

10

LIBERALISM AND THE LIMITS OF JUSTICE

Michael Sandel

The notion of possession leads naturally to claims of desert and entitlement. The argument over what people possess, and on what terms, has a direct bearing on the question of what people deserve or are entitled to as a matter of justice. It is to the issues of desert and entitlement that we now turn, to consider the second strand of Nozick's critique of justice as fairness. Rawls rejects the principles of natural liberty and liberal equality on the grounds that they reward assets and attributes which, being arbitrary from a moral point of view, people cannot properly be said to deserve, and adopts the difference principle on the grounds that it nullifies this arbitrariness. Nozick attacks this line of reasoning by arguing first that arbitrariness does not undermine desert, and second that, even if it did, a version of natural liberty and not the difference principle would emerge as the preferred result

Stated in terms of possession, Rawls' objection to natural liberty and liberal equality is that under these principles, persons are allowed unfairly to benefit (or suffer) from natural and social endowments that do not properly *belong* to them, at least not in the strong, constitutive sense of belonging. To be sure, the various natural assets with which I am born may be said to 'belong' to me in the weak, contingent sense that they reside accidentally within me, but this sense of ownership or possession cannot establish that I have any special rights with respect to these assets or any privileged claim to the fruits of their exercise. In this attenuated sense of possession, I am not really the owner but merely the guardian or repository of the assorted assets and attributes located 'here'. By failing to acknowledge the arbitrariness of fortune, the principles of natural liberty and liberal equality go wrong in assuming that 'my' assets belong to me in the strong, constitutive sense, and so allowing distributive shares to depend on them.

Expressed in terms of desert, Rawls' objection to the principles of natural liberty and liberal equality is that they reward assets and attributes that people cannot properly be said to deserve. Though some may think the fortunate deserve the things that lead to their greater advantage, 'this view is surely incorrect'.

It seems to be one of the fixed points of our considered judgments that no one deserves his place in the distribution of native endowments, any more than one deserves one's initial starting place in society. The assertion that a man deserves the superior character that enables him to make the effort to cultivate his abilities is equally problematic; for his character depends in large part upon fortunate family and social circumstances for which he can claim no credit. The notion of desert seems not to apply to these cases. (104)

Because no one deserves his good luck in the genetic lottery, or his favored starting place in society, or for that matter the superior character that motivates him to cultivate his abilities conscientiously, no one can be said to deserve the benefits these assets produce. It is this deduction that Nozick disputes. 'It is not true,' he argues, 'that a person earns Y (a right to keep a painting he's made, praise for writing *A Theory of Justice*, and so on) only if he's earned (or otherwise *deserves*) whatever he used (including natural assets) in the process of earning Y. Some of the things he uses he just may *have*, not illegitimately. It needn't be that the foundations underlying desert are themselves deserved, *all the way down*' (1974: 225).

Now what are we to make of this claim? If I do not necessarily have to *deserve* everything I use in producing a thing in order to deserve the thing, what *does* my desert depend on? Nozick says that some of the things I use I 'just may *have*, not illegitimately' (and, presumably, possibly arbitrarily). Once again, the notion of possession enters the scene. To see whether my having a thing, not illegitimately, can enable me to deserve what it helps me produce, we must explore in greater detail the relation between possession and desert, and sort out once more the sense of possession being appealed to.

For this purpose, it may be helpful to consider a discussion of justice and personal desert by Joel Feinberg, who analyzes the bases of desert with an admirable clarity in terms suggestive for the arguments before us (1970). Feinberg begins with the observation that no one can deserve anything unless there is some basis for the desert. 'Desert without a basis is simply not desert'. But the question immediately arises what *kind* of basis is necessary. As Feinberg writes, 'Not any old basis will do'. Once again, the notion of possession provides the key. 'If a person is deserving of some sort of treatment, he must, necessarily, be so in virtue of some *possessed characteristic* or prior activity' [emphasis added] (1970: 48).

A characteristic of mine cannot be a basis for a desert of yours unless it somehow reveals or reflects some characteristic of yours. In general, the facts which constitute the basis of a subject's desert must be facts about that subject. If a student deserves a high grade in a course, for example, his desert must be in virtue of some fact about *him* – his earlier performances, say, or his present abilities. . . . It is necessary that a person's desert have a basis and that the basis consist in some fact about himself.

(1970: 58–9, 61)

Feinberg's analysis, tying a person's desert to some fact about the person, would appear to support Nozick's claim that 'the foundations underlying desert needn't themselves be deserved, *all the way down*'. In fact, the reliance of desert on some possessed characteristic of the person suggests a thesis even stronger than Nozick's: that the foundations underlying desert *cannot* themselves be deserved, *all the way down*, any more than the foundations underlying possession can themselves be possessed, *all the way down*. We have already seen how the notion of possession requires that somewhere, 'down there', there must be a subject of possession that is not *itself* possessed (for this would deny its agency), a subject 'doing the possessing', so to speak. The analogy for desert must be a *basis* of desert ultimately prior to desert. For consider: if desert presupposes some possessed characteristic, and if possessed characteristics presuppose some subject of possession which is not itself possessed, then desert must presuppose some subject of possession which is not itself possessed, and therefore some basis of desert which is not itself deserved. Just as there must be some subject of possession prior to possession, so there must be some basis of desert prior to desert. This is why the question whether someone deserves (to have) his sterling character, for example, is notoriously difficult (for it is unclear who or what is left to judge once his character has been removed), and why, beyond a certain point, asking just wholesale whether someone deserves to be the (kind of) person he is becomes incoherent altogether. Somewhere, 'down there', there must be a basis of desert that is not itself deserved. The foundations underlying desert cannot themselves be deserved, *all the way down*.

