

For a link to this article, please click:

<http://www.tandfonline.com/eprint/sJ2JBVXkztyFMGmxS7tS/full>

Contractualism and Punishment

Hon-Lam Li

T. M. Scanlon's contractualism is a meta-ethical theory that explains moral motivation and also provides a conception of how to carry out moral deliberation. It supports non-consequentialism – the theory that both consequences and deontological considerations are morally significant in moral deliberation. Regarding the issue of punishment, non-consequentialism allows us to take account of the need for deterrence as well as principles of fairness, justice, and even desert. Moreover, Scanlonian contractualism accounts for permissibility in terms of justifiability: An act is permissible if and only if it can be justified to everyone affected by it. This contractualist thesis explains why it is always impermissible to frame an innocent person, why vicarious punishment is impermissible, and why there has to be a cap on sentences. Contractualism therefore allows us to take deterrence as a goal of punishment without the excess of utilitarianism. This paper further argues that the resulting view is superior to pure retributivism. Finally, it shows why legal excuses and mitigation can be justified in terms of the notion of negative desert.

Keywords: contractualism, non-consequentialism, punishment, deterrence, fairness, desert, negative desert, vicarious punishment, legal excuses, mitigation.

T.M. Scanlon's contractualism¹ is a meta-ethical² or second-order theory that explains permissibility in terms of justifiability: An act is permissible if and only if it can be justified to everyone affected by it. Thus understood, an impermissible act is one which we have a reason not to perform insofar as we wish to govern ourselves on terms that others could not reasonably reject and insofar as they are motivated by this ideal.³ Because we care about whether our acts are justifiable to others on grounds that they cannot reasonably reject, this theory provides an explanation as to how moral motivation is possible.

Moreover, Scanlonian contractualism also provides a plausible conception of how to carry out moral deliberation and how to resolve first-order moral issues.⁴ This

conception supports non-consequentialism, the theory that both consequences and deontological considerations are relevant factors to be taken into account in moral deliberation.

While Scanlon's contractualism is a form of hypothetical contractualism, the best-known contemporary version of hypothetical contractualism is John Rawls's argument in the Original Position.⁵ Despite its importance, Rawls's theory is not a suitable basis on which to ground morality. First, Rawls intends his theory to be one of social justice. The subject matter to which Rawls's theory applies is the basic structure of society, not moral actions as such. Second, whereas according to Rawls's theory principles of justice distribute "primary social goods," it is doubtful whether there are "moral primary goods."⁶ Thus, even if Rawls's theory is correct, it is too limited to ground all of morality.⁷

I shall put forward a partial theory of punishment within the framework of Scanlonian contractualism.⁸ There are two reasons for doing so. First, Scanlonian contractualism supports non-consequentialism⁹ -- the theory that both consequences and deontological considerations are relevant in moral deliberation. Because non-consequentialism can accommodate both consequences and deontological considerations, a non-consequentialist theory of punishment has room for considerations of deterrence, as well as for those of fairness, justice, and even desert.

Another reason is that Scanlonian contractualism has implications for first-order moral problems in general. Scanlonian contractualism is committed to the fundamental thesis that an act is morally permissible if and only if it is justifiable to everyone affected by it.¹⁰ This thesis has direct implications for punishment, because it sets limits on who should be punished and how much, and has relevance to the

question of whether certain forms of punishment are too cruel or unreasonable and should be excluded from the criminal justice system.¹¹

In what follows, it will be useful to keep track of two different types of arguments. At one level, the contractualist thesis that an act must be justified to all parties concerned supports various arguments on punishment. Yet, Scanlonian contractualism provides room within a non-algorithmic framework for exploring first-order arguments on punishment as well. Because such a kind of non-algorithmic exploration involves *phronesis* (practical wisdom), it is possible that two contractualists sharing the same meta-ethical framework might arrive at different first-order views on punishment.

1. Contractualism: Introduction

Scanlon strives to defend claims of individuals against aggregative reasoning. A utilitarian, for instance, is committed to the thesis of aggregation, that an individual's life could be outweighed by trivial claims (for example, to alleviate headaches) held by other people, if their number is sufficiently numerous. It seems, however, that an individual's life cannot be outweighed by other people's headaches, no matter how many people there are who want to have their headaches alleviated. Yet, a utilitarian would disagree.¹²

Confronted with this problem, Scanlon points out that we must first understand why normative (or first-order) utilitarianism still has its adherents despite its various highly counter-intuitive implications.¹³ This phenomenon, Scanlon observes, has to do with the intuitive appeal of the meta-ethical (or second-order) foundation of utilitarianism – “philosophical utilitarianism,” which is a philosophical thesis about the subject matter of morality, namely that the only fundamental moral

facts are facts about individual well-being. In order to undermine normative utilitarianism, Scanlon suggests, we must sap the force of philosophical utilitarianism by offering a more plausible alternative meta-ethical theory. Scanlon's contractualism provides such an alternative. As a meta-ethical theory, it informs us about what moral wrongness or impermissibility consists in.

To arrive at contractualism, two steps are necessary. First, the reference to moral wrongness has to do with the fact that we would feel remorse or self-reproach if we do something that we regard as wrong.¹⁴ While an action may be wrong for different kinds of reason, the inner core of morality has to do with "what we owe to each other" – the realm of morality in the narrow sense governing the kind of reasonable behavior that we expect of each other insofar as we all agree to treat others in a reasonable way. Morality in the broader, "impersonal," sense on the other hand, is concerned with impersonal grounds – or "reasons that are not tied to the well-being, claims, or status of individuals in any particular position."¹⁵ Morality in this broader sense covers wrongness relating to non-persons, such as non-human animals, trees, natural wonders, as long as people's well-being or interest is not affected.¹⁶ It is, however, morality in the narrow sense that is crucial to our relation with each other.¹⁷

The second step is to find the best account for wrongness (in the narrow sense) or impermissibility. Scanlon argues that utilitarianism does not account for our feeling remorse for failing to aid the needy in Bangladesh, for instance, because failing to maximize utility by itself does not account for our feeling remorse.¹⁸ Rather, we feel remorse because our failing to provide aid is unjustifiable, and our inaction is disallowed "by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement."¹⁹

Scanlon's strategy is analogous to the strategy of a chemist who explains why the theory that gold has the atomic number 79 is more successful than rival theories in accounting for the phenomenal properties of gold, such as being yellowish, heavy and resistant to rust.²⁰

Now suppose we are interested in finding out whether an act, *A*, if performed in the circumstances is morally permissible or not. To proceed, we need to find out whether there exists any principle which no one could reasonably reject, and which would disallow *A*. So we need to consider principles one by one, and there are an indefinite number of them.²¹ If there is a principle that no one could reasonably reject and that would disallow *A*, then *A* is impermissible.²²

In going through various principles, how do we know whether a principle, *P*, is reasonably rejectable or not? A principle is the "general conclusion about the status of various kinds of reasons for action."²³ In Scanlon's model, not all kinds of reasons have weight. The reasons must, firstly, be "generic reasons," or reasons based on generally available information of what individuals would want in certain situations, characterized in general terms, rather than detailed and particular information (which we do not usually know).²⁴ This is so because we want to be able to settle questions of right and wrong in the abstract, since the particular identity of the individuals affected is irrelevant.²⁵ Secondly, the generic reasons must be "*personal* (generic) reasons," which "have to do with the claims and status of *individuals* in certain positions."²⁶ Impersonal reasons fall outside the moral realm of "what we owe to each other."²⁷

To appeal to individuals' personal (generic) reasons, we would need to know, firstly, how *P* would affect different individuals' well-being, even though well-being

is not the only thing that matters.²⁸ But Scanlon's way of understanding well-being departs from a consequentialist or welfarist model. For one thing, "reasons for being concerned with how one is 'affected' by a principle are not limited to reasons having to do with one's welfare," such as one's wanting to be able to "determine by one's choice whether some result is likely to occur."²⁹ For another, there are reasons arising from the harmful effect caused if a certain type of action is generally permitted. It might be that, as in a case of environmental damage, "no individual action on its own causes a harm to which anyone could object,"³⁰ but that such kind of action is still impermissible because of the consequences, should it be generally permitted and performed.³¹ Secondly, a different basis on which individuals could object to a principle is that it is unfair, unjust, or rigged.³² Thus, principles should be set aside if they violate valid considerations of fairness, justice, or perhaps even desert, as well as other deontological moral principles.

Scanlon's contractualism is often misunderstood. One common misunderstanding is to interpret it as a first-order or normative theory and then to claim that what makes torturing babies wrong, for instance, is not the act's being reasonably rejectable, but rather its being a cruel act.³³ Thus interpreted, Scanlon's contractualism would indeed be viciously circular and redundant. Such objections, however, arise from a misunderstanding of Scanlon's view, for Scanlonian contractualism is not a first- but a second-order or meta-ethical view.³⁴

Another common objection compares Scanlon's theory with Rawls's, and then claims that whereas Rawls can derive his two principles of justice without being vague or circular, Scanlon lacks a clear criterion for deciding whether a "principle" is reasonably rejectable or not.³⁵ As Scanlon himself points out, Rawls can derive his

principles of justice only because a number of moral presuppositions are built into the design of the Original Position.³⁶ Those who raise this objection against Scanlon fail to realize the similarities of this part of Scanlon's contractualism with virtue ethics,³⁷ both approaches being receptive to the idea that an algorithm for resolving ethical issues does not exist.³⁸ In deciding whether a principle is reasonably rejectable, a contractualist should employ her practical wisdom (*phronesis*) or "practical judgment."³⁹

2. Rawls and Scanlon

For a contractualist, justification is built upon an understanding of impartiality. One important difference between Rawls (and John Harsanyi) on the one hand, and Scanlon on the other, has to do with their different understandings of the idea of impartiality.⁴⁰ Rawls believes that he could arrive at normative principles if self-interested parties in ignorance of who they are would adopt the same principles in an attempt to maximize their own self-interest. In trying to obtain normative principles, both Rawls and Harsanyi need to adopt some interpretation of ethical impartiality. Harsanyi interprets this condition as the one in which all parties (beyond the veil of ignorance) would have an equal chance of being anyone in society.⁴¹ Rawls, on the other hand, interprets this condition as the one under which self-interested parties are deprived of particular information, including any objective basis for estimating probabilities.⁴²

Yet, as Scanlon points out, the supposition – common to Harsanyi and Rawls – that we could obtain normative principles from self-interested parties under conditions of ignorance trying to maximize their self-interest involves a covert transition from the plausible contractualist idea that we must be able to justify our acts

to everyone concerned to the problematic idea that what is justifiable to everyone concerned is what would maximize an individual's self-interest behind the "veil of ignorance."⁴³ Why is this supposition is problematic? The short answer is that we cannot justify the situation to those who lose out in society ("the losers") on the grounds that average utility is high. Scanlon rightly points out that the fact that one option promotes higher average utility than another one does not settle the matter as to which option is morally justified.⁴⁴

A more detailed answer is that the kind of impartiality required by contractualism must ensure that a course of action be justified to every individual affected, no matter who he or she is. But Harsanyi interprets this condition as the one under which all parties would have an equal chance of being anyone in society. This interpretation, Scanlon rightly points out, is mistaken. For the losers can reasonably reject any principle that purports to justify their much worse condition by reference to the fact that others are better off (and hence average utility is higher).⁴⁵ Scanlon thinks that the problem lies in the possibility that a much worse condition suffered by a few can be justified by the fact that many people enjoy a somewhat higher level of well-being. He believes that his contractualism can avoid this problem of aggregation, and that the origin of the problem is the supposition common to Harsanyi and Rawls mentioned above.

