Rawls stabilizes and secures his liberal institutions with a strategy of reconciliation that progressively narrows and even closes the spaces of politics and contest. But the democratic politics and pluralism to which Rawls himself is committed thrive best in a setting that resists the closure that he postulates as their necessary condition. Rawls' theorization of punishment illustrates the point. In his effort to justify punishment and move it beyond the reach of contestation, Rawls is driven back to a discourse that his own theory of (distributive) justice disavows, the discourse of antecedent moral worth: Criminals are demonized as "bad characters" and misfits are viewed by Rawlsian citizens as outside agitators. These impulses touch every politics but they are aggravated, not alleviated, by Rawls' effort to contain contestation.

There are five pages on punishment in John Rawls' *A Theory of Justice*: Five pages (Rawls 1971, hereafter cited TJ: 240; 270; 314–15; 575–77). And in the myriad debates and amendments that have followed the publication of that founding text, the subject of punishment has attracted almost no attention at all. I focus on Rawls' account of punishment for two reasons: First, because it is a fair, if demanding, test of a regime to judge it by its way of dealing with its enemies (and, relatedly, by its way of distinguishing friends from enemies). Second, because the problems that plague Rawls' treatment of punishment are largely the same problems that plague Rawls' treatment of politics, writ large.

NOTE: I am grateful to the National Endowment for the Humanities and to Tulane University's Murphy Institute for providing the leave time during which this paper was written. Earlier versions of the paper were presented to Tulane University's NEH seminars on liberty in April 1991, and to the American Political Science Association Meetings in August of the same year. For their comments and suggestions, I am indebted to this journal's anonymous referees as well as to Rick Teichgraeber, Geoff Smith, Nancy Rosenblum, Richard Flathman, William Connolly, Michael Shapiro, Patrick Neal, Donald Moon, David Mapel, Peter Digeser, Glyn Morgan and Michael Whinston.
Rawls looks to politics and to political theory to provide justice—both distributive and retributive justice—with the closure that it needs to function smoothly. The original position, reflective equilibrium, and the practice of justification that they imply, are all part of an attempt to establish a common basis for the efficient (bureaucratic and jurocratic) administration of justice. For the sake of justice, Rawls seeks an end to political contest, a right and workable resolution of the fundamental rifts that mark American political culture. This administrative politics is enabled and supported by Rawls' expressivist assumption that a set of institutions can be so well—and so thinly—ordered that it fits and expresses, without excess or resistance, the selves it presupposes and the subjects it engenders. Once the right fit is found, the space of dissonance effaced, and conflicts settled, the burdens of political contest are lifted and citizens are left relatively free to live their lives in the vigorously and diversely engaging private realm of justice as fairness. Hence Rawls' wager that punishments, in justice as fairness, will be rare.

Working from a perspective indebted to Friedrich Nietzsche, Hannah Arendt, and Bernard Williams, I counter Rawls' expressivism with the assumption that no set of arrangements succeeds in expressing the self without excess, nor in displacing the ruptures of politics with the ease of administration. I point out that even in Rawls' own regime, the formation of subjectivity engenders resistances in the self while also enabling it. The ethical and political problem posed by this dynamic is not how to avoid it but how to engage its inevitability, how to resist Rawls' drive to depoliticize it, and how to recast the relationship between self and other in a way that allows citizens to be responsive to the remainders of politics rather than bewildered, angered and betrayed by them. I conclude that an ethical practice of punishment is one that eschews the closure of expressivist assumptions and exhaustive justification so as to preserve and proliferate space for resistance to and politicization of the state's use of violence. Moreover, I suggest that the democratic politics and pluralism to which Rawls himself is committed thrive best in a setting that resists the closure that Rawls postulates as their necessary condition.

1 Regarding the most fundamental rift—between liberty and equality—Rawls says that his goal is "to settle a fundamental disagreement over the just forms of social institutions within a democratic society under modern conditions" (Rawls 1980: 518).

2 From Nietzsche, I take the view that all sets of arrangements and all formations of subjectivity (to varying degrees) engender resistances to themselves even while they enable certain forms of life to be. From Arendt, I take the commitment to preserve and expand the spaces of politics in modernity. From Williams, I take the conviction that attempts to resolve moral or political conflicts are often plagued by unruly remainders to which ethics ought to call attention. I outline my position in more detail in Political Theory and the Displacement of Politics.
There is no shortage in *A Theory of Justice* of misfits and criminals who disappoint Rawls' expressivist expectations. These characters—and the dissonant parts of the self that they personify—keep surfacing in Rawls' text, in his thoughts. Focusing on Rawls' justification of punishment—in which he characterizes the criminal as a "bad character"—I ask why: Could it be that these are resistances, somehow engendered by the demands of the scheme itself? Could it be that instead of being marginal curiosities these figures are actually necessary components of a scheme that is continually defining itself against some "other?" Could it be that Rawlsian rationality is not sufficient to itself, that justice as fairness is forced to draw on otherness as a resource to secure itself even as it insists that one of its greatest merits is its refusal to treat others merely as means and not also as ends in themselves?

How should citizens treat the other—not just the criminal, but also the misfit or the dissonant impulse—that is produced by the institutional arrangements that many of them value precisely for their (re-) production of the identities they privilege as their norm? Rawls poses the problem most starkly in his brief remarks on punishment and criminality in a just regime. But the stage for those remarks is set by his theorization of politics as a practice of settlement, a device of reconciliation.

**RECONCILIATION OR POLITICIZATION?**

Rawls stabilizes and secures the institutions of justice as fairness with a strategy of reconciliation that progressively narrows and even closes the spaces of politics. Unfortunately, those closures make it more—not less—difficult for Rawls to sustain his founding ideals of democratic empowerment, pluralism, and mutual respect. Several of Rawls' founding insights illustrate the point: First, Rawls notes (with Weber) that the modern state exercises "a final and coercive authority over a certain territory" and population. It "affect[s] permanently men's prospects in life" (*TJ*: 222). One response to this circumstance might be to politicize the relationship between citizens and the state, to call on citizens—in the name of a radical democratization—to decenter the state as the privileged site of political action by creating (in Nancy Fraser's phrase) "multiple publics" and by recognizing those that already exist, even in the face of the state's longstanding denial of them (Fraser 1991: 56). But this is not a response that Rawls considers. In the name of a more democratic

---

3 Several critics note Rawls' closure of political space: Sandel 1982 (hereafter *LJJ*); Walzer 1981; Barber 1988; Benhabib 1988; Wolin 1989; Mouffe 1990. Of this group, only Mouffe calls for a strategy of politicization that proliferates political spaces.
politics, Rawls' recenters the state that other democratic activists are actively involved in decentering.⁴

Reconciliation, not politicization, is Rawls' goal in A Theory of Justice. Aided by the imaginary of the contract illustrated by the original position, Rawls identifies the conditions under which self-respecting citizens might be reconciled to the arbitrariness of state power and to the state's monopoly on the legitimated means of violence. By theorizing a state-centered arrangement that free and equal rational beings could consent to if they were ever given the opportunity to do so, Rawls distracts attention from the fact that the opportunity never actually arises. Satisfied that justice as fairness approximates a voluntary scheme, Rawls wagers that citizens will not experience the state as a coercive institution.⁵ As a result, Rawlsian citizens are allowed, indeed subtly encouraged, to become relatively passive consumers of the state's goods and services, impervious to the possibility that their survival as a democracy might require them to engage and resist—not simply reconcile themselves to—the state's status as the privileged and legitimate bearer of political power and coercion.