This result would seem amply to confirm Nozick's claim against Rawls that I do not necessarily have to *deserve* everything I use in producing a thing in order to deserve the thing, that some of what I use I 'just may *have*, not illegitimately'. And if this claim can be established, then it would appear that Rawls' argument from arbitrariness fails to undermine desert after all. To say, as Rawls does, that I do not deserve the superior character that led me to realize my abilities is no longer enough. To deny my desert, he must show that I do not *have* the requisite character, or alternatively, that I *have* it, but not in the requisite sense.

But this is precisely the argument Rawls' theory of the person allows him to make. For given his sharp distinction between the self, taken as the pure subject of possession,

and the aims and attributes it possesses, the self is left bare of any substantive feature or characteristic that could qualify as a desert base. Given the distancing aspect of possession, the self *itself* is dispossessed. On Rawls' theory of the person, the self, strictly speaking, *has nothing*, nothing at least in the strong, constitutive sense necessary to desert. In a move similar to the one invoked to show that the difference principle does not use a *person* as a means, only a person's *attributes*, Rawls can accept that some undeserved desert base is necessary to desert, only to claim that, on an adequate understanding of the person, this condition could never in principle be met! On Rawls' conception, the characteristics I possess do not *attach* to the self but are only *related* to the self, standing always at a certain distance. This is what makes them attributes rather than constituents of my person; they are *mine* rather than *me*, things I *have* rather than *am*.

We can see in this light how Rawls' argument from arbitrariness undermines desert not directly, by claiming I cannot *deserve* what is arbitrarily given, but indirectly, by showing I cannot *possess* what is arbitrarily given, that is, that 'I', qua subject of possession, cannot possess it in the undistanced, constitutive sense necessary to provide a desert base. An arbitrarily-given asset cannot be an essential constituent but only an accidental attribute of my person, for otherwise my identity would hang on a mere contingency, its continuity constantly vulnerable to transformation by experience, my status as a sovereign agent dependent on the conditions of my existence rather than epistemologically guaranteed. On Rawls' conception, no one can properly be said to deserve anything because no one can properly be said to possess anything, at least not in the strong, constitutive sense of possession necessary to the notion of desert.

A theory of justice without desert would seem a dramatic departure from traditional conceptions, but Rawls is at pains to show that it is not. In his opening pages, Rawls acknowledges that his approach 'may not seem to tally with tradition', but seeks to reassure that in fact it does.

The more specific sense that Aristotle gives to justice, and from which the most familiar formulations derive, is that of refraining from *pleonexia*, that is, from gaining some advantage for oneself by seizing *what belong to another*, his property, his reward, his office, and the like, or by denying a person that which is due to him. . . . *Aristotle's definition clearly presupposes, however, an account of what properly belongs to a person, and of what is due to him. Now such entitlements are, I believe, very often derived from social institutions and the legitimate expectations to which they give rise.* There is no reason to think that Aristotle would disagree with this, and certainly he has a conception of social justice to account for these claims. . . . There is no conflict with the traditional notion.

[emphasis added] (10–11)*

* All references to Rawls in this chapter not specifically cited are to Rawls 1971.

In comparing justice as fairness with traditional conceptions, Rawls confirms its novelty rather than denies it. What he presents as an incidental qualification to justice as classically conceived turns out on inspection to signal a striking departure. As Rawls suggests, traditional notions freely refer to ‘what properly belongs to a person’, institutions, presumably, aside; they presuppose thickly-constituted persons with a fixity of character, certain features of which are taken to be essential, ‘all the way down’. On Rawls’ conception, however, none of these concepts is available. In so far as a theory of justice ‘presupposes an account of what properly belongs to a person’ (in the strong sense of ‘belongs’), Rawls effectively acknowledges that he has none. Nor, he seems to imply, given the precedence of plurality, the priority of right, and the theory of the person they require, is it reasonable to think that such a theory of justice could be true. We are not essentially thick enough selves to bear rights and deserts antecedent to the institutions that define them. Given these constraints, the only alternative is to opt for a theory of justice based on entitlements to legitimate expectations, ruling out desert altogether. Rawls hedges this claim at first, saying only that ‘such entitlements are, I believe, *very often* derived from social institutions and the legitimate expectations to which they give rise’ [emphasis added] (10). But as the full consequences of Rawls’ view emerge, ‘very often’ becomes ‘always’, for it becomes clear that ‘such entitlements’ can arise in no other way. While Aristotle might not disagree that entitlements can arise in this way, it seems far from his view that they can arise in no other way. In denying that justice has to do with giving people what they deserve, justice as fairness departs decisively from the traditional notion after all.

Rawls’ apparent view that no one can properly be said to deserve anything, and the connection of this view with the notion of the self as ‘essentially unencumbered’, emerges more fully in his discussion of legitimate expectations and moral desert. He begins by acknowledging that justice as fairness, in rejecting desert, runs counter to common sense.

