

Egalitarianism, Responsibility and Desert*

Christopher Woodard

P. Grim, K. Baynes, P. Ludlow, and G. Mar (eds.)
Philosopher's Annual, Volume XXI (Atascadero: Ridgeview, 2000).

1. Introduction

There is currently a debate amongst egalitarians about the role of the concept of desert and, especially, the concept of responsibility, in explaining what justice requires. Interest in these two concepts has been stimulated by Rawls's remarks on the 'moral arbitrariness' of natural and social advantages (such as talents), and has been sustained and extended in the debate about what egalitarians should seek to equalise.¹

The current debate seems to be about two issues. The first issue is the plausibility of the claim made by Cohen, Arneson, and Roemer, that egalitarianism requires equality in circumstances for which people are not responsible.² This is a claim about the underlying rationale or driving thought of egalitarianism, not about the details of an egalitarian theory. There is a good deal of room for disagreement amongst those who accept it, about how exactly to formulate the aim, and, of course, about what exactly people are responsible for. But it is still a substantial issue whether egalitarianism is properly characterised as a style of thinking about justice which places responsibility in this central role.³

The second and related issue concerns the polemical merits of emphasising the role of responsibility in egalitarian arguments about justice, supposing that it does have a central role. Samuel Scheffler has drawn attention to a striking contrast between Rawlsian attempts to specify what justice requires without relying on the concept of responsibility, and the prominence of issues of responsibility in general public debate about justice. Liberal political philosophers, he suggests, risk detaching themselves from the terms of public debate, if they do not cast their arguments in terms of responsibility.⁴ On the other hand, there are good polemical reasons for not wanting to rely on the concept of responsibility, for to rely on it is to make arguments about justice vulnerable to all of the philosophical problems, and other kinds of persistent controversy, which surround the use of that concept.

Taken together, these recent arguments about the role of responsibility in egalitarian views about justice seem to constitute an important strategic debate about how to develop egalitarianism. One can view egalitarian thought from Rawls onwards as gradually uncovering the importance of the concept of

responsibility, and at present as considering whether to define egalitarianism in terms of the central connection it finds between responsibility and justice.

This article proceeds with this picture in mind, and it is intended to contribute to the discussion about the role of responsibility. I shall argue that it is crucial to distinguish two quite different types of view, both of which portray responsibility and justice as closely connected concepts. Unless we make this distinction, we are bound to overestimate the extent of agreement amongst egalitarians about the role of responsibility. And, if we make it, the salient issue for discussion seems to be not whether justice and responsibility are closely connected concepts, but whether we should seek to explain what justice requires in terms of (what I shall describe as) *agency-implicating concepts*.

If this analysis is correct, we should expect similar issues to arise concerning the relationship between justice and *desert*.⁵ Responsibility and desert are dissimilar in important respects, but they are both, in their usual interpretations at least, agency-implicating concepts, and they both admit of unusual, non-agency-implicating interpretations. The central issue for egalitarians, so I claim, is not whether their accounts of justice should make use of the concept of responsibility, but whether they should make use of *agency-implicating* conceptions of responsibility or desert, or instead seek to avoid doing so.

The first half of the article (sections 2-4) explains the proposal to relocate the central issue for egalitarians. The second half (sections 5-7) explores in detail Rawls's remarks on desert, as a case study of the strategy of trying to avoid relying on agency-implicating concepts.

2. The Connections Between Justice and Responsibility

First let us review the reasons why, it has been thought, egalitarianism is appropriately characterised by saying that it gives a certain central role to the concept of responsibility. This will provide the background for the argument, in later sections, that this characterisation is misleading in an important respect.