In addition, Rawls seems to have no strong justification for suppressing people's moral views in the Original Position. In contrast, in Scanlon's theory those contracting hypothetically to reach an agreement have access to moral reasons. These moral reasons are not just what the contracting parties' moral beliefs happen to be; rather, these are moral reasons as they objectively are. Thus, the "veil of ignorance" is

unnecessary in Scanlon's theory. Although Scanlon is clearly correct, I think that some version of the "veil of ignorance" behind which the contracting parties do not know who they are – though they have all other information, as well as the capacity to carry out moral deliberation – may be a useful *heuristic* device in some cases. **After all, the function of the "veil" in eliminating riggedness, unfairness, and injustice from the process of choosing a reasonably nonrejectable principle is similar to that of Scanlon's restriction of reasons to "generic reasons" alone.** Other ways to avoid riggedness, unfairness and injustice include the idea of "changing places" and the Golden Rule.⁴⁶

3. Non-consequentialism

What kind of normative or first-order theory would cohere with Scanlonian contractualism? It should be clear from the discussion above that the answer is non-consequentialism -- the theory that both consequences and deontological principles are significant considerations in moral deliberations. The kind of non-consequentialism that coheres with contractualism is, however, not just any version. Robert Nozick's entitlement theory, Scanlon points out, is a version of non-consequentialism.⁴⁷ Yet Scanlonian contractualism would not endorse the entitlement theory.

Before I explain the way in which Scanlonian contractualism enables us to deal with difficult issues in punishment, we have to be clear that there are two different types of questions that intersect and intertwine. Questions of the first type are first-order "local" issues, such as why we should have the institution of punishment at all. To answer such questions, we need to use our judgment with the aid of practical wisdom (*phronesis*). Adopting the correct broad theory is not

sufficient for resolving “local” issues. This is so in all areas of practical ethics, and not only with respect to the problem of punishment. For instance, even if we happen to adopt the correct “broad” approach (whether it is contractualism, utilitarianism, or virtue ethics), we cannot resolve the problem of animal ethics unless we also have the solution to the “local” issue of moral status (e.g., whether speciesism or its denial is correct). Questions of the second type have to do with our “broad” approach. Having the correct broad approach may be necessary, if not sufficient, for resolving a moral problem.

Not only do I accept contractualism, I also hold non-consequentialism, which implies that both consequences as well as deontological notions (such as justice, fairness and desert) are important for a theory of punishment. I shall argue that one important component in a theory of punishment has to do with the need to prevent crime and maintain safety in a community. On this basis it is justifiable to punish criminals if it would better protect the public via deterrence, provided that society is largely just. Even with the provision that society be largely just, my view is prone to the following objections: First, if the aim of punishment is to ensure public safety by means of securing deterrent effect through punishment, why is the court never justified to scapegoat an innocent person, if such scape-goating would maximize the deterrent effect in particular and collective well-being in general? Second, even given the need to promote public safety, would we not be using criminals – as Jeffrie Murphy argues⁴⁸ -- merely as a means to our end, not even when the accused is innocent, *but also when he or she is guilty*? In punishing a criminal for the sake of general deterrence, might we be using the criminal *merely* as a means? Third, if deterrence is the main aim of punishment for serious crimes, might the court be allowed to mete out excessively harsh sentences in the name of “collective utility” or

“public safety,” especially when the crimes are socially abhorred? Fourth, might the court ignore legal excuses, such as mistake, provocation, involuntary intoxication, duress, and insanity⁴⁹ as well as grounds for mitigation of sentence – on grounds of maximizing the deterrent effect?

To reply to these questions I shall show how Scanlonian contractualism allows us to develop a theory of punishment that accords importance to deterrence without the excess of utilitarianism. It does so because according to the contractualist thesis, our sentencing policy must be justifiable to everyone affected by it. It can hence explain why, if the policy is to deter crime, we should place an upper limit on sentences to be meted out for a particular type of crime, why we should not punish the innocent in order to deter, and why the court should not abandon legal excuses in order to maximize the deterrent effect.⁵⁰ I shall also supply non-consequentialist arguments -- based on the principles of fairness and negative desert – to further support the conclusion that a (mixed) deterrence theory need not entail the troubling ideas outlined above.

With certain exceptions (such as the theories of Andrew von Hirsch, T.M. Scanlon, and H.L.A. Hart), theories of punishment fall into three categories: 1) utilitarian or consequentialist theories, 2) retributivism or desert theory,⁵¹ 3) expressivism or communicative theory. Expressivism emphasizes the denunciation of the offender by the court or the public. Communicative theory builds on this basis and in addition places importance on the public apology⁵² expressed by the offender,⁵³ as well as the court’s communicating with the victim, the police, and the public.⁵⁴ Central to communicative theory is the idea that offenders should be treated as rational agents, and as members of a community that has a genuine commitment to certain values. The whole process of a criminal trial – including the giving of

evidence by witnesses in court, cross-examination, the application of criminal law to the case, the verdict, and the punishment – constitute ways of addressing offenders as rational agents.

I am sympathetic to Jonathan Glover's view that expressivism has a tendency to collapse into either a consequentialist or retributivist theory, depending on its rationale for the court or the public to denounce the criminal.⁵⁵ If the need to denounce the criminal has to do with consequences, such a theory can be classified as a consequentialist theory. If, on the other hand, the reason for condemnation has to do with the criminal's desert, the theory would be retributivist in nature – even though a retributivist theory does not necessary have to place any emphasis on communication.

I do not believe that either deterrence theory or retributivism alone can adequately account for the three central questions of criminal punishment: Why punish? How much? Whom to punish?⁵⁶ Non-consequentialism is a much more plausible theory when compared with these two positions. **It will be helpful to first consider the two opposite views about the relevance of consequences for moral consideration.** On the one hand, though one might hold the implausible view that consequences are morally irrelevant, most people would agree that considerations of consequences play a role in practical reasoning: Should I purchase a cellphone that is durable, user-friendly, or equipped with a good camera? It is difficult to think that considerations of this kind are irrelevant to practical reasoning in general or moral deliberation in particular. To give another example: If we can save one stranger's life, or alternatively the lives of five other strangers, then *ceteris paribus* we should save the lives of the latter. We should do so because well-being (and hence consequences) matters, even though it is not the only thing that matters.

On the other hand, one might hold the other extreme view that *only* consequences matter. But this is a strong claim because it either denies the relevance of fairness, justice and desert, or else insists that these considerations must be reducible to consequences. Yet, rejecting consequentialism does not mean that consequences would be irrelevant in practical reasoning. For if one acts on considerations of consequences, one is *not* thereby a consequentialist, not unless one also subscribes to the view that consequences are the *only* relevant considerations for all ethical issues.

Why is it often presupposed that retributivism and deterrence theory are incompatible? My conjecture is that this is the case because retributivism and deterrence theory are grounded on deontology⁵⁷ and utilitarianism respectively, which are indeed incompatible. But this is a false dichotomy. Non-consequentialism holds that both consequences and deontological considerations are relevant in moral deliberation. In developing a credible theory of punishment, non-consequentialism can accord proper weight to the need for deterrence, while also taking account of deontological concerns such as fairness and desert. The contrary view that considerations of deterrence and desert are incompatible or mutually exclusive is mistaken.

As I shall argue, the “mixed” or non-consequentialist theory of punishment consists of the following components: First, general deterrence is an important consideration in regard to punishment. Second, a criminal should be blamed or condemned by the court according to his positive desert.⁵⁸ Third, the ideas of fairness and negative desert⁵⁹ can explain why an innocent person and a defendant with a legal excuse should not be punished at all, and why a convicted offender with grounds for mitigation should be punished less. Fourth, given the contractualist idea that a

sentencing policy has to be justifiable to everyone affected by it, there must be a maximal limit of sentence for any given offense.

4. Treating Someone as Means vs. Treating Someone as *Mere Means*

Though general deterrence does not constitute the whole story to explain why a community should have criminal punishment, it is an integral part of the story. Surely, criminal statutes are enacted with a view to securing social goals. One such goal is to discourage certain acts that are regarded as socially disadvantageous. Pollution and tax evasion, as well as robbery and murder are the kinds of acts that the community has reasons to discourage or prevent. To think that such laws were passed *solely* with a view to punishing those who deserve to be punished seems implausible.⁶⁰ This is especially so regarding crimes that are *mala prohibita*,⁶¹ such as parking offenses.⁶²

General deterrence is not the only socially advantageous goal. Reform, habilitation, and treatment are also important goals for lesser crimes and for young offenders. General deterrence is, however, an especially important goal for serious crimes. One reason that motivates people to reject general deterrence as a basis for justifying criminal punishment is the thought that in punishing criminals for the sake of general deterrence we would be using them *merely* as a means. According to Jeffrie Murphy, if we punish criminals for the sake of deterrence, we are using them merely as a means, not only when they are innocent, but even if they are guilty.⁶³ Murphy's objection, if sound, would apply not only to utilitarianism, but also to any theory (such as non-consequentialism) that takes consequences to be morally important.

Murphy has not spelled out his reasoning as to why punishing criminals on grounds of general deterrence would be using them *merely* as a means. But one can make a conjecture. His reasoning is probably the following:

If we punish criminals not for their own sake, but only for the sake of general deterrence, then we would be using them for the sake of others, and hence using them *merely* as a means to promoting the well-being of other individuals or society as a whole. Therefore, we would be using them impermissibly and violate their rights as persons.

I shall argue that this view, though plausible, is mistaken. Murphy is inspired by Kant. Kant holds that when we use another person merely as a means, we use her as “a thing” and in a way that fails to treat her as a person or in accordance with her dignity. For Kant, to treat someone *merely* as a means is morally impermissible.⁶⁴

Consider the Transplant Case: In a small town, five patients are each dying of the failure of a different organ. One has heart failure, another pancreas failure, the third one liver failure, and so on. They will all die in two days. Their only chance for survival is through organ transplants among themselves, through a voluntary arrangement in which they will draw lots. In fact, family members will draw lots for each of them. The unlucky patient will have to sacrifice himself for the sake of the four lucky patients, whose lives will be saved. To avoid unnecessary complications, let us assume that the chances of the four lucky patients being saved via transplantation are 100%, that the lottery is a fair one, and that if the patients agree to enter into this arrangement they will fall into a temporary coma after receiving an injection of an anesthetic (so that there will be no difficulty in enforcing the agreement). Further, suppose that this arrangement is optional. Only those who agree to enter the lottery will have the prospect of being saved or being sacrificed. Those who do not enter into this arrangement will not be sacrificed, but will also not be saved.

Suppose that having deliberated with their families, all patients have agreed to enter into the lottery, and have agreed to be put into a temporary, anesthesia-induced coma, by signing an official document. Should they have agreed or not? And is the agreed arrangement morally permissible? Each patient has a self-interested reason to live longer than he or she otherwise would have. All of them will die within two days if they do not enter into the arrangement, but they will each have an 80% chance of survival upon agreeing. A simple cost-benefit analysis would show that each patient has a reason for agreeing. I shall call an arrangement (hypothetically) “consentable” if it is beyond a certain threshold of reasonableness – or if it is not too unwise or unreasonable for agents to consent upon. Everyone will be better off in terms of life expectancy and there will be no unfairness. Entering into such an agreement is reasonable, or at least not unreasonable. No one is being forced. No one is being cheated, since the lottery is fair. And there is no difficulty in enforcing the arrangement. I believe that a contractualist would think that it is morally permissible for these patients to enter into such an arrangement, should they choose to do so.⁶⁵ In other words, the arrangement is consentable.

But would each patient – and perhaps the surgeons as well – be using each other and even himself or herself merely as a means, in Kant’s language? Kant distinguishes two types of using people as a means, namely, the permissible and the impermissible use of others.⁶⁶ He calls them “using someone as a means,”⁶⁷ and “using someone *merely* as a means”⁶⁸ respectively. In any cooperative joint venture, we in a sense “use” our partners, but this is generally morally permissible – unless the agreement is illegal or tainted with fraud or unfairness. Such cases must be distinguished from cases where someone is used in a morally impermissible way.