There is a tendency for common sense to suppose that income and wealth, and the good things in life generally, should be distributed according to moral desert. Justice is happiness according to virtue. While it is recognized that this ideal can never be fully carried out, it is the appropriate conception of distributive justice, at least as a *prima facie* principle, and society should try to realize it as circumstances permit. Now justice as fairness rejects this conception. Such a principle would not be chosen in the original position. There seems to be no way of defining the requisite criterion in that situation. (310–11)

There seems to be no way of defining the requisite criterion of a person’s virtue or moral worth in the original position because no substantive theory of the person

antecedent to social institutions exists. For moral desert to provide an independent criterion of justice, there must be some substantive theory of the person, or of the worth of persons, to get it going. But for Rawls, the worth of persons is subsequent to institutions, not independent of them. And so a person's moral claims must await their arrival.

This leads to the distinction between moral desert and legitimate expectations. Once a person does the various things established institutions encourage him to do, he acquires certain rights, but not before. He is entitled that institutions honor the claims they announce they will reward, but he is not entitled that they undertake to reward any particular kind of claim in the first place.

A just scheme, then, answers to what men are entitled to; it satisfies their legitimate expectations as founded upon social institutions. But what they are entitled to is not proportional to nor dependent upon their intrinsic worth. The principles of justice that regulate the basic structure and specify the duties and obligations of individuals do not mention moral desert, and there is no tendency for distributive shares to correspond to it. (311)

The principles of justice do not mention moral desert because, strictly speaking, no one can be said to deserve anything. Similarly, the reason people's entitlements are not proportional to nor dependent upon their intrinsic worth is that, on Rawls' view, *people have no intrinsic worth*, no worth that is intrinsic in the sense that it is theirs prior to or independent of or apart from what just institutions attribute to them.

The essential point is that the concept of moral worth does not provide a first principle of distributive justice. This is because it cannot be introduced until after the principles of justice and of natural duty and obligation have been acknowledged. . . . [T]he concept of moral worth is secondary to those of right and justice, and it plays no role in the substantive definition of distributive shares. (312–13)

Rawls could agree with Feinberg that 'desert is a *moral* concept in the sense that it is logically prior to and independent of public institutions and their rules', but would deny that there is any 'antecedent standard for its definition', and so disagree with Feinberg that 'one of the aims of [a system of public bestowals] is to give people what they deserve' (1970: 86). For Rawls, the principles of justice aim neither at rewarding virtue nor at giving people what they deserve, but instead at calling forth the resources and talents necessary to serve the common interest.

None of the precepts of justice aims at rewarding virtue. The premiums earned by scarce natural talents, for example, are to cover the costs of training and to

encourage the efforts of learning, as well as to direct ability to where it best furthers the common interest. The distributive shares that result do not correlate with moral worth. (311)

To illustrate the priority of just institutions with respect to virtue and moral worth, Rawls suggests an analogy to the relation between the rules of property and the law of robbery and theft.

These offenses and the demerits they entail presuppose the institution of property which is established for prior and independent social ends. For a society to organize itself with the aim of rewarding moral desert as a first principle would be like having the institution of property in order to punish thieves. The criterion to each according to his virtue would not, then, be chosen in the original position. (313)

The analogy is intriguing, but one wonders whether it works entirely to Rawls' advantage. While it is apparent that the institution of property has a *certain* priority with respect to its correlative offenses, it is less clear why the dependence must run only in one direction, especially given Rawls' own commitment to the method of reflective equilibrium. For example, is our belief in the validity of the institution of property in no way enhanced by a conviction that robbery and theft are wrong? Would our confidence in the institution of property in no way be diminished if it turned out that those it defined as robbers and thieves were invariably good and virtuous men? And what of more extreme cases? While the norms and rules protecting human life can no doubt be defended on a variety of grounds, such as keeping people alive, avoiding suffering, and so on, is it logically mistaken to think that one justification of prohibitions against murder could be to punish murderers?

Rawls' position here appears especially perplexing in the light of a contrast he draws between distributive justice and retributive justice, suggesting that in the second case, some notion of moral desert may be appropriate after all. The view that distributive shares should match moral worth to the extent possible, writes Rawls, 'may arise from thinking of distributive justice as somehow the opposite of retributive justice'. But the analogy is mistaken. In a reasonably well-ordered society, 'Those who are punished for violating just laws have normally done something wrong. This is because the purpose of the criminal law is to uphold basic natural duties . . . 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 prescribed 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.

It is clear that the distribution of economic and social advantages is entirely different. These arrangements are not the converse, so to speak, of the criminal law so that just as the one punishes certain offenses, the other rewards moral worth. The function of unequal distributive shares is to cover the costs of training and education, to attract individuals to places and associations where they are most needed from a social point of view, and so on. . . . *To think of distributive and retributive justice as converses of one another is completely misleading and suggests a moral basis of distributive shares where none exists.*

[emphasis added] (314–15)

Unlike the benefits that flow from distributive arrangements, the punishments and prohibitions associated with the criminal law are not simply a non-moral system of incentives and deterrents designed to encourage some forms of behavior and discourage others. For Rawls, the pre-institutional moral notions excluded in distributive justice somehow find meaning for retributive purposes, and there is a tendency for punishment to correspond to them.

