing an account of just institutions of punishment primarily in terms of paying what one owes. Although we frequently speak of lawbreakers "serving their time" or "paying their debt to society," there is no straightforward way to explain how harsh treatment through institutions of punishment constitutes payment.

Morris could give an argument against a system of restitution in place of a system of punishment in terms of what is required to respect the status of persons. Restitution could have the effect of literally putting a kind of price tag on criminal actions, which would not be respectful of persons.

Consider, again, the case of rape in the context of a theory that replaced a system of punishment with that of restitution. Instead of sentencing rapists to a period of incarceration we sentence them to pay the victim some amount of money. There is something unsettling about putting a monetary price on something like rape, or murder, or torture. In the first place, how could we possibly decide how much money the lawbreaker should pay the victim? Moreover, wouldn't the victim justifiably feel further violated by the justice system haggling over how much the lawbreaker should pay for raping her?

The unsettling nature of doing this kind of calculation comes from a tension between monetary price and the value of persons that motivates both the benefits and burdens model and the bigger Kantian framework. There is a difficulty reconciling the notion of paying a price for failing to respect someone's personhood and a commitment to the fact that each

person is valuable in virtue of being a person and has an inalienable right to be treated as such. In part, the difficulty comes from the fact that a price system admits of degrees, it is scalar. We can rank things according to how much they cost. The value of persons, on the other hand, is about respecting a certain kind of status. It does not admit of degrees in the way prices do.

Although Morris may be able to argue against a system of restitution, this argument is not in terms of the nature of the benefits and burdens model itself. His possible response is in terms of what ultimately gives rise to and justifies the model. This response is similar in kind to the one Morris gives to possible objections about the intelligibility of transferring one's right to be treated as a person through a contract.

The Social Equilibrium

Part of Morris' difficulty in providing a robust account of the relationship between respect for the status of persons and institutions of punishment comes from his focus on the individual lawbreaker when he explains his benefits and burdens model. This does not make full use of the resources he can draw from the concept of a social equilibrium. This equilibrium need not be explained solely in terms of a sort of moral recordkeeping. Instead, it can be grounded in the free and equal members of a community and the normative status of personhood that is needed to justify the institution of punishment.

When someone has perpetrated a crime, particularly a crime like rape, murder or torture there is justifiable concern or outrage on the part of the community. This reaction is not necessarily fully accounted for by how the lawbreaker has taken advantage of the other members of the community. The concern (sometimes outrage) can also be a response to how the lawbreaker violated the personhood of a particular member of the group.

The community Morris invites us to imagine requires that each member be taken into appropriate account in reasoning about how the group should govern themselves, and in the policies and institutions that come from this decision. The status of the individuals as reasonable, free and equal gives rise to constraints on how the group should act. The community has a *prima facie* legitimate interest in the violation of the status of some member(s) by other members of the group, as respect for personhood is the foundation for how the community has organized itself.³⁰

In claiming that the violation of someone's personhood disrupts the social equilibrium that makes communal life possible, I am using social equilibrium somewhat differently than Morris. I am not directly speaking about the (possibly) quantifiable measure of benefits and burdens that he describes. I mean to refer, instead, to what legitimates the fairness of a just distribution of benefits and burdens. I will use social equilibrium to refer to the extension of recognition respect for each individual and from each individual in a community of free and equal agents.

This use of the social equilibrium refers to an ideal state of affairs. This state of affairs is more than the situation in which every member of the community has, in some sense, paid off their moral debts. The ideal state is one in which all members of the community recognize and respect the status of the other members as persons. The social equilibrium is not, itself, a process that tends toward some equilibrium point, although legal and penal institutions can be legitimate with respect to it.

³⁰ It could be the case that the community has other interests (perhaps in maintaining as great a sphere of freedom as possible) that override its interest in each kind of disrespectful action that persons might do. We could take something like committing adultery to be an instance of disrespect towards one's spouse. In that it is disrespectful the community may have an interest in it. However, the community's interest in allowing people as much freedom as possible to live according to their own conceptions of the good may weigh more than their interest in this way of disrespecting the status of personhood when decisions are made whether or not to legislate against adultery.