Rawl sees all the advantages of reconciliation—the state distributes the goods, services, and constitutional guarantees that secure justice and stability—but he is inattentive to its drawbacks. He seeks a reconciliation without remainders. Unfortunately, that means that when the scheme does find itself face to face with some of its own remainders (as it does, most dramatically, when it confronts and punishes criminals, the citizens whose experience of the state as a coercive power is perhaps the least subtle or ambiguous), it is surprised, it responds quite violently, and it judges itself to be thoroughly justified in doing so. Disempowered by their reconciliation to the state and its powers, Rawlsian citizens are hard put to articulate or politicize this violence but they are no less subject to it, no less parties to it.

Rawls' closure of political space continues apace as he moves further to secure the reconciliation that he seeks with a second founding insight: No set of arrangements should "encourage propensities that it is bound to repress and disappoint." Any regime that fails to elicit "the pattern of special psychologies" that "either supports its arrangements or can be reasonably

---

⁴ My point is not that the decentering strategy is necessarily and always better. Obviously, that depends on the form that it takes in any particular setting. But neither is the recentering strategy so obviously a boon to democratic politics that it need not defend itself against this powerful alternative. For a recent argument in favor of an extra-statist, democratic politics, see Connolly 1992.

⁵ Justice as fairness "comes as close as a society can to being a voluntary scheme." It "meets the principles which free and equal persons would assent to under circumstances that are fair" (TJ, 13).
accommodated by them” is unjust and unstable (TJ: 541). Rawls means to criticize the tendency of regimes hypocritically to preach one set of principles while practicing another. But the effect of his provision goes beyond that worthy intent.

The propensity of institutions to generate expectations that they then disappoint is a feature of political life. The generation of expectations is a process that resists the kind of control that Rawls seeks. Moreover, the gaps between promise and delivery, between expectation and experience, are spaces of politics in a democratic regime. They are undoubtedly, as Rawls fears, a site of resentment but they are also a site of idealism, of critical reflection and action. The promulgation of an ideal that closes these spaces has a depoliticizing effect that resonates throughout Rawls’ regime and makes it harder, not easier, for Rawlsian citizens to be true to (to rise to the challenge of) the spirit of the democratic scheme that Rawls envisions.

Rawls’ commitment to pluralism is also endangered by his closures of political space. Rawls insists that no single comprehensive vision of the good life may dominate the political space. But he sees no positive connection between politics and pluralism. He makes no provision for the politicizations necessary to maintain plurality in an age of social homogenization. Rawls imagines a private realm in which different forms of life co-exist as lifestyles, there is no contest among them. None feels threatened by the existence of the other. None maintains itself at the cost of the other’s existence. But in political life spaces of plurality and difference are often closed by the sedimentation or naturalization of a dominant culture’s form of life: Its own particular vision of gender identities becomes a standard by which all men and women are measured. Its family structures become a norm of health and safety. Its attitudes toward work and leisure become moralized. Its approved directions of sexuality are touted as natural, not contingent, and deviations from them are treated accordingly, as curiosities, evils, or illnesses. Without the resources of politicization, those who deviate from these norms find that their rights are protected in Rawls’ regime but that they themselves are (at best) disrespected. Rawls does not address this inequity. Unrelated to the basic structure, it is not the subject of justice. But what if it is a product of Rawls’ effective depoliticization of difference? What if genuine pluralism is a casualty of the (state-regulated) public/private distinction that Rawlsian justice postulates as the condition of pluralism’s possibility?

6 Communitarian theorists and some of those influenced by Nietzsche have been more insightful than most liberals in seeing the tendency of communities to treat their own ideals as norms by which others ought to live. “Politicizations” are any responses that engage or contest this tendency.
Rawls does not engage these questions but when he turns his attention to the practice of desert, his argument takes on a new, politicizing energy. Desert, Rawls argues, is incoherent, aporetic, unstable. It rewards differences (in talents and abilities) that are "arbitrary from a moral point of view" (TJ: 107, 102, 104). The bases of desert are not themselves deserved ("all the way down") (Nozick 1974: 225). Rawls' politicization of the practice, his deessentialization of its foundations, lead the parties in the original position to reject antecedent moral worth as a basis of distribution and to commit themselves to a potentially radical redistributive taxation scheme. This is Rawls' difference principle and it "expresses a conception of reciprocity" in which "men agree to share one another's fate, . . . to avail themselves of the accidents of natural and social circumstance only when doing so is for the common benefit" (TJ: 102).7

Rawls' difference principle and the rejection of desert upon which it is based have a powerful appeal. Unfortunately, however, so does the practice of desert. The mesmerizing pull of desert is not finally overcome by Rawls' critique; as I argue below, it returns to haunt justice as fairness when Rawls himself reaches for antecedent moral worth (or unworthiness) to account for the presence of criminality in a just regime and to justify its punishment. In so doing, Rawls relies on desert to serve its traditional function: It explains the inexplicable, it makes sense of evil and it justifies our dealing harshly with it. Problematically, the reappearance of desert also endangers the difference principle but Rawls, because he does not acknowledge desert's return, makes no provisions for its engagement. Another space of politics is closed precisely at the point at which justice as fairness most needs the resources of politicization.

THE POLITICS OF ORIGINATING POSITIONS

Rawls' reliance on the original position to provide the founding force for justice as fairness testifies to his preference for reconciliation over politicization. Some readers of Rawls have suggested that the original position has depoliticizing effects (or undesirable political effects). But no one has focussed on the drift of the position, on its movement within Rawls' text from a contractual situation, to a perspectival position, to a self-ordering mantra.8 The original

7 The willingness of those who are better off to share the fate of those who are less so enhances the voluntary character of the scheme. It gives the former reason to "expect others to collaborate with them when some workable arrangement is a necessary condition of the good of all" (TJ, 103).

8 Sheldon Wolin (1989) and Michael Walzer (1981) worry that the original position (in its contractual capacity) erases the past that politics must confront or express. Seyla
position’s drift does not mar the theory; it is one of the keys to the theory’s success. Indeed, the theory needs—it presupposes—the continued success of all three dimensions of the original position long after the founding of justice as fairness. As contract, the original position renews a sense of shared social purpose; as perspective, it implies an individual commitment to justice and to reconciliation; as mantra, it consolidates the self into the stable subject that Rawls takes as his starting point.

In its best known capacity, the original position represents a contractual situation in which Rawls’ two principles of justice are chosen and agreed to. The process is purely procedural: “there is no independent criterion for the right result; instead there is a correct or fair procedure such that the outcome is likewise correct or fair” (TJ: 86; cf. 136.) The fairness of the procedure is secured by several constraints, most famously by the veil of ignorance which restricts the parties’ knowledge of the details of their empirical identities and situations (“those contingencies which set men at odds and allow them to be guided by their prejudice” TJ: 19). The veil greatly simplifies the process of coming to agreement but simplicity is only part of the point. The veil also ensures that the parties’ choice is moral, not self-serving. From behind the veil, they are unable to choose principles of justice that advance their own particular interests or reward their own particular talents (TJ: 137, 141). The veil of ignorance allows Rawls “to define the original position so that we get the desired solution,” a solution “that nullifies the accidents of natural endowment and the contingencies of social circumstance as counters in the quest for political and economic advantage” (TJ: 141, 15).