One source of the view in question is a certain interpretation of Rawls's remarks on desert, which I shall discuss in more detail later. For the moment the point is that Rawls's claim that the basic structure of society should not be designed to reward natural and social endowments, since these are 'arbitrary from a moral point of view', can be massaged or extrapolated into the view that the basic structure should be designed to reward whatever it is that people are genuinely responsible for.⁶ Whilst Rawls officially disavows appeals to judgements about what people deserve or are responsible for, one can regard his conception of justice as embodying a prototype of the view that what justice

requires depends on what people are responsible for – or, perhaps, as trading rhetorically on the appeal of such a view.⁷

Another major source of the view that egalitarianism is well characterised by the central role it gives to the concept of responsibility, is the debate about what egalitarians should try to equalise (and in particular the discussions of expensive tastes).⁸ This debate addressed two separate questions, and, although they were not always clearly distinguished, the concept of responsibility featured in egalitarian answers to both.

The first question is this: what is the appropriate conception of advantage, for egalitarians? Most theories of justice, including most egalitarian theories, rely on some conception of advantage in order to characterise persons' circumstances. Such a conception is needed, for example, to say which of two people is *worse off*, in the sense relevant to justice. Two candidate conceptions of advantage are *welfare* and *resources*. Which of these, or which other measure, should an egalitarian use?

This issue may depend on considerations of neutrality; for example, we may choose *resources* as a measure of advantage because we do not want to rely on disputed assumptions about welfare in characterising advantage. Or it may turn on pragmatic considerations, about the availability of information.⁹ But it may also turn on considerations of responsibility. For example, should we treat someone with cheap tastes as advantaged, in the sense relevant to justice? Some people have thought that the answer depends on whether the person in question is responsible for those tastes: if she is, then her advantage is irrelevant to justice; if not, then it is relevant, and may be a candidate for compensation.

The second question addressed in the debate is this: of what does justice require equal amounts? Unlike the first question, this one is definitely egalitarian, since it assumes that justice requires equal amounts of *something*, and asks what that thing is. Answering it requires more than a conception of justice-relevant advantage – it requires also a view about which advantages or disadvantages should be compensated, in which circumstances. Some egalitarians have proposed the following answer: all and only those justice-relevant advantages for which persons are not responsible should be compensated. On this view, we may be unequal in things for which we are responsible, but we should be equal in those things for which we are not responsible.

This second idea has seemed to be a plausible formulation of the ideal that persons should be given a fair chance in life. Insofar as egalitarianism is motivated by that ideal, it is thought, it should operate with a distinction between the aspects of persons' circumstances for which they are, and those for which they are not, responsible. According to such a view, if the reason that I'm less

well-off than you is that I squandered my resources, while you prudently invested yours, and each of us is responsible for these different choices, justice does not require that what we have now is equalised. If, on the other hand, I'm less well off than you because of the class or country into which I was born, that inequality should be removed, because it is not something for which either of us is responsible. Cohen, Arneson, and Roemer have all offered versions of this idea.¹⁰

These two views – one about the appropriate sense of advantage, the other about what justice requires equal amounts of – are not equivalent. They do not cast responsibility in the same role. The first treats responsibility as a disqualifying condition from something's being an advantage (relevant to justice), and the second treats responsibility as a disqualifying condition for compensation, with respect to some conception of advantage. But both views place responsibility in a central role with respect to thoughts about justice.

A third source of the idea that egalitarianism is well characterised by the central role it gives to responsibility is the recent discussion of issues of incentives. Cohen has focused attention on the ways in which our attitudes to incentives may be shaped by underlying judgements about individual producers' responsibility for their behaviour.¹¹ Put very crudely, his idea is that it is much easier to justify incentive-providing inequalities in cases where individual producers would be *unable* to work just as hard in their absence, than in cases where they would be merely *unwilling* to work just as hard.¹² In the former case, the individual producers are not responsible for the fact that they require incentives to work at a given rate; perhaps it is just a fact of psychology that persons cannot work that hard, cannot summon the motivation, in the absence of incentives. But in the latter case, which Cohen thinks much more common, the fact that they would not work so hard without incentives *is* the producers' responsibility. And that affects the justification of incentives.