The normativity of the ideal state of social equilibrium comes from that of which it is an idealization. The social equilibrium describes how roughly equal members of a community who are vulnerable to the interference of other members and have an interest in pursuing their lives as they see fit would relate to one another. Insofar as our community should take this perspective to be a guide for how we should relate to one another, we have an interest in pursuing the sorts of institutions that are legitimate for this idealized community.

Wasserstrom's objections illuminate weaknesses with Morris' conception of the maintenance of a social equilibrium as a the maintenance of a fair distribution of benefits and burdens. The weakness of the model leaves Morris' retributivist framework without a way to explain the relationship between instances of punishment and the social equilibrium.

I think that the value of Morris' retributivist framework need not be called into question by the weaknesses of his particular model. Stepping back to focus on how Morris uses the nature of a community of free and equal agents to justify his benefits and burdens model suggests an alternative understanding of a social equilibrium.

Taking the social equilibrium to be an ideal state in which members of the community extend recognition respect to all other members can be used as the first step in an alternative interpretation of how punishment is related to the social equilibrium. It is here that I believe Joel Feinberg's expressive theory of punishment can be helpful in articulating the value of and need for a system of punishment, and the connection between punishing the wrongdoer and the violation of the social equilibrium of the greater community.

Feinberg's Expressive Theory of Punishment

Wasserstrom's objections to Morris illustrate the need for someone working in the broadly Kantian tradition who accepts retributivism to give a robust account of the relationship between maintaining the social equilibrium of respect for and from each individual and the punishment of lawbreakers. I suggested at the end of the last section that Joel Feinberg's expressive function of punishment may provide a way to account for this relationship. In this section I will follow out this suggestion.

I will begin by explaining the fundamental aspects of Feinberg's positive account of the expressive function of punishment. Next, I will explore how the expressive function can be used to explain the relationship between the status of persons and institutions of punishment. I will give an analysis of condemnation in terms of apt resentment and reprobation. I will also address how legitimate vengeance is related to these reactive attitudes and the relationship between apt condemnation on the part of an individual and apt condemnation on the part of the community. After presenting the nature of the reactive attitudes, I will discuss what it means for the community to express these attitudes. Finally, I will return to the nature of condemnation to show how it is related to respect for personhood and the social equilibrium, and how it places constraints on legitimate legal and penal institutions.

Penalties vs. Punishments

Feinberg introduces the expressive function of punishment to account for a difference between penalties and punishments. He argues that the standard definition of punishment,

the infliction of hard treatment by an authority on a person for his prior failing in some respect (usually an infraction of a rule or command),³¹

is not sufficient, as it cannot distinguish between penalties and punishments. Both penalties (like speeding tickets) and punishments (jail time for a felony) are types of harsh treatment, imposed by authorized officials on an individual for some action that violates a rule or command. Whether or not this distinction is plausible, there seems to be at least a pre-theoretical difference between fining someone for speeding or illegally parking, and incarcerating someone for rape or murder. The standard definitions of punishment cannot account for this difference.³²

To replace this inadequate definition of punishment Feinberg searches for a characteristic shared by all forms of punishment that is not shared by penalties. The characteristic Feinberg identifies is the expressive function punishment serves. More specifically, punishment is

a conventional device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, on the part either of the punishing authority himself or of those “in whose name” the punishment is inflicted.³³

For Feinberg, something does not count as punishment unless it carries with it a judgment of disapproval and reprobation of the action of the lawbreaker by the community or the authorized group making decisions on behalf of the community.

³¹ Joel Feinberg. “The Expressive Function of Punishment.” In *Doing and Deserving*. Princeton: Princeton University Press. 1970. p. 95. Feinberg takes this definition from the writings of A.G.N. Flew, S. I. Benn, and H.L.A. Hart.