But the veil of ignorance also subverts Rawls’ description of the original position as a meeting place, a choice situation, a site of negotiation and agreement. Rawls claims that “the principles of justice are the result of a fair agreement or bargain” but agreements or bargains postulate a plurality and difference that are absent in the original position. Rawls concedes the point when he says that “we can view the choice situation in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached.” Since the original position isolates a rational choice, “the differences among the parties are unknown to them,” the “deliberations of one person are typical of all,” and “each is convinced by the same arguments.” Thus, although Rawls suggests that we “imagine that the parties

Benhabib (1988) and Susan Okin (1990) worry that it homogenizes or erases gender difference, thereby disabling real political engagement or inclusion. None notes the tendency of the position to drift nor does any see the self’s isolation in the original position as a sign that the position serves a mantra-like function.
are required to communicate with each other through a referee," he knows that "such a referee is actually superfluous;" the "parties have no basis for bargaining in the usual sense" (TJ: 139; 12; 263).

Since the parties meet in the original position simply as beings possessed of the capacity to reason, the original position is guaranteed to produce not a mere agreement, but unanimity. "[O]nce knowledge is excluded, the requirement of unanimity is not out of place and the fact that it can be satisfied is of great importance. It enables us to say of the preferred conception of justice that it represents a genuine reconciliation of interests" (TJ: 141–42). Two moments in the account, thus far, signal the original position's first drift from contract to perspective: the twin requirements of correctness and unanimity. As Michael Sandel points out, "any agreements reached' in the original position are fair, not because the procedure sanctifies just any outcome, but because the situation guarantees a particular outcome" (LLJ: 127), in Rawls' words, a "correct" outcome. At the same time as the original position moves from a pure procedure to a guarantor of correct outcomes, the parties' plurality slides to unanimity, and the language of agreement and choice gives way to that of acceptance and acknowledgement (see also TJ: 129; Benhabib 1988: 89ff.; Walzer 1981: 389). "Ideally," Rawls says, the "acknowledgement [of the two principles] is the only choice consistent with the full description of the original position. The argument aims eventually to be strictly deductive" (TJ: 121). The ideal, then, is to close the space of politics and interpretation at the moment of founding. But that ideal is subverted, not secured, by Rawls' reliance on the incompatible languages of choice and acknowledgement.

Michael Sandel notes the incompatibility of choice and acknowledgement, as well as the disjunction in Rawls' references to agreement and unanimity in an original position that is occupied by a single party. But Sandel resolves these problems, noting that it makes sense to say that the (single party) original position produces an agreement if we think of each party as agreeing "to a proposition" rather than with others "with respect to a proposition." On Sandel's account, Rawls' language of choice is a decoy; acknowledgement is the deep truth of the original position (LLJ: 129–30). But Sandel moves too quickly to supplant choice and identify acknowledgement as the real "truth" of Rawls' original position. Perhaps the real truth of the position is not choice or acknowledgement but the position's drift from one to the other. Perhaps

---

9 I am well aware that Rawls does not think that he can achieve the ideal of strict deduction. He certainly knows that the two principles of justice do not follow perfectly from the original position. But he does idealize strict deduction and that affects the regime that lives under its (unrealized) spell. In particular, it affects citizens' attitudes towards eruptions of politics and dissonance in their regime.
the truth of the original position is its structural undecidability, its reliance on the overlapping and incompatible languages of (performative) choice and (constative) acknowledgement, of power and knowledge, or in more Rawlsian terms, of contract and perspective. On this reading, the second dimension of the original position supplements the first, it does not replace or supplant it.\(^\text{10}\)

The original position's second, perspectival dimension reassures the parties in the face of the disruptive, performative power of the first. Rawls' wariness of politics means that the founding of justice as fairness has to be quickly converted into an act of maintenance, the choice of principles domesticated into an acknowledgement of them, the radicality of desert's politicization stabilized by the reassurance of the familiar, a hasty return to intuitions.\(^\text{11}\) Rawlsian justice needs clarity, fixity and stability in order to function smoothly. It thrives on predictable administration. Hence Rawls' insistence that the choice of the principles in the original position is "final" (\textit{TJ}: 135).

Finality, one of the constraints of the fair choice situation, is meant to prevent the parties from choosing provisional principles of justice for now with the intention of revising them at some later date. Thus, once the practice of desert has been rejected, it must not resurface later. Once out of the original position the principles of justice are settled. They frame the process of constitutional construction, and the resulting constitutional principles, together with the two principles of justice, regulate the legislative process. Nothing settled in the first three stages of the founding is contestable in the last stage which is purely administrative, occupied solely with the "application of rules to particular cases by judges and administrators, and the following of rules by citizens generally" (\textit{TJ}: 196–98).\(^\text{12}\)

\(^{10}\) Nor, therefore, can it be understood to "come second" in any chronological sense. It is merely second on my list. Note that the undecidability of the original position means that the Rawls cannot attain the freedom from politics of (mis)interpretation that he sought at the moment of founding. The supplement advertises the inadequacy of both choice and acknowledgement as founding impulses, while also marking a rift in the would-be self-identical base of Rawls' new regime.

\(^{11}\) Excepting, of course, Rawls' de-essentialization of desert, which surpasses (and perhaps even violates) his commitment to set principles that "accommodate our firmest convictions" and "provide guidance where guidance is needed" (\textit{TJ}: 20).

\(^{12}\) The two principles constrain the process of founding in the second stage but the process is no longer strictly deductive at that stage. "It is not always clear which of several constitutions, or economic and social arrangements, would be chosen. But when this is so, justice is to that extent likewise indeterminate" (\textit{TJ}: 201). When this is so, justice is to that extent also politicized. In the frame of indeterminacy there is room for contest and for power to play a role in settling arrangements that the rationality of the original position leaves unsettled. Once the constitutional arrangements are settled, however, this space of contest is closed. There is some debate among Rawls' critics as to whether
The progressive closure of contest in Rawls' regime does not mean that the citizens may never again enter the original position. On the contrary, they may do so at any time. But their reentry is part of a continuing process of depoliticization, its function is not to open up alternative possibilities nor to investigate any unexpected effects of their initial agreement, but to consolidate existing practices and principles. The effect is to displace political action with introspection. If the original position is rightly conceived, each repetition of its operation will produce the selfsame outcome. As Rawls puts it, if the original position is to "be interpreted so that one can at any time adopt its perspective, [i]t must make no difference when one takes up this viewpoint, or who does so." No matter who enters the original position, no matter when, "the same principles are always chosen" (TJ: 139). Thus, the possibility of its perfect repetition is a condition of the original position's success as a perspective. But Rawls' celebration of the position's repeatability exceeds its perspectival needs and suggests that the position has drifted again, this time from critical, distancing, filtering perspective to a mantra-like ritual that orders the self into the form of rational subjectivity that the original position (in its other two dimensions) actually presupposes.

If the result is always the same, why repeat the experiment? One possible answer postulates both the original position's contractual and perspectival dimensions: Over the life of a regime or a person, new dilemmas arise, the frame of indeterminacy is too narrow or too broad. See Flathman (1987) and for the contrary view, Pogge (1989).

Since at each stage citizens are allowed access to more not less information it might seem that the movement from one stage to the next is not a closure but an opening up of political space. It is true that the space of political contest expands together with the breadth of indeterminacy and information in each successive stage of the founding, but it is also true that as citizens move from one stage to the next, the previous stage's contests are closed, settled, rendered inaccessible from their new vantage point.

In a critique of Rawls that is kindred to mine, albeit focused on Rawls' effective closure not of political spaces but of the spaces of moral argument and contest, Stanley Cavell argues that the finality condition is a product of Rawls' mistaken and unfortunate analogy between games and promising. The rules of games may be final and non-negotiable, but in "the moral life the equivalent finality is carried not by a rule but only by a judgement of moral finality, one that may be competently opposed, whose content may then enter into a moral argument, one whose resolution is not to be settled by appeal to a rule defining an institution" (1990: 113).