If an arrangement is (hypothetically) consentable, and if people enter into it without deception, and if in addition they actually agree to it,⁶⁹ and if the arrangement is carried out fairly, then their entering into the agreement and subsequent consequences is also morally permissible. Thus, in the Transplant Case, it is morally permissible for the patients to enter into the lottery, provided that each of them knows of the possible consequences.⁷⁰

It is plausible to view the consentable arrangement as one that patients (who wish to live longer) would have reason to agree on. In the Transplant Case, everyone has access to the probabilities of survival should they enter into the lottery, compared with the probabilities of survival should they stay away from it. The higher probability of survival created by the lottery cannot on its own generate an obligation to enter into it. There must also be actual consent in this case, and we need to investigate why this is the case.

Certainly, there are different scenarios that call for different types of consent. People might disagree over the considerations relevant for different decisions. First, there are cases where a hypothetical consent is necessary *and* sufficient for the arrangement. For instance, in an epidemic of a deadly disease, quarantine is justified because 1) the case is other-regarding, 2) quarantine is necessary to prevent dire consequences, and 3) the burden imposed on patients is likely to be bearable.

Second, there is the other extreme where the arrangement is hypothetically *non-consentable* and hence unjustifiable. Consequently, actual consent is unable to justify the arrangement. Dueling (with actual consent) is unjustifiable because the personal generic reason for justifying this sort of activity is weak compared with the personal generic reason against it, viz. the community has a strong reason for not

settling disputes by way of dueling. Murder contracts are also unjustifiable and void, because the community has a strong reason for prohibiting murder.

Third, in between these two extremes lie the cases in which it is acknowledged from the standpoint of a hypothetical consent that actual consent is also necessary to justify the arrangement. The purchase of self-regarding kinds of insurance, such as life-insurance, and most commercial transactions would typically require actual consent. I believe that the Transplant Case belongs to the third type of cases.⁷¹

Why does hypothetical consent have normative force at all? Consider this case: Suppose Mary and her partner, John, are meeting for dinner at a restaurant before an opera. Mary is running late. Knowing Mary's preference well, John orders her favorite dish for her so that they will not be late for the opera. The dish arrives as soon as Mary arrives at the restaurant. Mary is pleased. The question is whether Mary could have objected to John's ordering her favorite dish, even though she acknowledges that this is what she would have ordered? The idea of hypothetical consent would imply that the answer is no.⁷² This does not mean, however, that in all scenarios for hypothetical consent, no actual consent is relevant. In fact, as I said earlier, parties in the hypothetical scenario could agree that actual consent may be relevant or important in certain cases.⁷³

The presence of hypothetical consent helps ensure that people will enter into a *consentable* arrangement. If an arrangement is consentable, it will have to fulfill an *objective* threshold-requirement of reasonableness, while the threshold will depend on the nature of the practice concerned. If, on the other hand, a practice or arrangement is so unwise or unreasonable as to render it impermissible, no actual consent can render

the practice permissible. Conversely, when there is one but only one justifiable arrangement, the arrangement is not only permissible but also mandatory, in which case no actual consent would be necessary.

In a broad range of cases, actual consent is required. This is so because actual consent can be relevant in a number of ways. First, it can be relevant and necessary when hypothetical consent *underdetermines* the arrangement that we ought to enter into. Thus, in the Transplant Case, the parties could have consented in a number of ways as to whom to sacrifice, if they were behind the veil of ignorance: *ceteris paribus* 1) the oldest person (since he has already lived so long, which might make sense from the point of view of distributive justice), 2) the person with the least number of dependents (as this decision will minimize pain and sufferings), 3) the person who is least socially useful (from the perspective of social utility), 4) the person who is least likely to lead a happy life and, if there is a lottery, 5) the person who loses out in the drawing of lots. If these options result in underdetermination, even reason or logic cannot remove this kind of indeterminacy. In other cases, though there may be no indeterminacy in this strong sense, fierce disagreement may exist among (reasonable) people over what should be agreed to.

Still, in some self-regarding cases people should have the autonomy to choose whether to join an arrangement, such as whether to buy life- or home-insurance, because of variation of preferences among different people. A driver of a motor-vehicle must buy third-party insurance, however, because this decision is other-regarding.⁷⁴

A completely different function of actual consent is that it rightly takes account of the actual interest of the parties concerned. Self-interest and morality are often intertwined and mistaken for each other. If there were no actual consent, there would be no guarantee that self-interest and group-interest are not pursued — consciously or subconsciously — in the name of justice or morality. When parties argue for the best or most fair arrangement “behind the veil of ignorance,”⁷⁵ *they are not actually behind the veil of ignorance.*

Finally, parties must be assured that, whoever turns out to be the lucky or unlucky one, the contracted obligation will be performed. Otherwise potential participants might feel that the “agreement” is nothing more than an unenforceable deal. For all of these reasons, actual consent can render a consentable arrangements morally binding.

Now why should a contractualist agree that actual consent is sometimes necessary, or sufficient, to render a hypothetically consentable arrangement morally permissible, or even mandatory? According to Scanlonian contractualism, we must justify our act to every affected person. A hypothetically consentable arrangement is, by definition, one that fulfills the threshold of reasonableness, but one that participants can reasonably join in or opt out. This being the case, actual consent could be necessary, as well as sufficient, to render a (hypothetically) consentable arrangement morally binding in certain cases. If an arrangement is socially advantageous, potential participants can render it justifiable by offering their actual consent.

5. Punishment for the Sake of General Deterrence

Consider punishment for the sake of general deterrence (PGD). If PGD advances an important social goal, and if society is roughly just, then it is (hypothetically) consentable and permissible. In such a case, no one would be used *merely* as a means. However, whether PGD is consentable would depend on whether society is (at least) largely or nearly just. This proviso is important, for otherwise else the arrangement might not even be consentable.

This is so because it is widely held among sociologists and criminologists that poverty and/or extreme inequality⁷⁶ cause crime.⁷⁷ If society is not at least largely just, then, on Scanlonian terms, the practice of punishing people who (through no fault of their own) have a higher chance of committing crime for the sake of deterring others would be disallowed by the conjunction of 1) the contractualist thesis that the practice must be justifiable to everyone affected by it, and 2) the fact that such a practice cannot be justified to those who are badly off in society and who suffer as a result of poverty and/or socio-economic inequality.⁷⁸

It is true that the poor also suffer from robbery, theft and acts of violence by criminals who live in ghettos. But it does not follow that they must therefore consent to PGD, for, in addition to considerations of self-interest, they should also take account of whether an arrangement is fair or not. To see why this is the case, consider a variation of the Transplant Case: The dice is loaded, and consequently the patient with lung failure, say, has a higher chance of being sacrificed (say, 40%) than the others, but a lower chance of dying than that entailed in not joining the arrangement at all (which is 100%). Assuming that this patient knows of these facts, should he consent to the lottery? I think he has a reason for not consenting, namely, because the lottery is unfair. The situation is quite different if the dice has not been tampered with.

Analogously, if society is nearly just and/or equal, then everyone has a reason to support PGD. If there is to be criminal justice, there must be social justice as well.⁷⁹

On the other hand, if society is largely just, then PGD is probably morally permissible, and even mandatory in a society where a social order *without* PGD is not hypothetically consentable. And whether PGD is permissible, or mandatory, or neither, would depend on whether PGD or its absence would actually be consentable in a particular society. (A social order with PGD might not be hypothetically consentable in certain societies, for example if the crime rate is extremely low, or if PGD has no effect against crime, or if PGD is counter-productive.⁸⁰ It is not very plausible to say that PGD must be required in *all* societies regardless of their concrete conditions, even though general deterrence is relevant to a theory of punishment in our society.)

The idea of general deterrence should form an integral part of a theory of punishment for another reason as well: The “pure” retributivist theory (for example, Kant’s view), which claims that retribution or desert are the *only* grounds on which the court should mete out punishment, says that offenders deserve to be punished and that they must be punished. However, if desert or retribution were the only ground for punishment, then punishment must be carried out even if it actually provoked the undesired effect of increasing crimes.⁸¹ Such a consequence would render this view objectionable. In addition, diplomatic immunity in criminal cases is grounded on practical considerations.⁸² Unless we are willing to bite the bullet and deny that diplomatic immunity is ever a ground for not prosecuting a diplomat, and unless we want to say that we should punish criminals even if doing so would cause the “counter-deterrent effect,”⁸³ we should accept the view that the goal of deterrence should be an integral part of a theory of punishment. The most forceful ground for

holding this view is that, as Rawls points out, legislations are passed with a view to securing social goals, one of which is the prevention of crime.⁸⁴ Analogous to Rawls's view is the case of a professor who announces a penalty for late submission of term papers because she wants to deter lateness, rather than to mete out penalty because students who submit late papers deserve to be penalized.

6. A Thought Experiment

Retribution and its communicative variants maintain that a criminal deserves punishment, and that punishment should be carried out by order of the court. Retributivism became popular in the 1970s, a time when public opinion was critical of the need-based criterion of reform to mete out varying sentences to different offenders for the *same* offense. Andrew von Hirsch's *Doing Justice*, arguing that the court should mete out sentences according to the desert of the offenders, was a welcome response to the sentencing policy of an earlier era.

As a theory of punishment, retributivism is attractive for its power to explain why the court should not frame an innocent person, why vicarious punishment (e.g. the practice of vicariously punishing the parents, teachers, and friends of an offender for the most serious crimes in Ancient China) is morally impermissible, and why sentencing should be proportionate to the offense committed.⁸⁵ On the other hand, it has been argued that retributivism – at least an important version of positive retributivism – has not been successfully established.⁸⁶ (I shall discuss Scanlon's criticism of retributivism in section 8 below.)

To discuss retributivism further, let us consider the following "utopian" village: Suppose that a village is isolated from the rest of the world, and that there has been no serious crime for the past fifty years. Someone has recently committed a

serious crime: say, indecent assault. He is now truly and deeply remorseful. The victim has forgiven him. The court convicts him of the offense and condemns him for what he did. However, the court decides that under the circumstances the accused is to be spared of any hard treatment.⁸⁷ Is the court's ruling defensible? I believe that the answer is yes. To understand why, consider a slightly different case: Suppose that for the past twelve months a certain offense – say, car theft -- has been rampant and has caused social problems. Should the court increase the sentence in order to deter such an offense? Clearly, the court should do so, provided that the sentence is not excessive. (This proviso suggests that there should be a cap on the offense, an issue to which I will return below.) If this ruling is defensible, then the reverse is also defensible should the need to deter crimes no longer obtain. The example of the village is such a case in which there is no need to deter crimes.

The problem with *pure* retributivism is the following: If two persons have committed the same offense in two different communities, then provided that various factors (viz., their intentions, the harm caused, and the circumstances in which the offense was committed) are qualitatively identical, the retribution and hence the sentence should be the same. The circumstances to be considered here include, for instance, the harm caused to the victim, the intention of the criminal, and whether there were aggravating factors (e.g. a “breach of trust”) on the part of the criminal, or whether there were mitigating factors (e.g., “provocation” by the victim). The circumstances do not, however, include a social need for deterrence. Thus on this view an indecent assault (say, of a moderate kind) would call for a sentence of the same severity, whether the offense was committed in a metropolis or in the utopian village described here. A pure retributivist would argue that in this isolated “utopian” village the court had acted wrongly, and that it should have meted out a sentence of

imprisonment to the offender, even though he is truly remorseful and the victim has forgiven him

In this example, the offender's having deeply and sincerely repented indicates that his chance of recidivism is minimal. Second, the fact that the victim has forgiven him implies that she will not feel badly or treated unjustly if he is not punished.⁸⁸ Most importantly, until this case, no crime had been committed in this village for the past fifty years, and it is not foreseeable that there will be crime again, because everyone is conscientious and self-disciplined, so that there is no need for deterrence. Furthermore, if the offender were to be punished, it would be unclear what the proportionate sentence should be, since there is no precedent in the village. (Let us assume that the oldest person alive there cannot recall anything about sentencing in this village.) The small scale of the village (where everyone knows everyone else), the uncontaminated innocence, the traditional, simple way of life, as well as education, family, friends, and perhaps counseling have all contributed to the virtual absence of crime.