Cohen's target is a certain argument for inequality employing Rawls's difference principle. But if he is right about the significance of responsibility to the justification of incentives, his general point extends to all egalitarian theories which give weight to efficiency gains. For in all such cases, concern with efficiency is liable to conflict with concern with equality, especially in cases where, it is alleged, inequalities may cause efficiency gains. The difference principle simply provides one way of trying to balance equality against efficiency: the reason the issue arises does not have to do with the particular balance between equality and efficiency which is struck by that principle. So a third distinct way in which egalitarian views about justice may rely on the concept of responsibility, has to do with issues of incentives and the appropriate trade-off between efficiency and equality.

Reliance on the concept of responsibility has been found or perceived in several areas of egalitarian thought: in the conception of advantage appropriate to thinking about justice; in judgements about which inequalities of advantage should be removed; and in judgements about the appropriate trade-off to be made between considerations of efficiency and considerations of strict equality.

There are, then, good theoretical reasons for thinking that what characterises egalitarianism, in part at least, is the use it makes of the concept of responsibility – not just the bare fact that it relies on that concept, perhaps, but the particular ways in which it does so. One could agree with Roemer, when he says that,

The issue of responsibility was . . . germinal in the writings of Rawls and Sen, and became focal in the writings of Dworkin, Scanlon, Arneson, and Cohen . . . this new articulation of responsibility in the theory of distributive justice, and of egalitarian theory in particular, is the signal achievement in the field of the last fifteen years . . .¹³

And we must add to this comment on academic discussions of equality and responsibility, the fact that talk of responsibility seems to resonate with people's everyday intuitions about justice – which is one half of the story about the polemical advantages of emphasising its role.

For these various reasons, the current debate about whether or not egalitarianism should be characterised in terms of the role it gives to responsibility is appropriate and helpful. However, there is one respect in which looking at things this way is misleading, as I shall now try to explain.

3. The Underlying Debate

It is in one way misleading to think of egalitarians as divided between those who do, and those who do not, accord responsibility a central role in their conceptions of justice. The reason is not that the division is false; some who are arguably egalitarian, including some utilitarians, *do* think that there is no conceptual connection between justice and responsibility, whilst others, as we've seen, think there is such a connection. The problem, instead, is that the division as described is less important than another, from which it may distract our attention.

To see this, consider the following claim:

P: In just societies, people get what they are responsible for.

P seems to be just the kind of claim that we might use to characterise the difference between those egalitarian views which do, and those which do not, place responsibility in a central role with respect to justice. But *P* is crucially ambiguous. It may mean, first, that justice requires distribution of goods in accordance with persons' responsibility. Call this the *standard* interpretation. On such a reading, we are encouraged to suppose that we have an independent grasp of what persons are responsible for, which may be used to illuminate what justice requires. In order to apply *P*, we must know what persons are responsible for, which probably requires knowledge of the agency of persons with respect to different events. On the basis of such knowledge, according to the standard interpretation of *P*, we may calculate what justice requires.

According to what I'll call the *naturalist* interpretation, however, *P* claims something different. It claims that someone's being responsible for *X* simply consists of it being just for her to bear the costs or enjoy the benefits of *X*. On this reading *P* is a tautology, since responsibility has no content independent of justice – indeed, it is defined in terms of justice. If we assert *P* in this sense, we cannot expect to shed light on what justice requires through our understanding of responsibility, because our understanding of responsibility is parasitic on our understanding of what justice requires. In order to apply *P* interpreted this way, we do not seem to need knowledge of responsibility, or of persons' agency with respect to different events. Instead, we need to understand what justice requires, on the basis of which we may calculate what persons are responsible for.

'Responsibility' is often used in this second sense. When we say that someone has strict liability for something, for example, we imply only that, according to some institutional scheme, she is the person who should bear the costs of it; hence, dispute about her strict liability either will be a factual dispute, or it will take the form of dispute about the justice of that scheme.¹⁴ And sometimes the same conceptual relationship between justice and responsibility is supposed in less specialised attributions of responsibility: if I say that expensive tastes are their bearers' responsibility, I may mean simply that it is just for them to bear the costs of those tastes.¹⁵ Similarly, in institutional contexts, to say that someone is responsible for something often means simply that the institution ascribes the costs or benefits of the thing in question to that person: it needn't imply anything about their agency with respect to that thing.