The immediate puzzle is how this account can possibly fit with the analogy of property and theft. If retributive justice differs from distributive justice precisely in virtue of its prior moral basis, it is difficult to see how the example of property and theft could demonstrate the priority of social institutions with respect to virtue and moral worth, if this priority holds for distributive justice alone. This relatively minor confusion aside, the more basic question is how Rawls can admit desert in retributive justice without contradicting the theory of the self and related assumptions that ruled it out for purposes of distributive justice. If such notions as pre-institutional moral claims and intrinsic moral worth are excluded from a theory of distributive justice in virtue of an essentially unencumbered self too slender to support them, it is difficult to see how retributive justice could differ in any relevant way.¹

Do not the same arguments from arbitrariness exclude desert as a basis for punishment as for distributive shares? Is the propensity to commit crimes, any less than the propensity to do good, the result of factors arbitrary from a moral point of view? And if not, why would the parties to the original position not agree to share one another's fate for the purpose of criminal liability as well as distributive arrangements? Since under the veil of ignorance, none can know whether he shall have the misfortune to be born into the unfavorable social and family circumstances that lead to a life of crime, why would the parties not adopt a kind of difference principle for punishments as well as distributive shares, and agree, in effect, to regard the distribution of natural and social liabilities as a common burden?

Rawls holds that ‘those who are punished for violating just laws have normally done something wrong’, and so deserve their punishment (314). But suppose, by an act of vandalism, I deprive the community of a certain measure of well-being, say by throwing a brick through a window. Is there any reason why I deserve to bear the full costs of my destructiveness any more than the person who produced the window *deserves* to enjoy the full benefits of his productiveness? Rawls may reply that my ‘propensity to commit such acts is a mark of bad character’. But if the worker’s industriousness in making the window is not a mark of good character (in the moral, pre-institutional sense), why is my maliciousness in breaking the window a mark of bad character (in the moral pre-institutional sense)? To be sure (following Rawls, p. 103), given a just system of criminal law, those who have done what the system announces it will punish are properly dealt with accordingly and in his sense are ‘deserving’ of their penalty. ‘But this sense of desert presupposes the existence of the [retributive] scheme; it is irrelevant to the question whether in the first place the scheme is to be designed in accordance with the difference principle or some other criterion’ (103).

Some may think that the criminal deserves his punishment in the strong moral sense because he deserves the low character his criminality reflects. Perhaps this is what Rawls has in mind when he writes that ‘propensity to commit such acts is a mark of bad character’, and punishments properly fall on those who display these faults. Because the transgressor is less worthy in this sense, he deserves the misfortune that befalls him. But again (following Rawls, p. 104), this view is surely incorrect. It seems to be one of the fixed points of our considered judgments that no one deserves his place in the distribution of native endowments or liabilities, any more than one deserves one’s initial starting place in society. The assertion that a man deserves the inferior character that prevents him from overcoming his liabilities is equally problematic; for his character depends in large part upon unfortunate family and social circumstances for which he cannot be blamed. The notion of desert seems not to apply to these cases. None of which is to say that, generally speaking, a non-moral theory of distributive justice is incompatible with a moral, or desert-based theory of punishment, only that given Rawls’ reasons for rejecting desert-based distributive arrangements, he seems clearly committed to rejecting desert-based retributive ones as well.

The apparent inconsistency between Rawls’ retributive and distributive theories need not do serious damage to the theory as a whole. Given the method of reflective equilibrium, ‘justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view’ (21). From the standpoint of the overall theory, little hangs on Rawls’ retributive theory, apart from the measure of plausibility it lends justice as fairness for those committed to a strong, desert-based notion of punishment. If Rawls’ distinction succeeds, they need not choose between their retributive intuitions and the difference principle; if it does not, one or the other of those convictions must give way. If, on reflection, a non-moral theory of punishment

appears unacceptable, even in the light of the arbitrariness of criminal characteristics and dispositions, then the difference principle – rejecting as it does the notion of desert – would be called into serious question. If, on the other hand, our intuition that criminals deserve punishment proves no more indispensable than our intuition that virtue deserves reward (an intuition of common sense Rawls explicitly rejects), then we may adjust our intuitions in a direction that affirms the difference principle rather than opposes it. Desert would be rejected as the basis for both distributive and retributive arrangements, and so the inconsistency resolved.

But such a resolution returns us to the larger difficulties of a theory of justice without desert and a notion of the self as essentially dispossessed, or barren of constituent traits. Nozick argues against Rawls that the foundations underlying desert need not themselves be deserved, all the way down. But as we have seen, Rawls' denial of desert does not depend on the thesis Nozick refutes, but instead on the notion of the self as a pure, unadulterated, 'essentially unencumbered' subject of possession. Rawls is not committed to the view that a person can only deserve a thing he produces if he deserves everything he used in producing it, but rather to the view that no one possesses anything in the strong, constitutive sense necessary to a desert base. No one can be said to deserve anything (in the strong, pre-institutional sense), because no one can be said to possess anything (in the strong, constitutive sense). This is the philosophical force of the argument from arbitrariness.