³² This difference may begin to dissolve if we make the monetary amount of the fine very high and the jail time for an offense very low. Perhaps if we were comparing a \$50,000 fine for speeding to a 30 day jail sentence for shoplifting, our intuitions would be different. Although I think this is an important direction to explore, the primary focus of my paper need not work out a defense of Feinberg’s distinction, as I am primarily interested in the expression function, not the role Feinberg, himself, thinks it should serve in theoretical accounts of punishment.

³³ Feinberg, p. 98.

Although Feinberg introduces the expressive function of punishment to make sense of the difference between penalties and punishment, we need not confine its use to this distinction. The usefulness of exploring the expressive function of punishment need not hang or fall with whether or not we find Feinberg's formulation of the distinction between penalties and punishments compelling. As I am interested in the kind of work an expressive function can do independently of Feinberg's distinction between penalties and punishments, I will now turn my attention to an analysis of the expressive function itself.

Condemnation

Feinberg characterizes condemnation as “a kind of fusing of resentment and reprobation.”³⁴ Condemnation is what is expressed by the conventional sorts of treatment that constitute punishment. He takes resentment to be an umbrella term for emotions and attitudes associated with legitimate vengeance.³⁵ Resentment expresses an understanding that the vengeance is appropriate or justified. Although Feinberg defines resentment in terms of vengeance, it is not clear that there is a necessary connection between them. I will return to this point below, after explicating what Feinberg, himself, means by resentment. Reprobation is different than either resentment or vengeance, in that it is a term for the stern judgment of disapproval itself.

Reprobation is apt when we are in a position that allows us to make an informed judgment about the action of the other person. The judgment can only be legitimate if we have good reason to believe our judgment is correct. The action must also be something that is worthy of disapproval. It is the kind of thing that members of the community have a right to know about, and have the right to form judgments of approval or disapproval about.

³⁴ Feinberg, p. 101.

³⁵ Feinberg, p. 100.

Resentment on the other hand, has less to do with the action, and more to do with *who* performed the action. Take a non-criminal example: your rude colleague. Why do we think that it is appropriate to dislike and want to get back at (or hope that karma catches up with) a rude colleague? I suspect that we think resentment is appropriate in this case because we think the rude colleague should have known better. As colleagues you and she share (roughly) equal standing at your place of work. Moreover, the nature of collegial relationships is such that you should both feel and express a certain degree of respect for one another. If you view her in this manner and your actions reflect this fact, you feel justified in getting upset if she does not reciprocate this respect. After all, it is what you are due by your equal status and the expectations inherent in the colleague-relationship.

The “getting back at” part of resentment is appropriate because your colleague’s rudeness is a way of denying the nature of your relationship. The relationship is one of equality and respect. If she does not reflect this, she is denying that you are an equal footing. In a sense, she is holding herself above those she is rude to – she is making an exception of herself. Wanting to get back at her can be (in best case scenarios) a warranted desire to re-level the playing field. It is a desire to reestablish the appropriate relationship between the parties.

This is not to say that any action, whatsoever, can be done in the name of reestablishing equality. Only those actions that respect the colleague relationship are warranted as it is this relationship that was breached by the rude colleague in the first place. If respect for your colleague is violated by how you react to her, how “appropriate respect” can sanction the initial attitude of resentment is called into question. If respect and equality are what fundamentally justify the reactive attitudes what underwrites them should remain central to how one seeks recourse.

Contrast the rude colleague with a rude child under the age of five. What has changed with this example? Most obviously, there is a different relationship between you and the child. The equality between the parties in the previous example has disappeared. The expectations of the relationship have also shifted. You might feel as though the child owes you respect, but as the child cannot be thought of as fully culpable for her actions, you do not hold her responsible for knowing this fact.

You may still judge her to be responsible for her behavior in some sense, but the actions sanctioned by this judgment are different than in the colleague example. You may want to remind the child about basics of etiquette, or give her a time out so that she learns to associate being rude with unpleasantness. But if you feel resentment towards the child, something has gone wrong.

Resenting a rude child is to make a mistake about what children are. Resentment is appropriate for people who should have known better, the rude child does not fall under this group. Moreover, there is no sense in which you are justified in trying to re-level the playing field. You and the child did not start on equal footing, so desiring the sort of vengeance aimed at “making everything square” is completely mistaken.