Even civil disobedience is appropriated by Rawlsian justice. Its point, in justice as fairness, is to correct for the regime's departures from the two principles of justice that are chosen or acknowledged in the original position.

On the impossibility of perfect repetition, see Wittgenstein's discussion of rule-following (1967) and Derrida's treatment of iterability (1989).
new data emerges, new institutions come into being, and a renewed commitment to confront them together is called for or a fresh perspective on them is needed. But this answer works only if we ignore Rawls’ claim that the role of the veil of ignorance is to insure “not only that the information available is relevant, but that it is at all times the same” (TJ: 139). The perfect and unchanging repeatability of the position indicates that its operation is prized for something other than its outcome. If the operation is repeated, perpetually, perhaps that is because the Rawlsian subject is never a fait accompli. Perhaps it signals the regime’s perpetual need to reinscribe (unruly) subjects into the order. Perhaps Rawls counts on the original position to issue not only in an intersubjective agreement among selves but also (repeatedly) in an intrasubjective ordering of the self according to the dictates of Rawlsian rationality and justice.

Rawls’ reliance on the original position to produce the self that it presupposes means that the position is never really original; it is always already being repeated and this is a condition of its capacity to originate Rawls’ justice as fairness. But, more important, the myth of originality that surrounds the position affects the way citizens view political life in Rawls’ regime: Rawls’ reliance on the original position establishes a hierarchical divide between an original, pure, pre-political, noncontingent, introspective position and all other subject positions and positionings that constitute and condition human lives and possibilities. (The point of the original position is to “correct for” the latter TJ: 141.) The original position is the position, not a position and the result of Rawls’ enthronement of it as a single, sovereign perspective is a paradoxical combination of an enormous politicizing and subversive potential (illustrated by Rawls’ de-essentialization of desert) with a practice of repetition that betrays that potential and serves the regime’s needs of maintenance and consolidation with powerfully depoliticizing effects.

When Rawlsian citizens experience dissonance (in themselves or in others), the default is to return to the original position and confirm that from its perspective, the outlaw impulse, desire, or activity in question, is indeed irrational or unjust. In effect, the original position’s myth of origins encourages citizens to respond to dissonance with introspection, a practice that supports the regime’s broader efforts to privatize, naturalize, or dissolve the dissonant 17

When I say that Rawls advocates the original position, not an original position, I do not mean to imply that Rawls prohibits any tinkering with the version of the original position that he presents in TJ. On the contrary, Rawls invites his readers to adjust his account in reflective equilibrium and elsewhere. But whether or not we make adjustments, once we decide on the structure and constraints of the original position, what we have decided on is the structure of a sovereign perspective.
remainders of politics rather than politicize them. The point is illustrated by Rawls’ account of punishment and criminality in justice as fairness.

The Practice of Punishment

In On the Genealogy of Morals, Nietzsche envisions a society so powerful that it does not need to punish: “a society might attain such a consciousness of power that it could allow itself the noblest luxury possible to it—letting those who harm it go unpunished. ‘What are my parasites to me?’ it might say. ‘May they live and prosper: I am strong enough for that!’ ” Nietzsche’s imagined society does without punishment not because its laws are right nor because its citizens become increasingly obedient, but because it overcomes justice (the insistence on right) with mercy, a kind of magnanimity that is “beyond the law” (1969: ii, 10). The scenario is not merely naive. That it initially strikes the reader that way is one of the most telling points of the reflection. The image of a society so powerful that it does not need to punish is meant not to make us yearn unrealistically for the realization of the ideal, but to invite us to reflect critically on why and how societies punish, to think about punishment (its gravity, its sites, and its frequency) as a measure of societal need rather than individual deviance, as a signal of social weakness rather than administrative resolve.

In A Theory of Justice, Rawls also envisions a society that does not need to punish. But if his ideal society does without punishment, that is not because it is powerful enough to afford magnanimity but because its institutionalization of justice makes magnanimity unnecessary: Criminality does not surface in a well-ordered regime. Rawls wagers that citizens governed by relatively just institutions will acquire a “corresponding sense of justice” (TJ: 138). Well-ordered institutions engender well ordered selves and this secures the regime’s stability (TJ: 177; see also 6, 138, 572, 576–77). Gradually, the citizens’ felt need to (re)enter the original position disappears. Dissonance and resistance—both inter- and intrasubjective—diminish.

Hence Rawls’ wager that his scheme will pass the formal tests of stability and congruence. The “strains of commitment” will not be too burdensome in justice as fairness and the institutional order will be congruent with the members’ conceptions of the good (TJ: 580). But Rawls’ wager is premised on the assumption that the institutions of justice as fairness fit and express the (core) self without remainder. As Rawls puts it, the original position, by bracketing the contingent attributes and worldly features that set people at

———

18 Rawls’ regime does have a “public system of penalties,” however, even at the ideal level. Its purpose is not to punish violations of law (violations of law are not an issue until the move to partial compliance theory) but to solve the assurance problem. (TJ: 240).
odds, and the difference principle, by refusing to distribute scarce resources as moral rewards for contingently distributed natural talents, give "expression [to] our nature as free and equal rational beings" (TJ 584; see also 515). This expressivism is what underlies Rawls' belief that justice as fairness does not ask too much of its subjects. It also accounts for his preference for reconciliation over politicization.

But Rawls' expressivist ground is insecure. His identification of a human (moral) nature with a free and equal rationality soon gives way to the admission that some beings are free and equally rational but others are oddly irrational, even immoral, even when raised in a just regime. His identification of the core of the self with freedom and reason is soon tempered by the admission that there are some selves whose core constitution is anything but rational. Rawls tries to turn these figures to his advantage by making an example of them. He stabilizes his expressivist assumptions by contrasting his free and equal rational beings with a now spectral Other, the "bad character." Unfortunately, Rawls' expressivism leaves him unprepared for the character upon whom he relies. From Rawls' perspective, the bad character is so bewilderingly deviant that Rawls is forced back into a discourse that he disavowed: the discourse of antecedent moral worth.

The purpose of the criminal law, according to Rawls, is to uphold basic natural duties, those which forbid us to injure other persons in their life and limb, or to deprive them of their liberty and property, and punishments are to serve this end. They are not simply a scheme of taxes and burdens designed to put a price on certain forms of conduct and in this way to guide men's conduct for mutual advantage. It would be far better if the acts proscribed by penal statutes were never done. Thus a propensity to commit such acts is a mark of bad character, and in a just society legal punishments will only fall upon those who display these faults. (TJ: 314-15)

Rawls permits punishment, as a practice, to work as an incentive scheme but he insists that individual punishments be justified deontologically, not consequentially. The consequentialist justification of individual punishments leads too easily to a willingness to punish the innocent in order "to promote effectively the interest of society." That would undermine the priority of right as

19 The strategy serves another purpose too: It gives a new urgency to the Rawlsian insistence on the truth of expressivism. "Our" free and equal rationality is a source of protection, defence, and reassurance for the "we" that is faced with the other. Rawls never asks whether "our" insistence helps to produce the other against whom the "we" then defines and defends itself.

20 These arguments distinguishing individual punishments from the general practice are not made explicit in TJ though they are implied by what Rawls says in the passage
well as punishment's incentive structure which depends upon the public belief that only the guilty are punished.