Pure retributivism is not plausible, because it is not sensitive to context. In determining the right sentence, a judge should take into account the intention or *mens rea* of the offender, the type of offense committed, the harm caused to the victim, the circumstances under which the offense took place, which would include whether there are mitigating or aggravating factors. In addition, I believe that in exceptional circumstances the judge should take account of one further type of fact, which I call the "contextual fact." In the case above, the contextual fact is that the village, isolated from the outside world, has been extremely peaceful for the past fifty years, and hence it is extremely unlikely for any serious offenses to occur in the foreseeable future. *Pure* retributivism does not allow for this kind of contextual fact.

It is therefore plausible that an appropriate sentence for this crime in this “utopian” village should be different from the sentence for the same crime in a cosmopolitan city. The question is what accounts for this divergence of sentencing policy in these two places. The only plausible explanation takes recourse to the need for deterrence. In the hypothetical village there is no need for general deterrence, whereas in a cosmopolitan society there is such a need.⁸⁹

I hope the example of car theft above illustrates why the length of a sentence should vary in tandem with the need to deter the offense – subject to a maximum as I shall discuss. In the limiting case of the utopian village, the appropriate sentence could be one of discharge. A theory of criminal punishment should take account of this sort of “contextual fact.” One of the strengths of contractualism is that it is sensitive to this need.⁹⁰

7. Limits of General Deterrence

Probably many people reject the deterrence theory because they want to reject utilitarianism. Utilitarianism -- certainly act-utilitarianism -- entails the unpalatable conclusion that framing an innocent person as a scapegoat is permissible, or even mandatory, if doing so will maximize the deterrent effect in particular, or promote good consequences in general. However, I wish to point out that endorsing general deterrence as an important goal in criminal punishment, does not make us a utilitarian anymore than a driver who chooses a vehicle for its reliability, handling, and mileage is thereby a consequentialist.

Although consequences are relevant and important considerations in moral matters, according to non-consequentialism, they are not the only considerations.⁹¹ Considerations that cannot be reduced to consequences, such as fairness, justice and

desert, are also important moral considerations on this view. So one can be a non-consequentialist and assign importance to general deterrence, as well as fairness, justice, and even desert.

Even if the best consequences can be realized (when considerations of fairness, justice and desert are set aside), it does not follow that criminal punishment is thereby justified. Surely, harsh laws that inflict heavy penalties on offenders may be effective in deterring or preventing crimes. However, it does not follow that these laws are therefore justified, since it could be very unfair for the court to mete out undeservedly heavy sentences, such as those mandated by the “three strikes law” in various American states and especially in California.

Consider this “pickpocket” example: A city-state that relies heavily on tourism is on the brink of economic disaster because a small group of pickpockets (who take not only the money but also the travel documents in the wallets,) has caused the number of tourists to plummet. Consequently, the economy suffers heavily. Public schools have to close. Hospitals are no longer operative and many patients die as a result. These pickpockets are skillful and hard to catch. Suppose the police have arrested a pickpocket and the jury has convicted him of the offense. The court is now considering what sort of sentence to mete out. The judge thinks that she should mete out the heaviest sentence in order to generate the maximal deterrent effect. On the other hand, she also thinks that such a heavy sentence exceeds what the offender deserves. How should she decide? Quite obviously, meting out the death penalty (if the law allows it) or even life imprisonment is unjustifiably harsh. After all, the offender has only committed the offense of pickpocketing. It is equally obvious that a pure deterrence theory is unable to offer a satisfactory account, because (just as utilitarianism seeks to maximize utility) a pure deterrence theory attempts to

maximize the deterrent effect, at least in those cases where doing so would bring about the greatest collective well-being. It is true that a consequentialist might argue that the costs of punishment (i.e., the pain and suffering to be borne by the prisoners) must also be taken into consideration. However, there is no guarantee that the mathematics will always work out in favor of the pickpocket, no matter what the social circumstances turn out to be.

According to contractualism, the court cannot justifiably mete out excessively harsh punishment to an offender in the name of deterrence, because the court cannot justify this punishment to *him*. To inflict such a punishment on the offender would be using him merely as a means.⁹² *A fortiori*, a judicial act that frames an innocent person is not justifiable to this person. In the sections below, I shall argue that we can reach similar conclusions via the notion of negative desert.

8. Desert

Besides the dimension of general deterrence and consequences, considerations of desert, fairness and justice constitute another important dimension of any theory of criminal punishment. What do I mean by desert? To say that someone deserves X is to say that there is a deontological reason for giving X to her in respect of her acts or characteristics.⁹³ According to Positive Retributivism, a criminal should be punished by virtue of the fact that he has committed a crime and that he deserves to be punished for his crime. Yet whether Positive Retributivism is defensible is controversial. What are the grounds for saying that someone deserves to be punished? Why should someone be punished if this has no positive consequences? The debate is not yet settled in favor of Positive Retributivists.⁹⁴

In particular, Scanlon has argued that while it is reasonable to condemn criminals for their offenses on the grounds of desert, it is not reasonable to argue that they should be punished according to their desert. This is the case because, Scanlon points out, the appropriate response to desert – both for wrongdoings and beneficent acts – is attitudinal. We can rightfully change our attitude to people who perform deplorable acts - because of their desert - but we cannot *punish* them on grounds of desert *alone*.⁹⁵

I believe that Positive Retributivism is not yet established.⁹⁶ Nevertheless, I hold some *attenuated* version of positive retributivism, call it Blame Retributivism, which is more commonly known as “expressivism.” According to Blame Retributivism, a criminal deserves to be condemned for his wrongful offense. In his classic paper, Joel Feinberg has convincingly argued that condemnation is an essential aim in punishment.⁹⁷ Yet, has Feinberg convincingly shown that “hard treatment” can be grounded on his expressivist idea that, in the convention of our society, condemnation is carried out through such hard treatment?⁹⁸ I believe that an account is still needed to explain why we have such a convention. The most plausible, approximately true account is that hard treatment deters.

Scanlon says that we can condemn wrongdoers on grounds of desert.⁹⁹ But to go beyond condemnation and to mete out hard treatment requires that there be deterrent effect, according to Scanlon.¹⁰⁰ He further points out that hard treatment could not be justified as an expression of blame or condemnation on the grounds that the mere *convention* of expressing condemnation could not bear the necessary moral weight.¹⁰¹ This is true, I believe, because there are various ways to denounce a crime and its perpetrators without hard treatment – such as an announcement of someone’s

conviction on public media. To justify hard treatment, general deterrence seems to be necessary.

Therefore, while it makes perfect sense to condemn a cold-blooded murderer,¹⁰² it is not clear why society must also execute him *in order to* condemn him.¹⁰³ It is doubtful that onerous hard treatment, such as capital punishment and long sentences, could be justified *merely* as an expression of blame or condemnation. Thus, Feinberg's example that we often thank people not only verbally, but also by offering a gift, cannot be plausibly extended to the case of meting out hard treatment, such as long sentences. Similarly, while a department chair can thank his secretary with a (relatively) small present for her diligent performance for the past year, his giving her a house or even a new car would seem to be grossly excessive. Analogously, to explain why our penal system uses long sentences as hard treatment, the justification grounded on general deterrence seems necessary.

Part of the disagreement between Scanlon and Feinberg has to do with the question of whether hard treatment can be grounded on condemnation alone. Scanlon holds that condemnation must be limited to verbal communication and change of attitude on the grounds that the appropriate response to desert is attitudinal,¹⁰⁴ whereas Feinberg believes that condemnation can be extended to hard treatment. My view lies in between both positions. Since we give people gifts to express our gratitude to them, why can the court not mete out minor hard treatment, such as community service, at least partly as an expression of condemnation? Scanlon's objection that condemnation cannot bear the moral weight of justifying hard treatment is without force in the case of minor hard treatment, because community service is not so burdensome as to be unjustifiable on grounds of condemnation alone. Moreover, if in some cases minor hard treatment can be justified on grounds of condemnation, it

would seem strange that when we consider serious crimes, hard treatment would have no relevance whatsoever with respect to condemnation. Thus, it seems reasonable to think that just as minor hard treatment can be justified (at least to some extent) as condemnatory, hard treatment can also be justified to some small extent on grounds of condemnation, but not beyond this extent. Therefore, severe punishments for serious crimes are justified mostly by considerations of general deterrence.

9. Negative Desert

Another form of Retributivism that is both weaker and more plausible than Positive Retributivism is Negative Retributivism. According to Negative Retributivism, if one has not committed any crime, one does not deserve to be punished and hence should not be punished. This view is clearly correct. Can one accept Negative Retributivism without accepting Positive Retributivism? I believe that the answer is in the affirmative for two reasons. First, on any plausible understanding of desert, justice and punishment, one cannot deny Negative Retributivism. Even a utilitarian who thinks that we should frame an innocent person in order to avert a disaster from happening has to think that this innocent person does not deserve being punished, if she believes in desert at all, because an innocent person has not committed any offense. Second, and more importantly, Negative Retributivism (which is a much weaker claim than Positive Retributivism) does not presuppose Positive Retributivism. One might ask why, if Positive Retributivism is unintelligible or implausible, Negative Retributivism should be accepted. My answer is that Positive Retributivism is not unintelligible, but simply doubtful or controversial.¹⁰⁵ More important, if we are to accept Positive Retributivism and have it applied in a case, we must accept the following:

- 1) In this particular case, the defendant has committed a crime.
- 2) Because of the crime committed, he deserves to be condemned.
- 3) Because of the crime committed, he deserves to be punished.

The problem with sustaining Positive Retributivism is that 3) is doubtful or controversial, which is why I do not hold Positive Retributivism.

To accept Negative Retributivism, we need to show only that the defendant has not committed the offense, for, if the defendant is innocent, all three conditions fail to be satisfied. He has not committed a crime, consequently 1) does not hold, and hence 2) and 3) are necessarily unsatisfied. Therefore, we can conclude that an innocent person does not deserve to be punished – i.e., that Negative Retributivism is correct – without endorsing Positive Retributivism.

Whether one considers any version of Positive Retributivism, be it Hegel's theory,¹⁰⁶ or the unfair advantage theory, or Duff's secular penance theory,¹⁰⁷ one can cast doubt on some steps in their arguments, and hence doubt the conclusion that a criminal deserves to be punished. However, no one should doubt that an innocent person does not deserve to be punished.

Finally, let me point out that criticisms against Positive Retributivism typically do not apply to Negative Retributivism. For instance, Hart has argued against Positive Retributivism on the grounds that two wrongs do not make a right.¹⁰⁸ Scanlon has rejected it because it is repellant to hold that it is a good thing that some people get punished even if there are no good consequences.¹⁰⁹ I wish to point out that their arguments, if successful, would apply only to Positive Retributivism, and not to Negative Retributivism. This means that we can subscribe to Negative Retributivism without endorsing Positive Retributivism.

The only view that could possibly undermine Negative Retributivism is the no-desert view. It maintains that there is no such thing as pre- or extra-institutional desert, and consequently the claim “the innocent person does not deserve punishment” does not make any sense, because no desert-claim makes any sense. However, this view is very difficult to accept. Rawls holds that only institutional desert exists.¹¹⁰ Contrary to Rawls, it surely makes sense to say, for instance: “This team lost the last World Cup final, but deserved to have won; it was just extremely unlucky.”¹¹¹ In this statement, the desert referred to is pre- or extra-institutional desert.