That the argument from arbitrariness works in this way can be seen by viewing the moves from natural liberty to fair opportunity to the democratic conception, as traced by Rawls, as stages in the dispossession of the person. With each transition, a substantive self, thick with particular traits, is progressively shorn of characteristics once taken to be essential to its identity; as more of its features are seen to be arbitrarily given, they are relegated from presumed constituents to mere attributes of the self. More becomes *mine*, and less remains *me*, to recall our earlier formulation, until ultimately the self is purged of empirical constituents altogether, and transformed into a condition of agency standing beyond the objects of its possession. The logic of Rawls' argument might be reconstructed as follows:

At the far end of the spectrum, even before natural liberty appears, are aristocratic and caste societies; in such societies, a person's life prospects are tied to a hierarchy into which he is born and from which his person is inseparable. Here, the self is most fully ascribed, merged almost indistinguishably with its condition, embedded in its situation. The system of natural liberty removes fixed status of birth as an assumed constituent of the person, and regards each as free, given his capacities and resources, to compete in the marketplace as best he can, and to reap his reward. By shifting the basis of expectations from status to contract, the system of natural liberty repairs the arbitrariness of hierarchical societies by taking the person more narrowly, so to speak,

as distinct and separable from his surroundings. Still, some arbitrariness remains, most notably in the form of social and cultural contingencies. In the regime of natural liberty, a person's life prospects are governed by factors no more ascribable to the person (in the strong, constitutive sense) than his inherited status. Having relieved the person of his hierarchical baggage, the principle of natural liberty still conceives a thickly-constituted self, burdened by the accidents of social and cultural contingency. And so the move to fair opportunity, which strips the self of social and cultural accidents as well as inherited status. In a 'fair meritocracy', the effects of class status and cultural disadvantage are understood to reflect more on the society and less on the person. Those with comparable talents and 'the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system, that is, irrespective of the income class into which they are born' (73). In this way, the meritocratic conception extends the logic of natural liberty by ascribing less to the self and more to its situation.

But even the principle of fair opportunity, in rewarding individual effort, conceives the province of the self too expansively. For even 'the effort a person is willing to make is influenced by his natural abilities and skills and the alternatives open to him. The better endowed are more likely, other things equal, to strive conscientiously, and there seems to be no way to discount for their greater good fortune' (312). The self is still over-ascribed. Given its arbitrariness, even the character that determines a person's motivation cannot properly be regarded as an essential constituent of his identity. And so finally the move to the democratic conception, in which the self, shorn of all contingently-given attributes, assumes a kind of supra-empirical status, essentially unencumbered, bounded in advance and given prior to its ends, a pure subject of agency and possession, ultimately thin. Not only my character but even my values and deepest convictions are relegated to the contingent, as features of my condition rather than as constituents of my person. 'That we have one conception of the good rather than another is not relevant from a moral standpoint. In acquiring it we are influenced by the same sort of contingencies that lead us to rule out a knowledge of our sex and class' (Rawls 1975: 537). Only in this way is it possible to install the self as invulnerable, to assure its sovereignty once and for all in a world threatening always to engulf it. Only if the fate of the self is thus detached from the fate of its attributes and aims, subject as they are to the vagaries of circumstances, can its priority be preserved and its agency guaranteed.

This is the vision of the person that Nozick and Bell, as defenders of natural liberty and meritocracy, respectively, emphatically reject, even if they do not spell out in any detail the conception of the self they rely on instead. Both object that the argument from arbitrariness, consistently applied, leads ineluctably to the dissolution of the person, and the abnegation of individual responsibility and moral choice. 'This line of argument can succeed in blocking the introduction of a person's autonomous choices and activities (and their results) only by attributing *everything* noteworthy about the person

completely to certain sorts of “external” factors’, writes Nozick. Echoing his argument against the notion of common assets, Nozick questions whether, on Rawls’ account, any coherent conception of the person remains, and if so, whether it is any longer the kind of person worth the moral fuss deontological liberalism makes on its behalf.

So denigrating a person’s autonomy and prime responsibility for his actions is a risky line to take for a theory that otherwise wishes to buttress the dignity and self-respect of autonomous beings; especially for a theory that founds so much (including a theory of the good) upon a person’s choices. One doubts that the unexalted picture of human beings Rawls’ theory presupposes and rests upon can be made to fit together with the view of human dignity it is designed to lead to and embody.

(1974: 214)

Bell summarizes the objection in an epigram: ‘The person has disappeared. Only attributes remain’ (1973: 419). Where Rawls seeks to assure the autonomy of the self by disengaging it from the world, his critics say he ends by dissolving the self in order to preserve it.

To recapitulate our reconstructed version of the argument between Rawls and Nozick on the issue of desert: Nozick first argues that the arbitrariness of assets does not undermine desert, because desert may depend not only on things I deserve, but also on things I just *have*, not illegitimately. Rawls’ response is to invoke the distinction between the self and its possessions in the strongest version of that distinction, and so to claim that, strictly speaking, there *is* nothing that ‘I, qua pure subject of possession, *have* – nothing that is attached, rather than related, to *me* – nothing at least in the strong, constitutive sense of possession necessary to a desert base. Nozick’s rejoinder is that this defense cannot succeed for long, for it has the consequence of leaving us with a subject *so* shorn of empirically identifiable characteristics as to resemble once more the Kantian transcendent or disembodied subject Rawls resolved to avoid. It makes the individual inviolable only by making him invisible, and calls into question the dignity and autonomy this liberalism seeks above all to secure.

[. . .]

CONCLUSION

For justice to be the first virtue, certain things must be true of us. We must be creatures of a certain kind, related to human circumstance in a certain way. We must stand at a certain distance from our circumstance, whether as transcendental subject in the

case of Kant, or as essentially unencumbered subject of possession in the case of Rawls. Either way, we must regard ourselves as independent: independent from the interests and attachments we may have at any moment, never identified by our aims but always capable of standing back to survey and assess and possibly to revise them (Rawls 1979: 7; 1980: 544–5).