Justice as fairness boasts that it punishes persons only for acts that they commit but, on Rawls' own account, it is actually the brand of character from which these actions emanate that is outlawed. Actions are guided by punishment's incentives while characters are criminalized by punishment's retribution. In Rawls' regime, character is a thoroughly juridical concern, even if it is not actually punished in the absence of actions that speak it.

And yet, this concern with character seems to be very much at odds with Rawls' other commitments regarding the self, particularly with his rejection of antecedent moral worth as a basis of distributive justice and, related, with his treatment of the self as the bearer of attributes and talents that belong to it only contingently. Rawls does not say that the person with a "bad character" deserves his punishment, but he is markedly less charitable toward this person than he is toward other citizens of justice as fairness. The citizen who is incapable of holding a job because he is indolent is not, on Rawls' account, responsible for his situation. The fact that he has this attribute and not others is "arbitrary from a moral point of view." This is why Rawls citizens agree to "share" his "fate" and treat the distribution of talents and abilities as "common assets." They do not, however, take the same view of the criminal possessed of "bad character." They do not offer to "share" the burden of the criminal's "fate." They do not view his moral failings as a common liability. Rawls' account would be more consistent if he did allow punishment to work simply as an incentive scheme but he does not, for two reasons.21

First, the most important feature of Rawlsian justice is that it approximates a voluntary scheme and need not, therefore, (indeed must not) rely (at least not overtly) on coercion as an incentive for subscription. Second, there is a big difference between the fact that some people command higher wages and social rewards than others and the fact that some people heed their sense

---

21 These two reasons derive directly from Rawls' vision of justice as fairness. As I noted above, there is a third, more general reason for Rawls' rejection of the simple incentive scheme view of punishment at the outset; That view cannot rule out the punishment of the innocent in particular cases.
of justice while others do not. The difference between the indolent laborer and the criminal, on Rawls' account, is that the former is moved by certain attributes that are only contingently his while the criminal, by contrast, is unmoved by his sense of justice even though his sense of justice is deeply and rationally his, whether he affirms it or not. Indeed, by failing to affirm his sense of justice, by committing himself to a conception of the good that is incongruent with principles of right, the criminal denies his own moral personality.

This part of Rawls' argument is "connected with the Kantian interpretation," according to which "the desire to act justly and the desire to express our nature as free moral persons turn out to specify what is practically speaking the same desire" (TJ: 572). The desire is satisfied when we affirm the sense of justice. By affirming the sense of justice, we "realize our nature" (TJ: 574). By denying it, we do not: "What we cannot do is express our nature by following a plan that views the sense of justice as but one desire to be weighed against others. For this sentiment reveals what the person is" (TJ: 575). Contingent talents and attributes do not reveal what the person is. They tell us nothing about whether a person is of good or bad character. Only the sense of justice reveals that.

To some, however, the sense of justice is not so self-revealing. Rawls appeals to these people explicitly, giving them three reasons for compliance, but presumably these people are unmoved by his appeals: The desire to act justly and the desire to express their nature as free moral persons are not overriding desires to them. They do not care about the psychological costs of deception nor about "the loss of spontaneity and naturalness" that Rawls warns them are bound to result from the need to disguise their free-riding in a system whose conception of justice is public. Perhaps there are compensating pleasures to be had in beating the system, even when the system is just. Or perhaps they prefer the disguise that enables their free-riding to the one they would have to assume in order to fit themselves into the community of justice as fairness. Thus, they apparently are unmoved by Rawls' final appeal as well, the promise of sharing in the good of belonging to this co-operative scheme. Rawls may be correct when he says that "to appreciate something as ours, we must have a certain allegiance to it." Yet these people might respond that "to have a certain allegiance to something, we must first appreciate it as ours." The device of the original position is supposed to solve this problem of chronology and identification but Rawls admits that some people will just fail to identify with the regime (TJ: 571). For people such as these, there is nothing internal—no sense of loyalty, no identification, not even rational self-interest—to prevent or discourage them from devoting themselves to conceptions of the good that are not congruent with the principles of justice. For
people such as these, Rawls' three reasons for compliance simply do not persuade; perhaps these are people that will not be reasoned with.

This limitation on reason and reasoning is what forces Rawls to confront the problem of punishment. But it is not to those who are to be punished that Rawls addresses himself. Rawls justifies punishment to those who mete it out, to those who adhere to and identify with the scheme, to those whose rationality is modelled by the original position. The only concern here is whether citizens can punish others and exact retribution without dirtying their hands, without being haunted by remainders. Is it just for those "who do affirm their sense of justice" to require others who do not "to comply with just institutions?" (TJ: 575).

Rawls answers yes, without remainder; it is just to enforce compliance in justice as fairness. Since the question is posed behind the veil of ignorance, punishment is justified without explicit recourse to responsibility or desert; Rawls relies solely on rationality. The parties, "having agreed to the principles" see that "it is rational to authorize the measures needed to maintain just institutions." As for those who find, after the veil of ignorance is lifted, that "being disposed to act justly is not a good for them," well they "cannot deny these contentions." Rawls concedes that it "is, of course, true that in their case just arrangements do not fully answer to their nature, and therefore, other things equal, they will be less happy than they would be if they could affirm their sense of justice. But here one can only say: their nature is their misfortune" (TJ: 576).

Confronted by those unfortunate enough to experience a rift between themselves and the regime, Rawls comes as close as he ever does to admitting that justice as fairness is built on an ineliminable moment of arbitrariness: Some will feel at home, completed and realized by the constructions of justice as fairness, and others will not. To those who do not, to those who fail to affirm their sense of justice, "Their nature is their misfortune" is all that Rawls can say. However, when the problem is not the inability of a citizen to affirm his sense of justice but that citizen's inability to command a high income in the marketplace, Rawls instructs us to say something very different: his nature is our misfortune. If this is not an inconsistency it is only because, on Rawls' account, talents and attributes are randomly and inequitably distributed among persons, while the sense of justice is assumed to be equally available to everyone. The problem is that some cannot "affirm their sense of justice."

We might think that the distribution of this inability (to affirm) is, like the distribution of talents and attributes, arbitrary from a moral point of view but, again, there is a difference. Justice as fairness is not required to respond to the first arbitrariness because it does respond, rationally, to the second by
ordering its institutions in accordance with the two principles of justice. In so doing, it has done all that can reasonably be expected of it. It has set up a just basic structure according to principles that those not disposed to act justly would themselves affirm from the standpoint of the original position. It has eliminated thereby all the environmental factors, all the injustices both arbitrary and systemic that in other regimes motivate persons to criminal behavior.

Does this not account for Rawls' insistence that the object of punishment in justice as fairness is "bad character?" In justice as fairness, what else could it be? There is no perpetual underclass, here; homelessness is not a systemic problem. There is genuine equality of opportunity, as well as a public commitment to the equal consideration of others and reciprocity in the course of economic advance and development. In the absence of systemic injustice, there is nothing that can account for criminal behavior; criminality must be a symptom of sheer perversity, orneriness, a tic of some kind, a defective character. Criminality, in short, must be an assault on the system from some outside, from some mysterious and terrifying state of nature: It must be sociopathic.

The key to Rawls' account of punishment, and one probable reason for its brevity is that the system is simply not responsible for the production of criminality. Criminality is extra-systemic. What this means, however, is that the notion of antecedent moral worth (or, in this case, unworthiness) whose

---

22 This plays a crucial role in Rawls' justification of limiting the freedom of the intolerant. Their freedom may be limited "for the sake of equal liberty under a just constitution the principles of which the intolerant themselves would acknowledge in the original position" (TJ: 220). This means that some intolerant conceptions of the good are not just wrong (or intolerable), they are "irrational" (ibid.: 149–50). Since even a racist cannot deny this he, in a sense, wills his own suppression or at least he would will it—or give up his racist programmatic—were he to take up the standpoint of the original position. Thus, "by limiting his actions" we express our "respect for him as a person" (ibid.: 519).