10. Negative Desert, Excuse, and Mitigation

What follows from Negative Retributivism, which seems to be a very weak claim? I believe that it has interesting implications apart from the conclusion that an innocent person does not deserve to be punished. If someone commits a crime due to factors *beyond her control* (Hart)¹¹² or if she has not had *fair or adequate opportunity to avoid* committing it (Scanlon),¹¹³ she does not deserve conviction or punishment at all. One precondition for committing a serious crime and deserving the corresponding punishment is that one had control, or fair or adequate opportunity to avoid committing the crime. If (through no fault of one’s own) one does not have such control, or does not have fair or adequate opportunity to avoid committing the offense, one has a legal excuse to the charge and consequently does not deserve the conviction, let alone the punishment.

The condition of “having control” or having “fair or adequate opportunity to avoid” is an important safeguard to us as citizens because it protects us from being charged with a serious offense, just as the requirement that there be *mens rea* in order

for a serious offense to be committed is an important safeguard. In fact, they are the same safeguard.¹¹⁴ This is so for the following reason. To have control over my action is a prerequisite for being the agent of my action. If through no fault of my own I do not have control of an act, I am not responsible for it. In a very significant sense, it is not my action. This is so if someone was involuntarily intoxicated while committing the offense, for instance. The important requirement is that the intoxication was caused involuntarily. If a deplorable act was committed as a result, the person who committed it really was not the agent and hence should not be held responsible for what was done through him.¹¹⁵

Negative Retributivism holds that an innocent person, and those who have a legal excuse, do not deserve to be punished. Sometimes, however, whether one has control over one's action or whether one has fair or adequate opportunity to avoid running afoul of the law, is not clear-cut but a matter of degree. When one's control is only partial, or one's opportunity to avoid is diminished, one has grounds for mitigation of sentence.¹¹⁶ Thus if the defendant attacked the victim as a result of provocation, he would have grounds for a mitigation of sentence because (through no fault of his own) his "control" or "opportunity to avoid" was diminished, even though he still had some control over his act or had some opportunity to avoid committing it.

In this latter case, we should invoke Negative *Pro Tanto* Retributivism, the view that *to the extent that* through no fault of one's own one did not have fair or adequate opportunity to avoid committing the offense, then *to that extent* one does not deserve blame or punishment. This view can explain why an offender can have grounds for mitigation when he did not have fair or adequate opportunity to avoid committing the offense.

Note that on Scanlon's view one who commits deplorable acts deserves blame even if she did not have fair or adequate opportunity to avoid them.¹¹⁷ If someone had no control over becoming the sort of person that she now is (as manifested through her acts), Scanlon says, we can nevertheless still blame her because on his view a "desert base" itself need not be deserved.¹¹⁸ I do not, however, fully agree with Scanlon's view here. At least in a case where someone commits a deplorable act because of a psychological disorder, such as schizophrenia, or even autism, it seems wrong to criticize *her* - the person - even though it may be permissible or even right to criticize her *act* (insofar as the act is deplorable).

11. The Pickpocket Example Revisited

Let me return to the pickpocket example. I argued that a pure deterrence theory cannot avoid the consequence that the court should punish the convicted pickpocket very harshly, if doing so can maximize the deterrent effect and promote collective well-being. Yet, punishing a pickpocket with a long sentence is clearly too harsh and hence unjustified.

As I argued, a contractualist should endorse Negative Retributivism. I would like to show that a theory of negative desert can explain why we should neither punish an innocent person, nor punish people vicariously, even if doing so can save many people's lives. We should not do so, because an innocent person does not deserve a guilty conviction, let alone condemnation or punishment. Other theorists might argue that we should not punish an innocent person because such policies, were they to become public, would undermine the rule of law. This reasoning, however, does not do justice to the moral phenomenology of why we think it is wrong to punish an innocent person. Our thinking that punishing an innocent person is wrong is very

direct, and hence quite different from the indirect reasoning derived from the consideration of consequences. Further, consequence-based reasoning leaves too much to chance, because it is possible that punishing an innocent person could bring about the best overall consequences.

Moreover, I believe that the only way to avoid the conclusion that we should punish the pickpocket harshly is to insist that he does not deserve a long sentence. How can we make that argument? No doubt, the pickpocket intended to steal money from the tourists. The connection between such thefts and destabilizing the whole economy is, however, highly contingent and indeed unforeseeable. Thus, the pickpocket's "contribution" to such bad consequences was beyond his control. Even though he had the opportunity to avoid committing the theft, he did not have the opportunity to limit its highly contingent and unforeseeable consequences. Nor did he have the opportunity to know in advance that the unlikely consequences would unfold. Therefore, he should not be held accountable for the unforeseeable bad consequences at the stage of sentencing. Even though meting out a very harsh sentence to the pickpocket may maximize the deterrent effect and hence bring about the best overall consequences, doing so is unjustified because the court would be punishing him for consequences he could not foresee and hence such punishment would be undeserved.

12. Conclusion

Scanlonian contractualism provides a meta-ethical framework that supports non-consequentialism. More importantly, it entails the contractualist thesis that any act is permissible if and only if it is justifiable to everyone affected by it. This thesis allows

us to place limits on sentencing based on deterrence, as well as on the type of person who can be punished.

I argued that we should accept the notion of deterrence as well as those of fairness and negative desert as integral parts of a theory of punishment. Further, a plausible theory should also have room for the expressivist idea of censure.

Contrary to Murphy's view that punishing a guilty criminal for the sake of general deterrence would be using the criminal as a mere means, I have argued that this is not necessarily so because people would have reasons to consent to punishment for the sake of general deterrence, if society is nearly just. Considerations of deterrence must be constrained by contractualist concerns, however, so that innocent people will not be framed and an offender not be punished too harshly, even if doing so would maximize the deterrent effect.

I further argued that a theory of punishment should be sensitive to contextual facts (e.g., whether there is a need for deterrence), and that Scanlonian contractualism is sensitive to these facts. Pure retributivism by its very nature cannot take contextual facts into account.

Although I do not hold Positive Retributivism (the theory that a criminal deserves to be punished and that the court should punish him according to his desert), I accept Blame Retributivism, the expressivist view that the court should condemn the criminal according to his desert. More importantly, I embrace Negative Retributivism, the theory that an innocent person does not deserve any punishment at all. I argued that we should accept Negative Retributivism, even if we do not accept Positive Retributivism, because accepting Negative Retributivism in no way commits us to accepting Positive Retributivism.

Furthermore, I argued for Negative *pro tanto* Retributivism, the view that *to the extent that* (through no fault of one's own) one could not control one's act, or one did not have fair or adequate opportunity to avoid committing the deplorable act, then *to that extent* one does not deserve conviction or punishment. Negative Retributivism accounts for the case of a defendant who has a legal excuse, whereas Negative *Pro Tanto* Retributivism explains why a convicted offender can sometimes have grounds for mitigation. The idea of negative desert underlines the various versions of Negative Retributivism defended in this paper.

Finally, the idea of negative desert also helps to explain why the court is not justified to punish an offender (such as a pickpocket) very heavily in order to secure the maximal deterrent effect, because such a harsh sentence is undeserved. For this reason, a court that metes out excessively harsh punishment, or that punishes vicariously in order to deter would be using the offender as a mere means.

The framework of Scanlonian contractualism allows us to give significance to deterrence without giving up our intuitions that we should not punish offenders too heavily or mete out vicarious punishment. Retributivism is said to be able to account for these intuitions. I hope to have shown that Scanlonian contractualism can account for these intuitions, without the problems faced by pure retributivism.

Bibliography

Bazelon, David L. "The Morality of the Criminal Law." *Southern California Law Review* 49 (1976): 385-405.

Blau, Judith, and Peter Blau. "The Cost of Inequality: Metropolitan Structure and Violent Crime." *American Sociological Review* 47, no. 1 (1982): 121-22.

- Bohm, Robert. "Karl Marx and the Death Penalty." *Critical Criminology* 16, no. 4 (2008): 285–91.
- Brooks, Thomas. "Criminal Harms." In *Law and Legal Theory*, edited by Thom Brooks, 149-62. Leiden: Brill Publishers, 2014.
- Brownlee, Kimberly. "Dialogue." Chap. 7 in *Conscience and Conviction: The Case for Civil Disobedience*. Oxford: Oxford University Press, 2012.
- Burgh, Richard W. "Do the Guilty Deserve Punishment." *Journal of Philosophy* 79, no. 4 (1982): 193-210.
- Clear, Todd. "The Problem of Incarceration Concentrated in Poor Places." Chap. 3 in *Imprisoning Communities: How Mass Incarceration Makes Disadvantaged Neighborhoods Worse*. New York: Oxford University Press, 2007.
- Duff, R.A. *Punishment, Communication, and Community*. Oxford: Oxford University Press, 2001.
- Duff, R.A., and David Garland, eds. *A Reader on Punishment*. Oxford: Oxford University Press, 1994.
- Engels, Friedrich. *The Conditions of the Working Class in England*. Oxford: Oxford University Press, 1993, edited by David McLellan.
- Feinberg, Joel. "The Expressive Function of Punishment." *The Monist* 49, no. 3 (1965): 397-423.
- Gilligan, James. *Preventing Violence*. New York: Thames & Hudson, 2001.
- Glover, Jonathan. "Execution and Assassination." Chap. 18 in *Causing Death and Saving Lives*. Harmondsworth, UK: Penguin Books, 1977.
- Harsanyi, John. "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory." *The American Political Science Review* 69, no. 2 (1975): 594-606.

- Hart, H.L.A. "Legal Responsibility and Excuses." Chap. 2 in *Punishment and Responsibility*. Oxford: Oxford University Press, 1975.
- Hart, H.L.A. "Prolegomenon to the Principles of Punishment." Chap. 1 in *Punishment and Responsibility*. Oxford: Oxford University Press, 1975.
- Heffernan, William C. "Social Justice/Criminal Justice." In *From Social Justice to Criminal Justice*, edited by William C. Heffernan and John Kleinig, 66-68. New York: Oxford University Press, 2000.
- Hirsch, Andrew von. *Censure and Sanctions*. Oxford: Oxford University Press, 1996.
- Hooker, Brad. "Contractualism, Spare Wheel, Aggregation." In *Scanlon and Contractualism*, edited by Matt Matravers. London: Frank Cass Publishers, 2003.
- Kamm, Frances. *Intricate Ethics*. New York: Oxford University Press, 2007.
- Kamm, Frances. "Non-consequentialism." In *The Blackwell Guide to Ethical Theory*, edited by Hugh LaFollette and Ingman Persson, 205-26. Oxford: Blackwell Publishers, 2000.
- Kant, Immanuel. *Practical Philosophy*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.
- Kleiman, Mark. "Toward Fewer Prisoners & Less Crime." *Daedalus* 139, no. 3 (2010): 115-23.
- Korsgaard, Christine and Onora O'Neill. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.
- Loury, Glenn. *Race, Incarceration, and American Values*. Cambridge, MA: MIT Press, 2008.
- Marx, Karl. *Capital: A Critique of Political Economy, Book 1*. New York: Dover Publications Inc., 2011.