Deontology's liberating project

Bound up with the notion of an independent self is a vision of the moral universe this self must inhabit. Unlike classical Greek and medieval Christian conceptions, the universe of the deontological ethic is a place devoid of inherent meaning, a world 'disenchanted' in Max Weber's phrase, a world without an objective moral order. Only in a universe empty of *telos*, such as seventeenth-century science and philosophy affirmed,² is it possible to conceive a subject apart from and prior to its purposes and ends. Only a world ungoverned by a purposive order leaves principles of justice open to human construction and conceptions of the good to individual choice. In this the depth of opposition between deontological liberalism and teleological world views most fully appears.

Where neither nature nor cosmos supplies a meaningful order to be grasped or apprehended, it falls to human subjects to constitute meaning on their own. This would explain the prominence of contract theory from Hobbes onward, and the corresponding emphasis on voluntarist as against cognitive ethics culminating in Kant. What can no longer be found remains somehow to be created.³ Rawls describes his own view in this connection as a version of Kantian 'constructivism'.

The parties to the original position do not agree on what the moral facts are, as if there were already such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), *there is no such order*, and therefore no such facts apart from the procedure as a whole.

[emphasis added] (1980: 568)

Similarly for Kant, the moral law is not a discovery of theoretical reason but a deliverance of practical reason, the product of pure will. 'The elementary practical concepts have as their foundation the form of a pure will given in reason', and what makes this will authoritative is that it legislates in a world where meaning has yet to arrive. Practical reason finds its advantage over theoretical reason precisely in this voluntarist faculty, in its capacity to generate practical precepts directly, without recourse to cognition. 'Since in all precepts of the pure will it is only a question of the deter-

mination of will,' there is no need for these precepts 'to wait upon intuitions in order to acquire a meaning. This occurs for the noteworthy reason that *they themselves produce the reality of that to which they refer*' [emphasis added] (1788: 67–8).

It is important to recall that, on the deontological view, the notion of a self barren of essential aims and attachments does not imply that we are beings wholly without purpose or incapable of moral ties, but rather that the values and relations we have are the products of choice, the possessions of a self given prior to its ends. It is similar with deontology's universe. Though it rejects the possibility of an objective moral order, this liberalism does not hold that just anything goes. It affirms justice, not nihilism. The notion of a universe empty of intrinsic meaning does not, on the deontological view, imply a world wholly ungoverned by regulative principles, but rather a moral universe inhabited by subjects capable of constituting meaning on their own – as agents of *construction* in case of the right, as agents of *choice* in the case of the good. Qua noumenal selves, or parties to the original position, we arrive at principles of justice; qua actual, individual selves, we arrive at conceptions of the good. And the principles we construct as noumenal selves constrain (but do not determine) the purposes we choose as individual selves. This reflects the priority of the right over the good.

The deontological universe and the independent self that moves within it, taken together, hold out a liberating vision. Freed from the dictates of nature and the sanction of social roles, the deontological subject is installed as sovereign, cast as the author of the only moral meanings there are. As inhabitants of a world without *telos*, we are free to construct principles of justice unconstrained by an order of value antecedently given. Although the principles of justice are not strictly speaking a matter of choice, the society they define 'comes as close as a society can to being a voluntary scheme' (13), for they arise from a pure will or act of construction not answerable to a prior moral order. And as independent selves, we are free to choose our purposes and ends unconstrained by such an order, or by custom or tradition or inherited status. So long as they are not unjust, our conceptions of the good carry weight, whatever they are, simply in virtue of our having chosen them. We are 'self-originating sources of valid claims' (Rawls 1980: 543).

Now justice is the virtue that embodies deontology's liberating vision and allows it to unfold. It embodies this vision by describing those principles the sovereign subject is said to construct while situated prior to the constitution of all value. It allows the vision to unfold in that, equipped with these principles, the just society regulates each person's choice of ends in a way compatible with a similar liberty for all. Citizens governed by justice are thus enabled to realize deontology's liberating project – to exercise their capacity as 'self-originating sources of valid claims' – as fully as circumstances permit. So the primacy of justice at once expresses and advances the liberating aspirations of the deontological world view and conception of the self.

But the deontological vision is flawed, both within its own terms and more generally as an account of our moral experience. Within its own terms, the deontological self, stripped of all possible constitutive attachments, is less liberated than disempowered. As we have seen, neither the right nor the good admits of the voluntarist derivation deontology requires. As agents of construction we do not really construct, and as agents of choice we do not really choose. What goes on behind the veil of ignorance is not a contract or an agreement but if anything a kind of discovery; and what goes on in ‘purely preferential choice’ is less a choosing of ends than a matching of pre-existing desires, undifferentiated as to worth, with the best available means of satisfying them. For the parties to the original position, as for the parties to ordinary deliberative rationality, the liberating moment fades before it arrives; the sovereign subject is left at sea in the circumstances it was thought to command.

The moral frailty of the deontological self also appears at the level of first-order principles. Here we found that the independent self, being essentially dispossessed, was too thin to be capable of desert in the ordinary sense. For claims of desert presuppose thickly-constituted selves, beings capable of possession in the constitutive sense, but the deontological self is wholly without possessions of this kind. Acknowledging this lack, Rawls would found entitlements on legitimate expectations instead. If we are incapable of desert, at least we are entitled that institutions honor the expectations to which they give rise.