23 David Mapel suggests a more generous reading: The confinement of justice to the basic structure means that the system is loosely arranged; communities may educate their children into a vast array of belief systems, even into beliefs that are intolerant and unjust, as long as they do not act on them. Some will act on them, however, and when they do it is clear that social unions or the family are the source of criminality, not individual character and not the system itself qua basic structure. In my view, the more generous reading may apply to Rawls' justification for limiting the freedoms of the intolerant but it does not apply to the practice of punishing criminals, since Rawls explicitly attributes criminality to bad character. Mapel is right, though, to argue that Rawls aspires to looseness in the arrangements of justice; if my emphasis differs from his it is because I want to highlight the tight moments that enable the loosenesses that Mapel (1989) admires.
viability Rawls denied in his account of distributive justice is now assumed in order to account for and justify institutions of punishment and their regulation of compliance.\textsuperscript{24} By contrast with distributive justice, which responds only to institutionally generated demands ("legitimate expectations"), retributive justice is there precisely to respond to extra-institutionally generated demands, to irrational expectations and resistances. By blaming the latter on "bad character," justice as fairness effectively assumes responsibility only for the production of good characters, of those fortunate enough to find themselves well-fitted to the constructions of justice as fairness.\textsuperscript{25} Like many parents

\textsuperscript{24} Michael Sandel makes a similar point but he responds to Rawls' re-introduction of antecedent moral worth by arguing that it is an inconsistency that would have to be ironed out in reflective equilibrium. Sandel never asks why antecedent moral worth makes its reappearance here. Happy to mark its return, he does not try to identify the pressures that drive Rawls' account to this inconsistency nor is he any more concerned than is Rawls about the way that this theorization of punishment constructs criminality and affects those who are punished as well as those who are not (\textit{LLJ}; 89ff.). Thomas Pogge also notes the apparent inconsistency in Rawls' account but he, too, irons it out (1989: 76–85). Defending Rawls against the charge that in his regime "there is no Desert at all" (the capital 'D' denotes "moral deservingness"), Pogge points out that Rawls "counteracts this impression by insisting that entitlements arising under the criminal law do involve Desert" (83). The return of Desert in retributive justice is a virtue, for Pogge, its inconsistency with Rawls' earlier rejection of the practice, notwithstanding. Fortunately for Pogge's Rawls, the inconsistency is a product of Rawls' having burdened himself "with a stronger commitment than necessary" (76). Rawls' distributive scheme can be justified as "an unmoralized system of incentives;" it does not depend on a radical de-essentialization of Desert (83). Indeed, Pogge argues, it is rational to "unmoralize" economic benefits while maintaining a moralized penal system. Intent on buttressing Rawls' depoliticization of crime and punishment, Pogge, too, fails to ask what it is about criminality and punishment that drives Rawls back to Desert after his initial eagerness to dispense with it.

\textsuperscript{25} Stanley C. Brubaker (1988) argues that liberals cannot punish (by which he means that they cannot justify punishment) because they reject the ground of antecedent moral worth that, from Brubaker's perspective, makes sense of punishment. Brubaker is mistaken on two points: First, as we have just seen, Rawls (re)introduces antecedent moral worth to ground punishment and is therefore able, presumably, to punish to Brubaker's satisfaction. And second, there may be good ethical reasons to prefer a practice of punishment that is a little less justified than the practice that Rawls and Brubaker respectively envision. Rawls and Brubaker agree that punishment must be justified, that gaps in reasoning must be closed. It is possible, however, that less rather than more closure in justifying punishment is appropriate for a liberal democratic society. The achievement of closure in justifying punishment tends to depoliticize the practice, to demonize the wrongdoer, to treat the other merely as a means. Perversely enough, it is Rawls' attempt to escape this last problem (which he identifies with utilitarian justifications of punishment) that leads him to justify punishment with reference to the (antecedently) bad character—the otherness—of the wrongdoer.
in denial, justice as fairness will confess only to having raised its good children. It is mystified by the others. Who can tell where they came from? ("They must have made a mistake at the hospital.") Rawls is driven to this chimera because of his own insistence that justice as fairness not "encourage propensities that it is bound to disappoint" (TJ: 541). If justice as fairness is to pass this test, it must deny responsibility for the criminal propensities present in the regime because those propensities are the very ones that this regime is out to disappoint.

The same logic governs Rawls' discussion of envy, an "inexcusable" and irrational attitude in a regime so rationally ordered. Once again, Rawls' primary concern is to delineate the responsibility for the disruption. Does it emanate from the self or from the regime and its institutions? Is it a feature of political life or is it a private or natural (dys)function? Does the range of permissible inequalities and the terms of their permissibility give rise to this potentially destabilizing passion? Or is it generated, inexplicably, from an outside source? Rawls argues that envy is excusable and therefore politically significant only when it is the product of injustice; only then is it "not irrational" because only then is it occasioned by "circumstances where it would be unreasonable to expect someone to feel differently" (TJ: 534).

As a just society, justice as fairness "eliminates the conditions that give rise to disruptive attitudes" like envy by treating self-respect as a primary good, by limiting the range of inequalities, and by requiring that inequalities be justified to the least well off (TJ: 144). In this well-ordered society, then, the rational individual "is not subject to envy" (TJ: 530) and that means that when envy does surface in Rawls' regime, it is as a "disruptive attitude," irrational, inexcusable, inexplicable, and, most important, extra-systemic.

Rawls wagers that as the rational capacities of subjects in justice as fairness are heightened by living in a well ordered regime injustice and envy will

---

26 Of course, no regime can control the expectations to which it gives rise. That is why Rawls distinguishes between legitimate and illegitimate expectations—in order to stabilize this criterion of stability. Now the regime is only responsible for some of the expectations to which it gives rise. But, even so, how do we distinguish legitimate from illegitimate expectations? What difference is there between propensities that are regime-encouraged and those that are extra-institutional? Is a legitimate expectation one that was intended by the framers or the lawmakers? Or is it typical of the general effects that a particular piece of legislation has on the population? Should we rely on statistical or psychological studies to determine the frame of a reasonable response?

27 Envy is produced by a "lack of self-confidence in our own worth combined with a sense of impotence. Our way of life is without zest and we feel powerless to alter it." The self-respect that justice as fairness claims to secure provides both of the needed antidotes: self-confidence and empowerment (TJ: 440).
diminish and maybe even disappear. He assumes that "men's propensity to injustice is not a permanent aspect of community life; it is greater or less depending in large part on social institutions, and in particular on whether these are just or unjust." And so he concludes that "A well-ordered society tends to eliminate or at least to control men's inclinations to injustice" (TJ: 245). Rawls' claim that certain levels of inequality or other forms of injustice are bound to give rise to greater and more virulent forms of envy and other "disruptive" feelings is plausible. Less plausible are the rest of the assumptions in the Rawlsian package: that a set of institutions can assert such strict intentionalist control over the expectations to which it gives rise; that institutions and aspirations can be ordered in a way that fits and expresses the self without remainder; that resistances will not be aggravated but only alleviated by the demands and aspirations of a rational regime; that subjects in this regime do not, therefore, need the resources of politicization because the conditions for successful reconciliation can and have been met. These are the assumptions that lead Rawls to treat sources of disorder, forces of unsettlement (and politicization), as outside agitators. Envy and criminality disrupt a system that claims not to engender them, a system that insists that the outside agitators need only be removed—punished, banished, or suppressed—and order will be restored.