Vengeance

There are two important objections to my presentation concerning the nature of legitimate vengeance. The first is that there is no necessary connection between resentment and vengeance. These two things can (and often do) come apart. The second objection is that when you have the attitude of vengeance (even when its legitimate) it does not seem to be aimed at making everything equal between yourself and the person who wronged you. It

seems to be aimed at making the person who wronged you lower than you are. Let me address these objections in order.

The first objection is that there is no necessary connection between vengeance and apt resentment. At the very least it is intelligible to say that the feeling of resentment comes apart from the desire for revenge. One can imagine a case in which you have been wronged or disrespected by another person and resent them for doing so, but feel no desire to do something about it. You feel wronged and slighted, but are simply not motivated to get back at the person who wronged you.

It is more difficult to imagine a case in which you have been wronged by another person, but do not think you have a *reason* to do something about it. You could make the judgment that the slight to your status as a person is not worth the time and energy it would take to address the situation. However, this is just to say that you do not have an all-things-considered reason to do something about it. That does not imply that there is not *some* reason counting in favor of seeking recourse. It merely shows that your reason to do something is not overriding. In fact, if the resentment you feel towards the person who slighted you is appropriate, this implies that there is some legitimate reason for feeling it: she has disrespected your status as a person. This reason can be, and most often is, the same reason one would take to be an overriding reason in a case of legitimate vengeance.

Along with distinguishing the desire for revenge from a reason to seek recourse, it is important to differentiate private vengeance from the kind of legitimate vengeance that Feinberg believes is expressed in punishment. I would like to differentiate them by talking about what an individual has reason to do and what the community has reason to do. There may be a case in which resentment is legitimate but the individual does not have, all things

considered, reason to get back at the person who has wronged her. This does not imply that the community does not have a reason to, in some sense, get back at the person.

Imagine a situation in which someone has violated my status as a person by breaking into my house and stealing valuable possessions. I may personally feel no desire to get back at the person, or I could make the judgment that it is not worth my time or energy to do so. This does not mean that the community has no reason to punish the individual and thereby express legitimate vengeance. The fact that the lawbreaker was disrespectful of my status as a person is reason for the community to be concerned, as the social equilibrium, as I have described it above, is a concern of the entire community. The equilibrium has been disrupted, thus the community has some interest in responding to the perpetrator.

The second objection to incorporating vengeance into penal institutions concerns the aim of vengeance. I have characterized resentment (and the sort of legitimate vengeance that can be associated with resentment) as the desire to re-level the playing field or make everything square. However, it seems that when someone wants revenge her desire is not to make everything even, but instead, to make the object of revenge lower than herself. My response to this objection is similar in form to my response to the first objection: it is important to differentiate private vengeance from the legitimate vengeance expressed in punishment.

There are all manner of desires and inclinations that individual people may feel in response to being wronged. A private desire to “make low” is among them. This reaction, on the part of the community, is not appropriate. Not only is it difficult to understand how a community could have a desire, the community is committed to the extension of respect for and from each individual as it makes communal life possible. If some member of the community has wronged another member by disrespecting her personhood, then the community can have a

valid interest in reacting to the situation. However, the commitment to the social equilibrium acts as a constraint on what counts as an appropriate reaction on the part of the community.

The sort of legitimate vengeance that the community has an interest in expressing cannot be to make someone low, because this would violate the community's foundational commitment. Any conventional device (perhaps harsh treatment, incarceration, etc) that expresses legitimate vengeance on behalf of the community must respect the personhood of the punished. Failing to do so would not be appropriately aimed at reestablishing this social equilibrium between persons.

I have made much of the difference between an individual experiencing a reactive attitude like resentment and the community expressing something like legitimate vengeance. To understand this distinction we must understand not only what it means to express something like vengeance, but it requires an explanation of how the community can be said to express something like a reactive attitude. I will turn to the second of these problems before returning to the first.