But the difference principle requires more. It begins with the thought, congenial to the deontological view, that the assets I have are only accidentally mine. But it ends by assuming that these assets are therefore common assets and that society has a prior claim on the fruits of their exercise. This either disempowers the deontological self or denies its independence. Either my prospects are left at the mercy of institutions established for ‘prior and independent social ends’ (313), ends which may or may not coincide with my own, or I must count myself a member of a community defined in part by those ends, in which case I cease to be unencumbered by constitutive attachments. Either way, the difference principle contradicts the liberating aspiration of the deontological project. We cannot be persons for whom justice is primary and also be persons for whom the difference principle is a principle of justice.

Character, self-knowledge, and friendship

If the deontological ethic fails to redeem its own liberating promise, it also fails plausibly to account for certain indispensable aspects of our moral experience. For deontology insists that we view ourselves as independent selves, independent in the sense that our identity is never tied to our aims and attachments. Given our ‘moral power to form, to revise, and rationally to pursue a conception of the good’ (Rawls

1980: 544), the continuity of our identity is unproblematically assured. No transformation of my aims and attachments could call into question the person I am, for no such allegiances, however deeply held, could possibly engage my identity to begin with.

But we cannot regard ourselves as independent in this way without great cost to those loyalties and convictions whose moral force consists partly in the fact that living by them is inseparable from understanding ourselves as the particular persons we are – as members of this family or community or nation or people, as bearers of this history, as sons and daughters of that revolution, as citizens of this republic. Allegiances such as these are more than values I happen to have or aims I ‘espouse at any given time’. They go beyond the obligations I voluntarily incur and the ‘natural duties’ I owe to human beings as such. They allow that to some I owe more than justice requires or even permits, not by reason of agreements I have made but instead in virtue of those more or less enduring attachments and commitments which taken together partly define the person I am.

To imagine a person incapable of constitutive attachments such as these is not to conceive an ideally free and rational agent, but to imagine a person wholly without character, without moral depth. For to have character is to know that I move in a history I neither summon nor command, which carries consequences none the less for my choices and conduct. It draws me closer to some and more distant from others; it makes some aims more appropriate, others less so. As a self-interpreting being, I am able to reflect on my history and in this sense to distance myself from it, but the distance is always precarious and provisional, the point of reflection never finally secured outside the history itself. A person with character thus knows that he is implicated in various ways even as he reflects, and feels the moral weight of what he knows.

This makes a difference for agency and self-knowledge. For, as we have seen, the deontological self, being wholly without character, is incapable of self-knowledge in any morally serious sense. Where the self is unencumbered and essentially dispossessed, no person is left for *self*-reflection to reflect upon. This is why, on the deontological view, deliberation about ends can only be an exercise in arbitrariness. In the absence of constitutive attachments, deliberation issues in ‘purely preferential choice’, which means the ends we seek, being mired in contingency, ‘are not relevant from a moral standpoint’ (Rawls 1975: 537).

When I act out of more or less enduring qualities of character, by contrast, my choice of ends is not arbitrary in the same way. In consulting my preferences, I have not only to weigh their intensity but also to assess their suitability to the person I (already) am. I ask, as I deliberate, not only what I really want but who I really am, and this last question takes me beyond an attention to my desires alone to reflect on my identity itself. While the contours of my identity will in some ways be open and subject to revision, they are not wholly without shape. And the fact that they are not enables me to discriminate among my more immediate wants and desires; some now

appear essential, others merely incidental to my defining projects and commitments. Although there may be a certain ultimate contingency in my having wound up the person I am – only theology can say for sure – it makes a moral difference none the less that, being the person I am, I affirm these ends rather than those, turn this way rather than that. While the notion of constitutive attachments may at first seem an obstacle to agency – the self, now encumbered, is no longer strictly prior – some relative fixity of character appears essential to prevent the lapse into arbitrariness which the deontological self is unable to avoid.

The possibility of character in the constitutive sense is also indispensable to a certain kind of friendship, a friendship marked by mutual insight as well as sentiment. By any account, friendship is bound up with certain feelings. We like our friends; we have affection for them, and wish them well. We hope that their desires find satisfaction, that their plans meet with success, and we commit ourselves in various ways to advancing their ends.

But for persons presumed incapable of constitutive attachments, acts of friendship such as these face a powerful constraint. However much I might hope for the good of a friend and stand ready to advance it, only the friend himself can know what that good is. This restricted access to the good of others follows from the limited scope for self-reflection, which betrays in turn the thinness of the deontological self to begin with. Where deliberating about my good means no more than attending to wants and desires given directly to my awareness, I must do it on my own; it neither requires nor admits the participation of others. Every act of friendship thus becomes parasitic on a good identifiable in advance. ‘Benevolence and love are second-order notions: they seek to further the good of beloved individuals that is already given’ (191). Even the friendliest sentiments must await a moment of introspection itself inaccessible to friendship. To expect more of any friend, or to offer more, can only be a presumption against the ultimate privacy of self-knowledge.