One result of this approach is Rawls' refusal to recognize the phenomenon of criminality as a valid test of the arrangements he advocates. The fact that some (or even many) persons exist for whom the right and the good are not congruent does not have the standing to call into question the justice of justice as fairness. 28

It can even happen that there are many who do not find a sense of justice fits their good; but if so, the forces making for stability are weaker. Under such conditions penal devices will play a much larger role in the social system. The greater the lack of congruence, the greater the likelihood, other things equal, of instability with its attendant evils. Yet none of this nullifies the collective rationality of the principles of justice; it is still to the advantage of each that everyone else should honor them. (TJ: 576)

Here Rawls' argument shifts as the tests of stability and congruence give way to rationality, but it is a rationality that need not move anyone. Even if many people find themselves ill-fitted to the scheme, it would still be the case that it is "to the advantage of each that everyone else should honor" it. Even if the propensities to injustice and envy are never diminished and are perhaps even

28 The same argument applies to envy: The fact that many people in justice as fairness feel envious does not mean that there is anything amiss in the arrangements. It probably means that the envy in question is of the inexcusable variety.
aggravated, even if punishment plays a heavy-handed role in maintaining stability, the rationality of the scheme is not in question. What would have the power to nullify the collective rationality of the scheme? How frequent and how grave would punishments have to be before the rationality of the scheme would be problematized, or even affected, by phenomenal outcomes? Rawls says that it would have to be frequent and violent enough to make the scheme practically untenable. But that could just as well mean that as long as punishment works, as long as it keeps the lid on those whose nature is their misfortune or succeeds in putting them away, then their nature will not become our misfortune.29

This is not the scenario that Rawls envisions, however. His confidence in his expressivist wager is so great that he speculates that “in a well-ordered society sanctions are not severe and may never need to be imposed” (TJ: 240). His confidence is Kantian. Kant, too, thought that

the more closely the legislation and government were made to harmonize with this idea [that a constitution ought to allow “the greatest possible human freedom in accordance with laws which ensure that the freedom of each can coexist with the freedom of all the others”], the rarer punishments would become, and it is thus quite rational to maintain (as Plato does) that none would be necessary at all in a perfect state. (1970: 191)

In the Rawlsian (and Kantian) ideal, the state punishes rarely not because it is magnanimous or powerful (in Nietzsche’s senses) but because its citizens are law abiding, its subjects are, for the most part, well ordered. Unfortunately, the ideal has an ambiguous effect on punishment. Because Rawlsian punishment is so thoroughly justified by the bad character, because it is so well deserved, those who punish are not sullied, not even touched, by their actions. By justifying punishment so well, Rawls removes or relaxes all doubts and reservations, all ambivalences, all remnants of power and arbitrariness. He makes life neat and easy after the founding . . . for those whose nature is their good fortune. He does not explain, however, why we should idealize a practice of punishment in which the citizens who punish have no hesitations about the use of power, no ambivalence about inflicting punishment, no remorse about depriving others of their liberty nor about exacting time from a human life.

Rawls seeks closure for punishment but it is elusive. His return to antecedent moral worth to justify punishment and define criminality is ill at ease

29 There is no reason to reconsider the basis of the scheme “so long as the conception of justice is not so unstable that some other conception would be preferable” (TJ: 576–577). Since punishment is a stabilizing institution, the citizens’ reliance on it is not, on Rawls’ account, in itself a reason to think that some other conception would be preferable.
with his admission that some (and possibly many) citizens are simply unlucky to be ill-fitted to the regime’s needs. Rawls knows that they do not deserve their ill fittedness; it is their contingent misfortune. And yet it is at precisely this moment that desert reappears in Rawls’ account. Why? Is it because this is the moment at which Rawls is most anxious to disavow any relation between his well ordered subjects and the other? Were it not for Rawls’ disavowal, there might be room to consider the possibility that his expressivist wager actually plays a role in the production of bad characters (as well as good ones). There might be room to think about the bad character as a remainder of Rawls’ politics. Is this the possibility that Rawls is anxious to rule out: that he himself turns the unfortunate misfit into the bad character who deserves what he gets in order to protect and stabilize the just order that the Rawlsian wager has enabled up to this point? Or is the role of the expressivist wager in the production of the deserving criminal more subtle than that? Does the citizens’ public faith in the success of Rawls’ wager simply leave them so unprepared for the appearance of misfits and criminals that they react (as does Rawls himself) with a shock, surprise, and condemnation that serve both to justify the suppression of the other and to consolidate the citizens themselves into a “we” that can punish confidently, without remorse or remainder.

30 Rawls’ distinction between the bad character and the person who is hard hit by bad luck is reminiscent of the nineteenth century English distinction between the deserving and the undeserving poor. The undeserving poor did not deserve help (of public or private charities) because they deserved to be where they were. Their poverty was due to character flaws (like alcoholism) or weaknesses that they could do something about if they wanted to. The poverty of the deserving poor, by contrast, was due solely to bad luck. They were poor through no (character) fault of their own and so they had a claim to help or charity. As in Rawls, the pairing of the contrasting categories stabilizes each one, making both charity and condemnation possible, satisfying both of the urges experienced by the relatively well off when they are confronted by the poor. In contemporary American politics, the character/luck distinction continues to operate, crystallized in the figures of the urban teen mother contrasted with the unemployed auto worker.

31 From similar concerns regarding Rawls’ focus on the “idea of expressing our nature,” Stanley Cavell reasons to a somewhat different conclusion: “But doesn’t this assume a limitation in the kinds of scoundrel or saint there may be, in particular that there are none touched with the satanic, with an intolerance precisely for membership, for reciprocity in the intelligible, or any other, realm?” My own argument suggests that although Rawls does indeed begin from an expressivist assumption, the effect of the assumption is (pace Cavell) precisely to turn mere scoundrels or rogues into beings that are touched with the satanic or, as in Rawls’ vocabulary, with the mark of bad character, and are therewith treated accordingly, which is to say, they are exorcised. Perhaps the argument developed here can provide Cavell with the clarification he seeks “of the
POLITICIZING ALTERNATIVES

If the demonization of the other into an antecedently bad character is one of the effects of the institutionalization of the Rawlsian wager, that is a powerful reason to reject the project. Perhaps a theory of justice should strive not to justify punishment so well that it is moved beyond the reach of politicization but to insist that although justification is always a part of the practice of punishment, it is never seamless, never complete.\(^{32}\) The problem with Rawlsian punishment is its aspiration to univocality. Rawls imagines a practice in which there is no moral anguish, no unruly excess, no joy in another's suffering, no troublesome doubts, only a sense of justice. But justifications of punishment are not so well ordered. They always draw on a range of conflicting assumptions and beliefs (vengeance, rehabilitation, self-preservation, or, as in Rawls' own case, desert and the need for stability) and this marks punishment as a tragic situation, in Bernard Williams' sense: It is never simply the right thing to do. It is not something that we ever get right. Punishment draws on a palimpsest of justifications because no single set of assumptions or beliefs is capable of putting our doubts to rest. And this is appropriate. Because of its pervasive, violent effects on human and political life, punishment is one of the most important sites of politicization in any set of arrangements. Its rifts should not be sealed. These are spaces of politics, power, and resistance and they should be preserved, even aggravated, for the sake of the remainders of politics.