- Marx, Karl. "Capital Punishment. — Mr. Cobden's Pamphlet. — Regulations of the Bank of England," *New-York Daily Tribune*, February 17, 1853. Available at <https://www.marxists.org/archive/marx/works/1853/02/18.htm>
- Morris, Herbert. "Persons and Punishment." *Monist* 52, no. 4 (1968): 475-501.
- Murphy, Jeffrie G. "Marxism and Retribution." *Philosophy and Public Affairs* 2, no. 3 (1973): 217-43.
- Murphy, Jeffrie G. and Jean Hampton. *Forgiveness and Mercy*. Cambridge: Cambridge University Press, 1990.
- Nagel, Thomas. "Rawls on Justice." In *Reading Rawls*, edited by Norman Daniels, 1-15. New York: Basic Books, 1975.
- Nathanson, Stephen. *An Eye for An Eye?* Oxford: Rowan & Littlefield, 2001.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
- Pallikkathayil, Japa. "Deriving Morality from Politics: Rethinking the Formula of Humanity." *Ethics* 121, no. 1 (2010): 116-47.
- Parfit, Derek. "Justifiability to Each Person." *Ratio* 16, no. 4 (2003): 368-90.
- Pettit, Philip. "A Consequentialist Perspective on Contractualism." *Theoria* 66, no. 3 (2000): 228-36.
- Pettit, Philip. 1999. "Doing unto Others." *Times Literary Supplement*, June 25, 7-8.
- Pettit, Philip. "Two Construals of Scanlon's Contractualism." *The Journal of Philosophy* 96, no. 3 (2000): 148-64.
- Primoratz, Igor. *Justifying Legal Punishment*. Amherst, MA: Humanity Books, 1989.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press, 1971.
- Rawls, John. "Two Concepts of Rules." In *John Rawls, Collected Papers*, edited by Samuel Freedman, 20-46. Cambridge, MA: Harvard University Press, 1999.

- Royal Commission on Capital Punishment. *Report of the Royal Commission on Capital Punishment, 1949 -1953*. London: H.M.S.O., 1953.
- Scanlon, T.M. "The Aims and Authority of Moral Theory." *Oxford Journal of Legal Studies* 12, no. 1 (1992): 1-23.
- Scanlon, T.M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.
- Scanlon, T.M. "Contractualism and Justification." Paper presented at the Workshop on Contractualism and Justification, Porto Alegre, Brazil, March 2014.
- Scanlon, T.M. "Contractualism and Utilitarianism." Chap. 7 in *The Difficulty of Tolerance*. Cambridge: Cambridge University Press, 2003.
- Scanlon, T.M. "Desert, Blame and Punishment." Paper presented at the International Conference on the Philosophy of Criminal Punishment, Hong Kong University, Hong Kong, June 2013.
- Scanlon, T.M. "Giving Desert its Due." *Philosophical Explorations* 16, no. 2 (2013): 1-16.
- Scanlon, T.M. *Moral Dimensions*. Cambridge, MA: Harvard University Press, 2008.
- Scanlon, T.M. "Punishment and the Rule of Law." In *Deliberative Democracy and Human Rights*, edited by Harold Hongju Koh and Ronald Slye, 257-71. New Haven: Yale University Press, 1999.
- Scanlon, T.M. "Rawls on Justification." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 139-167. Cambridge: Cambridge University Press, 2003.
- Scanlon, T.M. "Replies." *Social Theory and Practice* 28, no. 2 (2002): 337-58.
- Scanlon, T.M. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

- Scanlon, T.M. "Wrongness and Reasons: A Re-examination." In *Oxford Studies in Metaethics*, volume 2, edited by Russ Shafer-Landau, 5-20. Oxford: Oxford University Press, 2007.
- Shelby, Tommie. "Justice, Deviance, and the Dark Ghetto." *Philosophy & Public Affairs* 35, no. 2 (2007): 126-60.
- Sher, George. "Deserved Punishment." Chap. 5 in *Desert*. Princeton: Princeton University Press, 1987.
- Southwood, Nicholas. *Contractualism and the Foundation of Morality*. Oxford: Oxford University Press, 2010.
- Tasioulas, John. "Mercy." *Proceedings of the Aristotelian Society* 103, no. 1 (2003): 101-32.
- Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1992.
- Wacquant, Loïc. "Class, Race and Hyperincarceration in Revanchist America." *Daedalus* 139, no. 3 (2010): 74-90.
- Westen, Peter K. and James Mangiafico, "The Criminal Defense of Duress: A Justification, Not an Excuse - And Why It Matters," *Buffalo Criminal Law Review*, Vol. 6 (2003), 833.
- Western, Bruce and Becky Pettit. "Incarceration and Social Inequality." *Daedalus* 139, no. 3 (2010): 8-19.

Notes

[I carried out the research for this article when I was a Fulbright Senior Visiting Researcher in the Department of Philosophy at Harvard University during the academic year of 2010-11. I am extremely grateful to Tim Scanlon for hosting me and for numerous helpful discussions on contractualism, punishment, and aggregation. I am grateful to the editor of *Criminal Justice Ethics* and anonymous readers for excellent sets of comment. I am also very grateful to Kimberley Brownlee, John G. Bennett, Peter Chau, and Joe Lau for particularly helpful comments on earlier drafts of this paper. I would like to thank David Archard, Joseph Chan, Jiwei Ci, Sin-Yee Chan, Chris Fraser, Roger Lee, Win-chiat Lee, Tommie Shelby, Kai-yee Wong, and Allen Wood for discussion or comments.]

¹ Scanlon elaborates his contractualism in *What We Owe to Each Other* and in subsequent publications, especially “Replies,” “Rawls on Justification,” *Moral Dimensions*, “Wrongness and Reasons,” *Being Realistic about Reasons*, and “Contractualism and Justification, as well as papers published prior to *What We Owe to Each Other*, such as “Contractualism and Utilitarianism,” and “Aims and Authority.”

² The characterization is Scanlon’s. As he explains: “What I am calling Philosophical Enquiry is addressed to questions about the nature of morality (or of some part or aspect of it) rather than about which actions or practices are right or wrong. It can therefore be classed as a kind of meta-ethics if this term is taken in a broad sense which does not imply that the method to be employed is necessarily the analysis of the meaning of moral terms.” “Aims and Authority,” 5.

³ See Scanlon, *What We Owe to Each Other*, 154, 157.

⁴ Scanlonian contractualism offers a substantive account as to what kind of reasons are to be given weight, and why consequentialism should be rejected. See section 1 of this paper.

⁵ See Rawls, *A Theory of Justice*.

⁶ Scanlon makes this point in *What We Owe to Each Other*, 242.

⁷ Rawls's theory also faces other formidable challenges. See Nagel, "Rawls on Justice," and Harsanyi, "Can the Maximin Principle Serve." See also Scanlon's criticism in section 2 below.

⁸ See Scanlon, "Punishment and the Rule of Law;" "Giving Desert Its Due;" "Desert, Blame and Punishment;" and "Contractualism and Justification." The view I shall argue for is not a summary of Scanlon's first-order view on punishment, but departs from his view at various points. In particular, I differ (to some extent) from Scanlon's view with respect to the issue of desert, and I propose to rely on the idea of negative desert 1) to place limits on the maximal penalty for an offense, 2) to offer a rationale for legal excuses and for 3) grounds for mitigation.

⁹ I follow Frances Kamm's usage of the term here. See Kamm, "Non-consequentialism," and chapter 1 of *Intricate Ethics*.

¹⁰ See Scanlon, "Contractualism and Justification."

¹¹ Although I do not have room for elaboration here, I believe that based on the contractualist thesis that we should justify our act to everyone affected by it, capital punishment should be excluded from the legal system because of inevitable and unjustifiable mistakes.

¹² See Scanlon, *What We Owe to Each Other*, ch. 5, and "Replies." See also Parfit, "Justifiability to Each Person." Scanlon tries to resolve the problem of aggregation by means of what is now widely known as the Individualist Restriction. To illustrate Scanlon's approach, suppose Annie, Bernie and Cathy prefer a course of action A, whereas Derek prefers B. Scanlon compares the strongest personal generic reason for taking A (assuming that the personal reasons Annie, Bernie and Cathy have are equally strong) with Derek's personal generic reason for B. On Scanlon's view, we are to compare Annie's reason with Derek's reason and completely discount Bernie's and Cathy's. (In case Annie's and Derek's reasons are equally weighty, Bernie's reason for A – if weighty enough, or "relevant" – can "break the tie.") I do not accept the Individualist Restriction myself. Suppose we can save one person from death, or alternatively two hundred people from "near death" (a condition almost as bad as death). Who should we save? According to the Individualist Restriction, we must save the one person. It seems clear, however, that the right thing to do is to save two hundred people from "near death." I agree with Scanlon's conclusion that headaches each suffered by a different person – no matter how numerous -- cannot possibly outweigh a life. I argue for this conclusion without recourse to Scanlon's position in my forthcoming book (tentatively titled *Practical Ethics: A Contractualist Approach*). In

any case, the issue concerning the Individualist Restriction is less relevant to the problem of punishment, because we can fall back on the fundamental contractualist thesis that an act is permissible if and only if it is justifiable to everyone affected by it.

¹³ Scanlon, “Contractualism and Utilitarianism,” 124. This is especially so with act-utilitarianism, while rule-utilitarianism is widely regarded as an unstable compromise.

¹⁴ See Scanlon, “Wrongness and Reasons,” 8, 11, 13 and 17-18.

¹⁵ See Scanlon, *What We Owe to Each Other*, 219.

¹⁶ *Ibid.*, 174-175. Scanlon also includes sexual morality under morality in this broader sense.

¹⁷ Scanlon writes: “It is one thing to take into account the (as it were impersonal) fact that it is a bad thing for a person to be harmed, and quite a different thing to take into account the fact, when it is a fact, that that person could reasonably complain about being treated in this way. Taking the latter into account—taking into account the claims that they have on us—is crucial to our relations with each other. This difference between personal and impersonal reasons seems to me to be clearly revealed by the remorse test. Violating a principle supported *only* by impersonal reasons does not damage my relations with the person who has these reasons in the same way as when the reasons against a principle permitting an action are personal reasons of the kind I have described (such as reasons to want to preserve the opportunity to experience these wonders in their unspoiled state)” (Scanlon, “Contractualism and Justification,” 11-12).

¹⁸ See Scanlon, *What We Owe to Each Other*, 152.

¹⁹ *Ibid.*, 153-154. In a slightly differently worded formulation the reason contractualism emphasizes is “the reason we have to live with others on terms that they could not reasonably reject insofar as they are motivated by this ideal.” These are both formulations of contractualism because both have to do with the kind of wrongness or impermissibility that an affected person can reasonably reject, and which would damage the relation between the wrongdoer and others. We are motivated – even motivated to go to great lengths – to justify our actions to others on terms that are reasonable insofar as they also share this ideal. But whether or not others *actually* share this ideal is irrelevant. Even if other members of my community do not share this ideal, Scanlon holds, one still wants to be able to justify one’s action as being reasonable. Thus, the kind of “agreement” in his formulation of contractualism is merely hypothetical.

²⁰ See Scanlon, “Wrongness and Reasons,” 12; *What We Owe to Each Other*, 12.

²¹ See Scanlon, *What We Owe to Each Other*, 201.

²² Even if we have gone through many principles and none of them would disallow *A*, it does not follow that *A* is permissible. It may just be that we are not imaginative enough. Thus, the more principles we consider that do not disallow *A*, the more confident we are in saying that *A* is permissible. It does not follow that we *know* with certainty that *A* is permissible.

²³ Scanlon, *What We Owe to Each Other*, 199.

²⁴ *Ibid.*, 204.

²⁵ Scanlon elaborates as follows: “The relevant reasons are ‘generic reasons’—reasons that a person has simply in virtue of being in a certain position—for example, the reasons that someone has as a person to whom a certain assurance has been given, or the reasons of someone who has given assurance may have to want to act differently, or, to take another example, the reasons that someone has not to be harmed in a certain way versus the reasons that individuals have not to want to take certain precautions against causing harms of that kind. These generic reasons that are relevant to determining matters of right and wrong because we need to be able to settle questions of right and wrong in the abstract, before we know which individuals, if any we will actually interact with in these ways.” “Contractualism and Justification,” 6. Note that this focus on abstract requirement does not mean that the circumstances under which an act takes place are morally irrelevant to the determination whether a principle is reasonably rejectable or not.