For persons encumbered in part by a history they share with others, by contrast, knowing oneself is a more complicated thing. It is also a less strictly private thing. Where seeking my good is bound up with exploring my identity and interpreting my life history, the knowledge I seek is less transparent to me and less opaque to others. Friendship becomes a way of knowing as well as liking. Uncertain which path to take, I consult a friend who knows me well, and together we deliberate, offering and assessing by turns competing descriptions of the person I am, and of the alternatives I face as they bear on my identity. To take seriously such deliberation is to allow that my friend may grasp something I have missed, may offer a more adequate account of the way my identity is engaged in the alternatives before me. To adopt this new description is to see myself in a new way; my old self-image now seems partial or occluded, and I may say in retrospect that my friend knew me better than I knew myself. To deliberate with friends is to admit this possibility, which presupposes in

turn a more richly-constituted self than deontology allows. While there will of course remain times when friendship requires deference to the self-image of a friend, however flawed, this too requires insight; here the need to defer implies the ability to know.

So to see ourselves as deontology would see us is to deprive us of those qualities of character, reflectiveness and friendship that depend on the possibility of constitutive projects and attachments. And to see ourselves as given to commitments such as these is to admit a deeper commonality than benevolence describes, a commonality of shared self-understanding as well as 'enlarged affections'. As the independent self finds its limits in those aims and attachments from which it cannot stand apart, so justice finds its limits in those forms of community that engage the identity as well as the interests of the participants.

To all of this, deontology might finally reply, with a concession and a distinction: it is one thing to allow that 'citizens in their personal affairs . . . have attachments and loves that they believe they would not, or could not, stand apart from', that they 'regard it as unthinkable . . . to view themselves without certain religious and philosophical convictions and commitments' (Rawls 1980: 545). But with public life it is different. There, no loyalty or allegiance could be similarly essential to our sense of who we are. Unlike our ties to family and friends, no devotion to city or nation, to party or cause, could possibly run deep enough to be defining. By contrast with our private identity, our 'public identity' as moral persons 'is not affected by changes over time' in our conceptions of the good (Rawls 1980: 544–5). While we may be thickly-constituted selves in private, we must be wholly unencumbered selves in public, and it is there that the primacy of justice prevails.

But once we recall the special status of the deontological claim, it is unclear what the grounds for this distinction could be. It might seem at first glance a psychological distinction; detachment comes more easily in public life, where the ties we have are typically less compelling; I can more easily step back from, say, my partisan allegiances than certain personal loyalties and affections. But as we have seen from the start, deontology's claim for the independence of the self must be more than a claim of psychology or sociology. Otherwise, the primacy of justice would hang on the degree of benevolence and fellow-feeling any particular society managed to inspire. The independence of the self does not mean that I can, as a psychological matter, summon in this or that circumstance the detachment required to stand outside my values and ends, rather that I must regard myself as the bearer of a self distinct from my values and ends, whatever they may be. It is above all an epistemological claim, and has little to do with the relative intensity of feeling associated with public or private relations.

Understood as an epistemological claim, however, the deontological conception of the self cannot admit the distinction required. Allowing constitutive possibilities where 'private' ends are at stake would seem unavoidably to allow at least the possibility that 'public' ends could be constitutive as well. Once the bounds of the self are no longer

fixed, individuated in advance and given prior to experience, there is no saying in principle what sorts of experiences could shape or reshape them, no guarantee that only 'private' and never 'public' events could conceivably be decisive.

Not egoists but strangers, sometimes benevolent, make for citizens of the deontological republic; justice finds its occasion because we cannot know each other, or our ends, well enough to govern by the common good alone. This condition is not likely to fade altogether, and so long as it does not, justice will be necessary. But neither is it guaranteed always to predominate, and in so far as it does not, community will be possible, and an unsettling presence for justice.

Liberalism teaches respect for the distance of self and ends, and when this distance is lost, we are submerged in a circumstance that ceases to be ours. But by seeking to secure this distance too completely, liberalism undermines its own insight. By putting the self beyond the reach of politics, it makes human agency an article of faith rather than an object of continuing attention and concern, a premise of politics rather than its precarious achievement. This misses the pathos of politics and also its most inspiring possibilities. It overlooks the danger that when politics goes badly, not only disappointments but also dislocations are likely to result. And it forgets the possibility that when politics goes well, we can know a good in common that we cannot know alone.

NOTES

- 1 In a footnote (315), Rawls cites Feinberg in apparent support of this claim, but Feinberg allows a role for desert in both distributive and retributive justice. Feinberg's point is that retributive justice involves what he calls polar desert (where one either deserves good or deserves ill), whereas distributive justice involves nonpolar desert (where, as a prize, some deserve and others do not). But both cases involve desert in the moral, pre-institutional sense (Feinberg 1970: 62).
- 2 For discussion of the moral, political, and epistemological consequences of the seventeenth-century scientific revolution and world-view, see Strauss 1953; Arendt 1958: 248–325; Wolin 1960: 239–85; and Taylor 1975: 3–50.
- 3 As one liberal writer boldly asserts, 'The hard truth is this: There is no moral meaning hidden in the bowels of the universe. . . . Yet there is no need to be overwhelmed by the void. We may create our own meanings, you and I' (Ackerman 1980: 368). Oddly enough, he insists nonetheless that liberalism is committed to no particular metaphysic or epistemology, nor any 'Big Questions of a highly controversial character' (356–7, 361).

REFERENCES

- Ackerman, B.A. (1980). *Social Justice in the Liberal State*. Yale University Press, New Haven.
 Arendt, H. (1958). *The Human Condition*. University of Chicago Press, Chicago.