Rawls' ideal, however, is to leave no space for these sorts of questions, alternatives, and remainders to come up. By treating punishment as an expression of right, not power, Rawls seeks to reconcile citizens to a practice that might otherwise disturb them. The reconciliation empowers citizens to maintain and protect their sets of arrangements. But it disempowers those (bad characters as well as others) who might seek to politicize punishment. Moreover, Rawls' apparently thoroughgoing rejection of desert as a principle of distribution leaves the citizens unprepared to engage it upon its return as the ground of retribution. These depoliticizing effects are problematic because they are justified, after all, by a wager. If Rawls' expressivist wager turns out to be wrong, if the regime fails to take, if some or even many citizens find that their nature is their misfortune, the result is not more politicization and resistance, by Rawls' own admission, it is more punishment (TJ, 576). The frequency of punishment, in itself, signals nothing more than the fortuitous

\(^{32}\) Justification not in Rawls' sense of "everything fitting together into one coherent view" (TJ: 579) but in the less strict sense of giving reasons.
presence in the regime of a large number of bad characters. It does not testify to any failure of the regime or its aspirations. And the citizens, dispossessed of the cognitive and political resources to analyze, trace, or engage this sequence of events, are disempowered to resist or politicize them within the frame of justice as fairness.

I have been reading Rawls through the lens of a counterwager generated by the Nietzschean claim that the self, an original multiplicity, has no univocal core and is therefore not only enabled by its construction into a subject but is always resistant to it as well. Up to this point I have granted Rawls his assumption that the space of punishment is intersubjective but the counterwager invites further reflection on this front: Is there not an intrasubjective dimension to punishment too? The spectral power of the bad character is diminished somewhat and the force of the counterwager enhanced if we think of the criminal and the bad character as personifications of those aspects of the self that are resistant to the formation of responsible subjectivity. From this perspective, resistance is located not only in distinct individuated subjects possessed of (or by) bad character; it is experienced

---

33 The counterwager does not insist on its own truth in opposition to Rawls'; it calls attention to the effects of the Rawlsian wager in the day to day life of the scheme. The question, then, is not which wager is right but what effect does each wager have on the society that is founded on its ground? Which one is more sensitive to the remainders of politics? Which one produces remainders less insistently? Likewise, the question raised by the need to punish is not whether punishment can be justified but what effect does its being justified have on the way that punishment is practiced in Rawls' regime?

As these questions indicate, my (broadly) Nietzschean critique of Rawls should by no means be confused with the early wave of Rawls criticism that also drew on Nietzsche. Allan Bloom, for example, saw in the aristocratic elements in Nietzsche's thought the resources for a critique of Rawls that is very different from my own (1975).

34 Rawls implies that punishment simply maintains inter-subjective order, that the problem is to establish the conditions of cooperation among already well-ordered selves. The intra-subjective dimension of punishment is most apparent when Rawls turns his attention from bad characters to bad faith. The punishment that is directed at bad faith (not characters) is part of a daily, intra-subjective, self-ordering process. In providing for the penal mechanism, the parties seek to insure themselves against being the sort of character who is unwilling or unable to share in the goods of justice as fairness. The problem of bad faith—the assurance problem—is posed by the citizens' suspicion that others might be free-riding and that is a problem even at the ideal level (TJ: 240; see also 497, 577). A well-ordered society is especially vulnerable on this count. It relies heavily on rational self-interest as an incentive for allegiance but, of course, rational self-interest is also a great incentive for free-riding. For a reading of Hobbes that emphasizes the role of self-interest in creating, not solving, the free-rider problem, see William Connolly (1989).
(to a greater or lesser extent) by all subjects and so the criminality that Rawls punishes retributively is, contra Rawls, an integral part of the subject. Likewise, the inexcusable envy that rises up unaccountably in some, on Rawls' account, might be taken to represent a range of "disruptive" attitudes and impulses that are rationalized into silence by the would be well ordered subjects of this well ordered regime.35

If dissonance and resistance mark all subjects to some extent, then the parties deliberating in the original position are underinformed on some important points. They know that they may turn out to be black or white, middle or lower class, highly talented or unremarkable. But they are not told that they may turn out to be one of those unlucky louts whose nature is his misfortune, a bad character who is committed to a conception of the good that is incongruent with the principles of right and is unable to affirm his sense of justice. They do not consider the principles of justice from the perspective of the outlaw even though each of them is likely to experience outlaw impulses and even though some of them are likely to find that they are themselves outlaws in this regime. They do not know that there is an ineliminable element of arbitrariness at the very foundation of this regime and that its effects and reverberations will be felt by most of them daily. The parties assume that they will be well and easily fitted to the Rawlsian order and that expectation will often be disappointed. When that happens, Rawlsian citizens are once again disempowered, unable to politicize their disappointment. In justice as fairness, the potentially political space, the space between expectations and experience, is closed, disappointment is privatized, and introspection—not political action or organization—is counselled.

The counterwager suggests that a set of arrangements that chastens its aspirations to closure is less likely than justice as fairness to convert resistances into criminality, otherness into bad character. The renunciation of closure invites the articulation of resistances instead of branding them as antecedent irrationalities for which the system is not responsible and to which it need not respond responsibly. By relaxing the demand that subjects "fit" identifies—by renouncing expressivism—it hopes to diminish the propensity to self-loathing, vengeance, and violence fostered by regimes that profess

35 Once again, Cavell's reading of Rawls dovetails with mine. He, too worries that one of the effects of Rawls' particular model of conversational justice is "the deprivation of a voice," the subtle silencing of the voice of the other (xxxvii). Cavell explores these effects through Ibsen's Nora and raises powerful doubts about the Rawlsian project (101-26). But, by turning to Ibsen, Cavell also leaves room for Rawls' readers to dismiss the critique as one that runs parallel to Rawls' concerns without intersection. Hence my decision to trace the deprivation of voice as it manifests itself in Rawls' own text, repeatedly.
themselves free of any part in subjectivity failures. And, in its insistence that ill-fittedness is not antecedent to any scheme, and neither is well-fittedness, this politicizing alternative is far more consistent than is Rawls in rejecting antecedent moral worth. It is also more open and honest about the element of arbitrariness that marks the practice of punishment in any regime.

The counter wager is just a wager, too, but if it fails, the result will be more politics than one might anticipate, not more punishment. Or, better, since any increase in punishments is assumed to be politically significant, the increased frequency of punishment is, contra Rawls, taken as a signal that something political is amiss and increases in punishment are, therefore, seen as a reflection of a societal need that bears examination, not as an independent measure of the number of bad characters resident in the regime. Moreover, those engaged in this political contest will find empowerment in the unruliness of punishment's justifications.36 Thus, if a more politicizing theory can do better than Rawls it is not because it escapes the problems that Rawls faces but because it (unlike Rawls) recognizes the impossibility of such an escape and engages that impossibility politically, institutionally, and discursively.

REFERENCES


36 Rawls could argue that, in the event of his expressivist wager's failure, the result is also likely to be not more punishment but more politics since when punishment ceases to function well as a stabilizing mechanism, citizens are well advised to seek alternative arrangements, to re-found their regime. But, since Rawls does not permit the frequency of punishments, in itself, to be taken as a signal of the regime's inadequacy, and since punishment is so well justified in and by the regime and its citizens, the regime could conceivably maintain itself as a stable entity even if it punished quite frequently. Thus, the problem is that punishments would have to be quite frequent and ineffective before re-founding could become a viable alternative.


Received: September 22, 1991

Accepted for Publication: February 3, 1992