²⁶ Scanlon, *What We Owe to Each Other*, 219. (Emphasis on “individuals” added.)

²⁷ In Scanlon’s words: “[I]mpersonal grounds ... [are] reasons that are not tied to the well-being, claims, or status of individuals in any particular position.” *Ibid.*, 219. First, one kind of impersonal reasons pertain to our actions on eco-systems, natural wonders (such as the Grand Canyon), non-human animals, trees as well as works of art. This does not, however, mean that actions damaging nature or works of art *necessarily* (or even likely) fall outside the scope of contractualism. Such actions may harm the interests of other individuals, in which case their claims would have to be taken into account even though the force of their “personal reasons” “depends ... on further judgments of impersonal values, namely the judgment that these objects are worth seeing and should be admired.” *Ibid.* 220.

There remain, however, actions harmful to nature or animals that do not harm other individuals' interests. Such actions would fall completely outside the scope of "what we owe to each other."

²⁸ Scanlon is clear that "[c]ontractualism is not based on the idea that there is a 'fundamental level' of justification at which only well-being ... matters and the comparison of the magnitudes of well-being is the sole basis for assessing the reasonableness of rejecting principles of right and entitlement." Ibid., 214.

²⁹ Scanlon, "Contractualism and Justification," 6.

³⁰ Ibid., 7.

³¹ See Ibid.

³² See Scanlon, *What We Owe to Each Other*, 206-217, 229.

³³ See Thomson, *The Realm of Rights*, 30, n.19; Pettit, "Doing unto Others," 7-8; Pettit, "Two Construals of Scanlon's Contractualism," 148-164; Pettit, "A Consequentialist Perspective on Contractualism," 228-236. Scanlon reports that Frances Kamm was the first to press this objection on him.

³⁴ Scanlonian contractualism is no more circular or redundant than the Ideal Observer Theory or God's Command Theory.

³⁵ See Hooker, "Contractualism, Spare Wheel, Aggregation," 58, 64, and Southwood, *Contractualism and the Foundation of Morality*, 61-70.

³⁶ See Scanlon, *What We Owe to Each Other*, 242-246.

³⁷ I would like to thank Tim Scanlon for discussing this point with me.

³⁸ Scanlon also thinks that there does not exist any algorithm in ethics for several reasons. His contractualist formula is anything but algorithmic. First, when we object to someone's action performed under certain circumstances, we offer reasons because we cannot object to someone's action without having some idea of why his action is objectionable. See Scanlon, *What We Owe to Each Other*, 197-8. Such reasons are the grounds on which to base a general conclusion against his action under the circumstances. But the process of drawing such a general conclusion – which Scanlon calls "a principle" – is anything but mechanical. It requires the endless seeking of reasons, the weighing of them, and sometimes the silencing of some kind of reasons by a reason of a different order, and the eventual drawing of the correct conclusion from conflicting kinds of reasons. Even in the case of an

uncontroversial principle, there is much room for interpretation and judgment as to whether it is applicable to the case under consideration. The apparently simple or uncontroversial moral principles often disguise complex issues. Consider the principle or rule: “Thou shalt not kill.” This prohibition seems relatively simple and straightforward. But does this principle or rule prohibit abortion, voluntary euthanasia, capital punishment, or killing in self-defense? And why is it permissible to kill nonhuman animals, if it is permissible at all? In order to understand this principle or rule more fully, we must try to understand the complex networks of issues involving the right to life and the circumstances under which such a right might be forfeited (if at all), as well as the issue of moral status when the problems of abortion and animal ethics are considered. In interpreting what “Thou shalt not kill” means and what kinds of cases it covers, we have to draw on this complex understanding, rather than applying a storable rule mechanically. It is this understanding, rather than the storable rule, that enables us to arrive at conclusions about new kinds of cases See *What We Owe to Each Other*, 201.

³⁹ It is quite widely held that Kant’s “practical judgment” is similar to Aristotle’s *phronesis*.

⁴⁰ Scanlon, “Contractualism and Utilitarianism,” 142-44.

⁴¹ Harsanyi arrives at the Principle of Average Utility. See *ibid.*, 142.

⁴² Rawls derives the Difference Principle from this basis. See Rawls, *A Theory of Justice*, 150-61.

Scanlon, however, seems to doubt that Rawls has offered sufficient justification for suppressing probabilities in the Original Position. See Scanlon, “Contractualism and Utilitarianism,” 148.

⁴³ See Scanlon, “Contractualism and Utilitarianism,” 143-47.

⁴⁴ See *ibid.*, 145-48.

⁴⁵ See *ibid.*

⁴⁶ See *ibid.*, 138. Suppose the contracting parties have access to moral reasons as well as other facts except information about their own identity. The hypothetical agreement reached should be fair. Although suppression of the information about identity is inessential (as Scanlon rightly holds), this might well be heuristically or instrumentally useful for the parties to reach the hypothetical agreement, at least in some cases (for example, when their self-interest might overshadow their moral reasons). See the case of the loaded dice in section 4 of this paper.

⁴⁷ See Scanlon, “Contractualism and Justification.” Nozick’s entitlement theory is discussed in his *Anarchy, State, and Utopia*.

⁴⁸ See Murphy, “Marxism and Retribution,” 217-43.

⁴⁹ H.L.A. Hart points out that utilitarianism cannot satisfactorily explain why the legal system should allow for legal excuses and has proposed a plausible answer to this last question. See Hart, “Legal Responsibility and Excuses.” My proposal is compatible with but different from Hart’s.

⁵⁰ Based on the contractualist thesis that an act must be justifiable to everyone affected by it, it is highly arguable that capital punishment cannot be justified in our society because convictions are inevitably marred by mistakes, and hence capital punishment would be reasonably rejected by the innocent person who is erroneously convicted of murder.

⁵¹ See Scanlon, “Giving Desert its Due.” Scanlon points out that a desert theory need not be retributivist, because the desert theory can be limited to condemning (but not punishing) the criminal.

⁵² See Duff, *Punishment, Communication, and Community*, 94-114.

⁵³ Thomas Brooks also regards the difference between expressivism and communicative theory in a similar way. See Brooks, “Criminal Harms.”

⁵⁴ See Tasioulas, “Mercy.” John Tasioulas argues for a broader view of communication .

⁵⁵ See Glover, *Causing Death and Saving Lives*, ch. 18. Brooks also holds that expressivism collapses into retributivism. See “Criminal Harms.” Along this line, we can see why some expressivist or communicative theories can be “mixed,” viz. that combine elements of consequentialism and retributivism.

⁵⁶ See Hart, “Prolegomenon to the Principles of Punishment.” Hart suggests that a good theory of punishment should be able to answer all three questions satisfactorily.

⁵⁷ By “deontology” here, I mean an ethical theory that takes account of deontological considerations. A pure deontological theory would disregard consequence-based considerations, whereas a moderate (or “mixed”) theory would take account of such considerations. Retributivism is often associated with the pure version of deontological theory.

⁵⁸ By “positive desert” I mean that someone *deserves* something (whatever this turns out to be) by virtue of his acts or characteristics.

⁵⁹ By “negative desert” I refer to the idea that someone *does not deserve* anything (to the usual extent or at all).

⁶⁰ See Rawls, “Two Concepts of Rules.” Surely, insofar as the law has the power to create *mala prohibita*, not all laws can exist to punish people who do what is antecedently wrong. Therefore, not all laws are created with the view to punishing people who deserve to be punished.

⁶¹ *Mala prohibita* are offenses that are “wrong because prohibited” by law. *Mala prohibita* contrast with *mala in se*, which are offenses that are “wrong in themselves.” Murder and rape are *mala in se*.

Retributivism is most plausible for *mala in se*.

⁶² It is currently an offense in Hong Kong for a retailer not to charge customers a levy for plastic bags obtained or used. This is a *malum prohibitum*.

⁶³ See Murphy, “Marxism and Retribution.”

⁶⁴ As Scanlon has plausibly argued, this is a meta-ethical doctrine that explains what acting wrongly consists in. See Scanlon, “Aims and Authority.”

⁶⁵ Would Kant think that entering into this agreement is to some extent like committing suicide and hence wrongful (since Kant thinks that suicide is impermissible)? Kant raises the casuistic question of whether someone who decides to be vaccinated against smallpox, thereby putting her life in danger, is committing suicide, even though she does it in order to preserve her life. See *Practical Philosophy*, 548 (6:424). Though Kant does not answer the question, we can safely assume that the answer is no. The patients in the Transplant Case are in a rather similar situation, with each person running a 20% risk of death in order to avoid death. Furthermore, Kant holds that someone who runs from the enemy to save his life and “leaves his comrades in the lurch” (ibid.) is a coward. If, on the other hand, someone dies defending herself and her fellows, the action is no suicide but “noble and gallant” (ibid.) For Kant, life itself is not to be highly prized, and one should seek to preserve one’s life only insofar as one is worthy to live, because it is better to sacrifice life than to forfeit morality since “it is necessary that, so long as we live, we do so honourably; but he who can no longer live honourably is no longer worthy to live at all” (ibid., 147 [27:373]). If the parties to a “mutual help” arrangement attempt to maximize their chances of survival in order to take care of their families, or to contribute to society, or to complete whatever worthwhile projects they have committed themselves to, they are not committing suicide but are doing something rational and reasonable.

⁶⁶ For Kant, we should always treat someone as an end, that is, as “a man” (ibid., 147 [27:373]) or “[a] person . . . whose actions can be imputed to him” (ibid., 378 [6:223]) because “[h]umanity . . . is

worthy of respect” (ibid., 147 [27:373]). What Kant means is that a person has dignity (which is a non-derivative value) which is inherent in a person’s property of being a rational agent, and that because he possesses such inherent value, a person is worthy of respect.

⁶⁷ Insofar as a person is useful, she can be a means to someone else’s end. Being a means is compatible with being an end. We can say that a person has derivative value, but since she is a person endowed with rational capacity, she also has non-derivative value. See Scanlon, “Moral Dimensions,” 91-3.

⁶⁸ Being *merely* a means (or being a *mere* means) is incompatible with being an end. On Kant’s view, a nonhuman animal has only derivative value because it does not have the rational capacity of a person and hence does not have the inherent or non-derivative value (the “dignity”) that is worthy of our respect. He thinks that it is right to treat an animal *merely* as a means, which is to treat it as “a thing” but wrong to treat a person *merely* as a means. “A thing is that to which nothing can be imputed. Any object of free choice which itself lacks freedom is therefore called a thing (res corporalis).” *Practical Philosophy*, 378 (6:223). It is widely held that Kant is mistaken in thinking that only human beings can have non-derivative value.

⁶⁹ One might ask whether an actual agreement alone would not be sufficient. Christine Korsgaard and Onora O’Neill have argued that to treat someone merely as a means is to treat him in a way to which he could not possibly consent. When someone is deceived or coerced to do X, they argue, he could not possibly consent to doing X. See Korsgaard, and O’Neill, *Sources of Normativity*. However, as T. M. Scanlon points out, there is such a thing as permissible deception, or even permissible coercion. Japa Pallikkathayil agrees and argues further that Korsgaard’s view is implausible as an interpretation of Kant’s Formula of Humanity because, for one thing, Kant says that force is allowed in some situations. Moreover, Pallikkathayil points out, Kant says that one cannot (actually) consent to being punished. See Pallikkathayil, “Deriving Morality,” 116-25. Since Kant thinks that criminals should be punished, it follows that punishment is not based on actual consent. In other words, the idea that one cannot consent to being punished does not entail that one cannot consent to following the law with the understanding or consequence that if one breaks it, one will be punished. The institution of punishment is, for Kant, grounded on hypothetical consent, or reason. See also Murphy, “Marxism and Retribution,” 54-7.