Reactive Attitudes and the Community

My explanations of vengeance require an account of the connection between the personal reactive attitudes and how social institutions can manifest or express reactive attitudes. I will draw from Peter Strawson's account of the reactive attitudes in "Freedom and Resentment"³⁶ to build this connection. Strawson defines the reactive attitudes (like resentment) in the following way:

The reactive attitudes...are essentially reactions to the quality of others' wills towards us, as manifested in their behavior: to their good or ill will or indifference or lack of

³⁶ Peter F. Strawson. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*. London: Methuen & Co Ltd. 1974

concern. Thus resentment, or what I have called resentment, is a reaction to injury or indifference.³⁷

Resentment seems, in the first instance, to be deeply personal as it is a reaction to how someone treats me.

To build the connection between individual reactive attitudes and communal expression of reactive attitudes, we need first to move from my reactions to how I am treated to reactions to how someone else is treated. If we are to give an account of the communal expression of attitudes, it seems that I need to be able to talk about my having a reactive attitude toward how Sarah treats Myles, where I am neither of the parties involved.

Strawson refers to these third person attitudes (those that do not concern how *I* am treated) as sympathetic or vicarious generalized analogues of the reactive attitudes. These attitudes are “reactions to the qualities of others’ wills, not towards ourselves, but towards others.”³⁸ Strawson argues that the personal reactive attitudes are “humanly connected” to the vicarious reactive attitudes. By this he means that they “have common roots in our human nature and our membership of human communities.”³⁹

Identification of the generalized analogues of the reactive attitudes gives us third personal analogue attitudes of particular individuals. We can assess the appropriateness of these attitudes in much the same way we assessed the aptness of personal resentment and reprobation. I legitimately have this vicarious attitude when what Sarah does to Myles manifests inappropriate ill will, indifference or lack of concern. The appropriateness of the generalized analogues of the reactive attitudes will also be related to those things in which a

³⁷ Strawson, p. 14.

³⁸ Strawson, p. 14.

³⁹ Strawson, p. 16.

third person has a legitimate interest. This may change depending on the nature of the third person's relationship to the two parties. If I am Myles' friend I may have a greater scope of justified interest in how he is treated by Sarah. If he is a stranger, my scope of justified interest may be smaller.

If we focus exclusively on the interest I can justifiably take in how Sarah treats Myles when they are both strangers, there is a straightforward way to explain how their interactions are of interest. Rather than talk of a particular third person observer, we can talk of a somewhat idealized third person observer. We can make this move because the broadly Kantian framework I am working in claims there is no salient difference between me and some else insofar as we are members of the community.

This third person is ideal in that she is a free and equal member of the community who has an interest in preserving her freedom and maintaining her equal status as a person. This status is empty without means by which she can exercise her freedom and equality within her community. So, this idealized third person will be interested in how Sarah treats Myles to the extent that how Sarah treats Myles violates the rules, institutions and principles that makes it possible for the third person to exercise her freedom and maintain her equal status. That is, she is interested in those actions Sarah might do that violate the social equilibrium as I have defined it.

One might object that moving from an interest in reacting or responding to someone who shows ill will to others to an interest in maintaining a social equilibrium of respect for and from each member of the community is too quick. I would like to suggest that although Strawson characterizes what underwrites the personal reactive attitudes in terms of demanding or expecting "the manifestation of a certain degree of goodwill or regard on the

part of other human beings towards ourselves,”⁴⁰ this description captures what Darwall means by recognition respect. Saying that I have a reasonable demand on others to treat me with a minimal degree of goodwill, or at least not to regard me with ill will, is one way of explaining what it could mean to demand that others take my status as a person seriously in their practical deliberations.

Understood this way the generalized or vicarious analogues of the personal reactive attitudes are a demand that all persons be treated with recognition respect. One way to understand what this means is found in Strawson’s account of what underpins the generalized reactive attitudes:

the demand for the manifestation of a reasonable degree of goodwill or regard, on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt, i.e., as we now think, towards all men.⁴¹

In line with my suggestion that the personal reactive attitudes are a demand that I be treated with respect, Strawson’s characterization of the generalized reactive attitudes as a demand for the manifestation of a reasonable degree of goodwill or regard towards all men is one way of explaining the relationship of recognition respect for and from each member of the community which is, in part, how I have defined the ideal state of the social equilibrium.