⁷⁰ This conclusion is compatible with Kant's view that sacrificing oneself in a battle for one's country is a noble act. Someone who is willing to die for her country is not thereby committing suicide, which Kant would have condemned.

⁷¹ Some people believe that it is wrong to enter into this kind of lottery. We must respect their belief, as long as they do not interfere with others who want to enter into the lottery.

⁷² I am grateful to Thomas Nagel for a discussion on this point.

⁷³ See Scanlon, *What We Owe to Each Other*, 155, 176. I am grateful to David Archard and Tim Scanlon for independently suggesting this point.

⁷⁴ The personal preference to purchase insurance of various types seems to fall on or around the blurred boundary of the realm of consentable practice. One relevant consideration is whether the practice is self-regarding or other-regarding. If a certain type of insurance (e.g., third-party insurance) is important as well as unwise not to buy it, and if it is other-regarding, then it would be only right to force people to purchase this type of insurance. However, it may be completely optional whether one wishes to purchase the type of insurance that is self-regarding (e.g., fire insurance, home or renters insurance, even life insurance), in which case actual consent is required. Another relevant consideration is the magnitude of the consequence. Everything else being equal, the greater the consequence, the less likely it is that we can leave the decision to an individual's actual preferences; conversely, the lesser the consequence, the more likely it is that we can leave it to individual choices.

⁷⁵ See section 2 and note 46.

⁷⁶ I realize that there is considerable disagreement among sociologists about whether poverty or extreme inequality is the real cause of crime. But I leave this issue aside. For references on this point, see Judith R. Blau and Peter M. Blau who argue: "income inequality in a metropolis substantially raises its rate of criminal violence. Once economic inequality is controlled, the positive relationship between poverty and criminal violence disappears. . . . Apparently the relative deprivation produced by much inequality rather than absolute deprivation produced by much poverty provides the most fertile soil for criminal violence" ("Cost of Inequality," 121-22). David Bazelon, Richard Delgado, and Elliott Currie all endorsed the idea that "relative deprivation" causes crime. See also Heffernan, *From Social Justice to Criminal Justice*, 66-68. Other relevant sources include: Kleiman, "Toward Fewer Prisoners"; Wacquant, "Class, Race & Hyperincarceration"; Western and Pettit, "Incarceration and Social

Inequality,” 12-17; Clear, *Imprisoning Communities*; and Loury, *Race, Incarceration, and American Values*.

⁷⁷ The evidence for the view that crime is positively correlated with poverty is overwhelming. As Murphy points out, “[o]f 1.3 million criminal offenders handled in USA everyday, 80% belong to the lowest 15% income level – the ‘poverty level’” (“Marxism and Retribution,” 60).

Tommie , notes that “despite making up only 13 percent of the male population in the United States, black men constitute almost half of the male prison population, and on any given day, nearly a third of all black men in their twenties are in prison, on probation, or on parole. These black men are overwhelmingly from ghetto communities” (“Justice, Deviance, and the Dark Ghetto,” 142). The theory is that, because of economic hardship in an inequitable society, those below the poverty line feel the pressure to commit crime to supplement their income, since they are labeled by living “in the ghetto”), lack proper education, and consequently have extremely limited opportunity for lawful earnings. The theory recognizes that the rich also commit crimes, due to factors such as greed and jealousy, but not because of economic hardship. (Indirectly, the poor are driven to commit crimes—and not only property crimes—because of the lack of opportunity for proper education.) The theory also takes account of the fact that some poor people can successfully refrain from committing crime, while a smaller percentage can even rise above their social class. Notwithstanding these adjustments, the theory claims that economic hardship in an inequitable society is a major cause of the proliferation of crime in capitalist systems.

See also Marx, *Capital* and “Population Crime and Pauperism”; Engels, *Conditions of the Working Class* and Gilligan, *Preventing Violence*, esp. 39-46 and 91. Further, Marx and Engels have argued for a positive correlation between property crimes and the rise of capitalism. The positive correlation between the rise of unemployment and crimes is a truism nowadays.

⁷⁸ We can also reach a similar argument if we adopt a quasi-Rawlsian “veil of ignorance” design that applies on a micro scale while other information remains available: if society is not nearly just, the parties behind the veil of ignorance as to which class or stratum of society they might belong have reasons to reject PGD. For they might turn out to be at the bottom of the social structure, and suffer a higher chance of committing crime as well as of being the primary target, and hence bearing the brunt, of PGD.

⁷⁹ See Bazelon, “Morality of the Criminal Law,” and Heffernan, *From Social Justice to Criminal Justice*.

⁸⁰ This result caused by deterrence is known as the “brutalization effect.” See Nathanson, *An Eye for An Eye?* 13-14, 28-29, 43; and Marx, “Capital Punishment.”

⁸¹ Robert Bohm calls this the “counter-deterrent effect.” See Bohm, “Karl Marx and the Death Penalty.”

⁸² Many sovereign states have a long tradition of granting diplomats immunity from prosecution for criminal offenses. The international treaty that recognizes this tradition is the Vienna Convention on Diplomatic Relations (1961).

⁸³ See Bohm, “Karl Marx and the Death Penalty.”

⁸⁴ The case of *mala prohibita* provides another reason why general deterrence is an important goal of criminal punishment.

⁸⁵ See Hart, “Prolegomenon to the Principles of Punishment.”

⁸⁶ The retributivist arguments offered by Herbert Morris and Jeffrie Murphy have been refuted and their authors now disown them. See Morris, “Persons and Punishment” and Murphy, “Marxism and Retribution.” See also Burgh, “Do the Guilty Deserve Punishment”; Sher, “Deserved Punishment”; and Dagger, “Playing Fair with Punishment.”

⁸⁷ In British law, “discharge” is a legal option that has been used in cases where the court deems it inappropriate to mete out hard treatment. Discharge means that no hard treatment is meted out, though the offense is recorded. In a Hong Kong case, a desperately poor beggar stole from the public moneybox. The magistrate convicted him of theft, but the sentence meted out was “discharge.” In addition, the magistrate gave a HKD 100 bill to the beggar.

⁸⁸ Murphy and Hampton argue that forgiveness (given by the victim) and mercy (shown by the court) are two different things, and hence that whether one should be punished is a separate issue from whether one has been forgiven by the victim. See *Forgiveness and Mercy*, 20-22.

⁸⁹ See also Hirsch, *Censure and Sanctions*. Hirsch uses a somewhat similar example in his book.

⁹⁰ According to contractualism, whether an act is wrong depends on whether *its performance in the circumstances* is disallowed by a principle to which no one could reasonably reject.

⁹¹ See Kamm, *Intricate Ethics*, ch. 1.

⁹² See Scanlon, “Contractualism and Justification.”

⁹³ See Nathanson, *An Eye for An Eye*, ch. 6-7. Nathanson argues that a claim of what someone deserves should include her effort spent in achieving or avoiding an act.

⁹⁴ See note 88.

⁹⁵ See Scanlon, “Giving Desert its Due.” The point is that deterrence is necessary to justify punishment.

⁹⁶ I have no *deep* commitment to its denial either. I believe, however, that in its full-fledged form (viz. “an offender deserves to be punished and ought to be punished because of his/her wrongdoing”)

Positive Retributism is hard to defend. See T. M. Scanlon’s argument on desert in this section.

⁹⁷ See Feinberg, “The Expressive Function of Punishment.”

⁹⁸ See *ibid.*, 402. Moreover, Lord Denning says: “The ultimate justification of any punishment is not that it is a deterrent, but that it is the emphatic denunciation by the community of a crime: and from this point of view, there are some murders which, in the present state of public opinion, demand the most emphatic denunciation of all, namely the death penalty” (quoted in the *Report of the Royal Commission on Capital Punishment*).

⁹⁹ See Scanlon, *Moral Dimensions*, chap. 4.

¹⁰⁰ See Scanlon, “Punishment and the Rule of Law,” and “Desert, Blame, and Punishment.”

¹⁰¹ See Scanlon, “Desert, Blame, and Punishment,” 22.

¹⁰² See Scanlon, *Moral Dimensions*, chap. 4. Scanlon argues that there is pre- or extra-institutional desert for blaming or condemning someone for a wrongful act.

¹⁰³ See *Report of the Royal Commission*. Lord Denning says here that capital punishment is the most emphatic denunciation of a murderer.

¹⁰⁴ See Scanlon, “Giving Desert its Due.”

¹⁰⁵ As I said earlier, I have not encountered any argument that is persuasive. Moreover, Scanlon has a plausible argument to show that the appropriate response to desert should be only attitudinal, and that (as a response to desert) we can only condemn the wrongdoer. See “Giving Desert its Due.” See also Note 86.

¹⁰⁶ According to Igor Primoratz’s reading in *Justifying Legal Punishment*, Hegel says that we should treat a criminal as a person. As a person, a criminal must have acted on the principle that it is permissible to commit murder when she killed someone. Since she acted on this principle, it must be

permissible for society to act on this same principle vis-à-vis her. Consequently, society must be able to apply this principle when meting out the death penalty. I believe that this argument does not hold because the principle upon which the criminal acted might apply to the victim but not to herself. Suppose the principle she uses is: “Eradicate human garbage from this filthy world.” This principle would allow her to kill the victim (whom she considers to be human garbage), but does not allow others to kill her, since she is not – on her own understanding -- human garbage. While it is doubtful that Hegel’s argument works to support the conclusion that this criminal deserves to be executed, it is clear that an innocent person does not deserve to be punished.

¹⁰⁷ See Brownlee, *Conscience and Conviction*, chap 7 for a critique of Duff’s theory.

¹⁰⁸ Hart, *Punishment and Responsibility*, 234-5.

¹⁰⁹ See Scanlon, “Giving Desert its Due.”

¹¹⁰ See Rawls, “Two Concepts of Rules;” and Scanlon, *Moral Dimensions* and “Giving Desert its Due.” Although Scanlon once held a similar view (as in “Punishment and the Rule of Law”), he now thinks that there is pre-institutional or extra-institutional desert regarding blame or condemnation.

¹¹¹ According to the rules of soccer, a team that scores more goals is the winner. In this institutional sense of “desert,” the winner deserves to win. This does not add anything to the statement that a team wins according to the rules of soccer. Therefore, this sense of desert is not significant. However, a team that is very unlucky – imagine that its strikers hit the post twenty times without a goal – could deserve to win in the extra-institutional sense of desert, even though it does not win the game.

¹¹² See Hart, *Punishment and Responsibility*, chaps. 1 and 2.

¹¹³ See Scanlon, *What We Owe to Each Other*, chap. 6.

¹¹⁴ Such a safeguard does not apply to some minor offenses known as strict-liability offenses.

¹¹⁵ If someone is coerced under a threat of death to commit a deplorable act, he may have the legal defense of duress. According to English law, the threat must be illegitimate. Second, the deplorable act must have been committed “not wholly voluntarily.” Finally, the deplorable act committed must be less serious than the threatened harm the defendant was facing. Perhaps because of its similarity to the defense of necessity, duress should be seen as a justification rather than a legal excuse. (See Peter K. Westen and James Mangiafico, “The Criminal Defense of Duress: A Justification, Not an Excuse - And Why It Matters,” *Buffalo Criminal Law Review*, Vol. 6, p. 833, 2003.)

¹¹⁶ See Tasioulas, “Mercy,” 116-18. Surely, there are other kinds of grounds for mitigation. One type of grounds has to do with the fact that the defendant is already suffering from great misfortunes. Another is based on the remorsefulness of the defendant.

¹¹⁷ See Scanlon, “Giving Desert its Due,” 105, 108.

¹¹⁸ Ibid. Scanlon agrees with Feinberg on this point.