We are now in a position to make the connection between the idealized third person and social institutions, on our way to explaining the connection between individual personal reactive attitudes and the expression by communal institutions of reactive attitudes. Several traditions in political philosophy have stories about how this connection works. However this goes in the details, be it a form of Social Contract Theory or Rawls’ Original Position, the stories involve some mechanism for safeguarding the freedom and equality of the members.

⁴⁰ Strawson, p. 14.

⁴¹ Strawson, p. 15.

This mechanism reflects the concerns than an ideal person would have in creating these institutions – namely an interest in maintaining what I refer to as the social equilibrium of respect for and from each member of the community.

For the most part this mechanism takes the form of the institutions of criminal law. There are several ways legal and penal institutions can reflect respect for persons. I have outlined some of these in Section I of this paper. Requirements of due process, prohibitions on punishing the innocent, and the requirement that one be shown to be guilty beyond a reasonable doubt are some of the policies that reflect the appropriate respect for persons. I will return to these institutional requirements below in discussing how apt condemnation can serve as a constraint on legal and penal institutions.

These institutions are what ideal persons would agree to and serve as the kind of disinterested third person I described above. These institutions are such that they incorporate and promote the interests of ideal disinterested third persons, and create and preserve the conditions under which persons can exercise their distinctive capacities. Insofar as they are constituted such that they preserve and promote these interests, there is a sense in which they express the generalized analogues of the reactive attitudes of the disinterested ideal third party. This does not require that the community, itself, possess those attitudes, only that the community contain institutions that serve to express the attitudes of an ideal free and equal member. This claim requires that I say more about what I mean by “expression.” I will turn to that now.

What Expression Is and What Expression Is Not

My explanation of the communal expression of generalized analogue reactive attitudes requires an explanation of the nature of expression. By expression I mean the manifestation

in action of a principle, commitment, or belief. I do not want to explain expression in terms of an emotional or affective state, nor an act of communication. These preliminary remarks do not do much to illuminate how a principle, commitment or belief is manifested in an action. Before moving on to what role the expressive function plays in legitimate institutions of punishment, I will attempt to make this notion of expression more clear.

Let me begin by addressing the emotional element of expression that many people appeal to in explaining the nature of expression. The sense in which social institutions express something in action is not tied to some affective state. I do not know of an intelligible way of explaining how a group of people can have an emotional state that does not reduce to some account of the emotional states of all or most of its constituent members. The sense in which the community expresses a reactive attitude is related to how the community's institutions reflect certain principles, commitments and beliefs.

Consider a case in which the legal and penal institutions, at least partially, function as a way for the community to disavow certain kinds of actions. There are some actions that conscientious citizens may deem criminal, but that the government does not legislate against. For example, until 1976 there was a provision in the rape laws called the Marital Rape Exemption, which made it impossible for any husband to be charged with the rape of his wife.⁴² That is, a husband forcing his wife to have sex against her will was not considered rape and could not be prosecuted under the criminal law.

If we, as the community, react to this as a miscarriage of justice we do not necessarily feel a sort of "frustrated *Schadenfreude*"⁴³ toward the perpetrator. To demand the punishment of

⁴² Wellesley Centers for Women, "The Wife Rape Information Page," <http://www.wellesley.edu/WCW/mrape.html>.

⁴³ Feinberg, p. 103.

husbands who force their wives to have sex against their wills may represent the belief that these actions should be condemned, or a commitment to the sexual autonomy of all members of the community, based on a principle of respect for persons that holds that regardless of the relationship between the parties someone cannot force another person to have sex against that person's will.

The demand for instituting a legal prohibition on marital rape could represent the community's commitment to punishing those who violate the personhood of others. Demands for a change in the law can be understood as a demand that the legal institutions *go on record* by recognizing that spouses can wrongfully force their partners to have sex, and the wrongness of this sort of action should be testified to in the law.⁴⁴

The communal demand for the government to manifest the commitment to respect for persons by legislating against marital rape, is more than a demand that the legal and penal institutions communicate the wrongfulness of forcing one's spouse to have sex against her will. Elizabeth Anderson and Richard Pildes argue that communicative acts are only a subset of expressive acts. They define communicative acts in terms of an agents' intentions:

To communicate a state of mind is to act with the intention of inducing others to recognize that state of mind by recognizing that very communicative intention.⁴⁵

I would add to this definitions the requirement that communicative acts be successful in getting others to recognize one's communicative intention. This more accurately reflects what we mean by communication and further limits the scope of the subset of communicative acts within the group of expressive acts.

⁴⁴ Feinberg, p. 103.

⁴⁵ Anderson and Pildes, "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review*, May 2000, Volume 148, Issue 5, p. 26.

When we punish someone for disrupting the social equilibrium of respect for and from each member of the community we are manifesting our value of this equilibrium in action. The success of expressing this commitment is not measured by whether we actually communicate this state of mind to any designated members of the community. If doing so actually has the effect of communicating this value, this can be a positive result, but it is not a necessary one for successful expression.

The Expressive Function's Role in Legitimate Institutions of Punishment

It is important to begin by noting that the expressive function of punishment does *not* justify the institution of punishment by itself. The justification for the institutions comes from the legitimacy of viewing a community as a group of free and equal moral agents and from the validity of taking a community to value the roughly equal and free standing of its members. If communities are not like this, then punishment (as I have described it) would not be justified.

Expression of condemnation or respect for persons in punishment is not a goal or a justified aim of institutions of punishment. Rather, theories of apt expression tell us that when we are speaking, reasoning, and acting we should do so subject to certain regulative constraints. If we are conscious of expression in punishment we do not think that the goal of punishing is to maximize the expression of condemnation in the community. Instead, the proper expression of condemnation and respect for personhood, as described above, are both constraints on how we go about pursuing the end of justified punishment.

An explanation of the kinds of constraints the appropriate expression of condemnation (and thus, respect for personhood) places on institutions of punishment would serve as an account of the relationship between the status of persons and the institutions of punishment

that Morris needs in response to Wasserstrom's objections. In the following section of this paper, I will explore ways in which the proper expression of condemnation places constraints on the kinds of penal institutions a community adopts, in hopes of sketching the kind of account I was in search of at the beginning of this section.

Condemnation and Constraints on Institutions

If the characterizations I have given above are apt, then resentment is only justifiably felt toward persons and reprobation is a way of expressing respect for an agent's personhood. They are intricately tied to the relationship of freedom and equality between citizens under the law. The institutionalization of the generalized analogues of these personal reactive attitudes in punishment can reflect the necessary sort of respect for the parties involved. Let me explain how this is the case.

If punishment is defined in part by the expression of condemnation, then punishment is reserved for only the appropriate recipients of reprobation and resentment. Appropriate resentment is felt by those who stand in a relationship of freedom and equality to one another. There is an implicit acknowledgement of the personhood of those punished by the very fact that they are subject to punishment, as being subject to punishment requires that one be a free and equal member of the community.

That punishment be of persons and express recognition respect for the status of both the victim (if there is one) and the lawbreaker as persons gives rise to a host of constraints on legitimate institutions of punishment. Only those who are fully free and equal members of the community can be found criminally liable. This legitimizes exemptions for the mentally ill, children and others who do not have the capacity for rational and reasonable decision

making. It also requires that the social institutions under discussion be those of punishment rather than rehabilitation or therapy (as discussed in Section I).

Reprobation is not directly about who someone is, but about what is a legitimate sort of judgment to make about them. In the context of the community we have been discussing it places constraints on the policies of the sorts of institutions and kinds of laws of the community. The first aspect of apt reprobation is the requirement that the person feeling reprobation have good reason to believe her judgment is correct. This requirement underpins the criminal trial system, in that the criminal judicial system attempts to assure the fact that those instituting punishment have good reason to believe they are correct in their judgments.

This is reflected most acutely in the United States' legal system by the burden placed on the prosecution to show that the accused is guilty beyond a reasonable doubt. To convict someone as a criminal, the jury (and by extension the rest of society), must determine whether they have good reason to do so. Moreover, this is good reason in a non-consequentialist sense. The prosecutor does not attempt to show that (beyond a reasonable doubt) sending someone to jail will result in better consequences. Rather, they attempt to show that (beyond a reasonable doubt) the accused performed the action of which she is accused.

In addition to the safeguards of the trial system to justifiably express reprobation institutions must have procedures that take into account only what is, in principle, knowable. This would prohibit judgments based on the intrinsic moral character of an individual. It would prohibit punishing someone for being a "bad person" rather than for something they actually did. The restriction of the realm of criminality to what is knowable carries through to

legislation as well. Legislation that takes into account the intrinsic moral worth of some person will be similarly prohibited.

The second requirement of apt reprobation is that the action be worthy of disapproval. This requires that the actions we deem criminal be legitimately open to scrutiny by the community. This requirement plays a central role in the defense of the right to privacy. Those who believe in a right to privacy argue that there are things that the government has no right to know about, let alone a right to disapprove of. For example, under a system that took seriously the requirements of apt reprobation, anti-sodomy laws would most likely be prohibited, because citizens' sex lives are not (usually) the sorts of things that are apt objects of either approval or disapproval by the government.

Apt reprobation can also play an important role in a defense of some of the United States' First Amendment rights. Part of the motivation for instituting freedom of religion, speech and association, is an acknowledgement that what the amendment protects should be beyond the scope of the community's scrutiny. For example, it is not appropriate for the community to criticize (or even take an interest in) a citizen's religious affiliation.

Conclusions and Future Directions

Having laid out some ways the proper expression of condemnation can serve as a constraint on the kinds of penal institutions a community adopts, I would like to conclude by way of describing how the retributivist-expressive framework I have laid out for understanding the institution of punishment can be a fertile ground for future research.

The point of view prescribed by the retributivist-expressive framework I have described can be a useful tool for analyzing specific policies and institutions. For instance, states like California and Texas have recidivism statutes that give mandatory life sentences to people

who have committed multiple felonies. The details of such statutes are too complex to address in this paper, but I would like to suggest that they can (and should) be discussed in terms of respect for persons. The justification for these sorts of statutes often cites the dispositions of recidivist offenders. If making judgments about dispositions concerns not the moral status but the moral worth of the lawbreaker, these policies would be deemed illegitimate by my account.

I have been working within a broadly Kantian framework in order to address particular worries that arise within it concerning the justification of punishment. To the extent that account is plausible, it removes a potential source of worry that this framework cannot give a viable justification for the institution of punishment. To the extent that the picture I have offered is independently plausible as an account of the legitimacy of penal institutions, it may provide an independent source of support for the broadly Kantian framework.

Works Cited

- Elizabeth Anderson and Richard Pildes, "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review*, Volume 148, Issue 5 (May 2000).
- Steven Darwall, "Two Kinds of Respect," *Ethics* Volume 8, Number 1. (Oct. 1977).
- Joel Feinberg. "The Expressive Function of Punishment." In *Doing and Deserving*. Princeton: Princeton University Press. 1970.
- Immanuel Kant. *The Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge Texts in the History of Philosophy Series. Cambridge University Press. 1996.
- Karl A. Menninger. *The Crime of Punishment*, New York: Viking Press, 1969.
- Herbert Morris. "Persons and Punishment," In *Punishment and Rehabilitation*. 3rd Edition. Edited by Jeffrey Murphy. Belmont, CA: Wadsworth Publishing Company. 1995.
- Peter F. Strawson. "Freedom and Resentment." In "*Freedom and Resentment*" and *Other Essays*. London: Methuen & Co Ltd. 1974
- Richard Wasserstrom. *Philosophy and Social Issues*, Notre Dame: University of Notre Dame Press, 1980.