The ambiguity in *P* is a result of the fact that the concept of responsibility is open to quite different interpretations. Standardly it is interpreted as an *agency-implicating concept*; as a concept, that is, the correct application of which relies upon our understanding of persons as agents.¹⁶ Standardly, we think that whether or not 'Jones is responsible for *X*' is true, depends on facts about Jones's agency or lack of it. But responsibility may be interpreted

naturalistically, where that means that the standard reliance on our understanding of agency is avoided, or minimised. For naturalists about responsibility who are not also sceptical about use of that concept, some other kind of consideration must explain the difference between correct and incorrect application. One candidate is the concept of justice, which, as we've seen, is sometimes thought to explain responsibility.

For any given claim about responsibility and justice, it can be difficult to work out which concept is understood as having explanatory priority. For example, two people may agree that it is just for the bearers of expensive tastes to bear the costs of those tastes, but each may hold this view for quite different reasons. One may think that bearers of expensive tastes typically *choose* to have those tastes, say, whilst the other believes that justice requires, on grounds independent of persons' agency, that the costs of expensive tastes fall on those who have them. Often, the real test is not the verbal form of the claims made; it is the relationship which is asserted between claims of the form, 'A is responsible for *X*', and claims of the form, 'it is just for A to bear the costs/enjoy the benefits of *X*'. The first type of view treats the former kind of claim as (potentially) explanatory of the latter type of claim, whilst the second type of view treats them as synonymous.

When the concept of responsibility is used as a term derivative of the concept of justice, it may not be agency-implicating. We can see this clearly in the case of strict liability, where the claim that Jones is strictly liable for the barn's burning down implies nothing about Jones's involvement in that event, or anything else, apparently, about Jones's capacities or actions as an agent. In general, if we understand the concept of responsibility as derivative of justice, we may escape problems of agency in its use. But by the same token, we forgo any opportunity of explaining what justice requires in terms of persons' responsibility.

These considerations suggest that we need a three-way classification of egalitarian views about justice and responsibility. First, there are those who see no conceptual connection between justice and responsibility. Second, there are those who see only tautological connections of the kind just described. And thirdly, there are those who see a genuine explanatory connection between justice and responsibility (which implies that responsibility is conceptually prior to justice).¹⁷

The problem with the proposed division between those egalitarian views which connect justice to responsibility, and those which do not, then, is that it tends to obscure the difference between tautological and genuine explanatory connections between these concepts. The difference between these two types of connection is very important, both for the details of a theory and for our broader

impression of its type. We do not understand a theory which asserts a connection between justice and responsibility, unless we know whether the connection is supposed to be explanatory of justice, or simply definitive of responsibility.

Indeed, for many purposes, those theories which find only tautological connections between justice and responsibility may be most usefully classified with those which find no connection at all. The reason for this is that neither of these types of theory hopes to shed light on the nature of justice by reference to responsibility; only the third type of theory tries to do this. For theorists of justice, the important issue is the viability of explaining what justice requires in terms of responsibility and similar agency-implicating concepts.

4. The Attractions of Naturalism

Why try to explain justice in terms of responsibility? Common-sense morality seems to suppose that judgements of responsibility are relevant to justice: it supposes, for example, that one thing which is relevant to the justice of someone's having great wealth, is whether or not she earned it. But we needn't simply follow common-sense; we can explain how responsibility could illuminate justice. A plausible view of justice is that it consists of something like *fair treatment of persons*, and it's plausible too to think that, in order to treat persons fairly, we need to respond to them as *agents*, not simply as parts of the natural world. If so, we may expect to illuminate aspects of justice by investigating responsibility, since responsibility is one of our most central concepts used for describing persons as agents.¹⁸

Responsibility is an agency-implicating concept: in its standard interpretation, at least, its use implicates our understanding of persons as agents. That fact explains why it is plausible to seek illumination of justice from an understanding of responsibility. Justice involves the fair treatment of persons, which involves treating persons as agents, and the concept of responsibility is central to our understanding of persons as agents. But the fact that responsibility standardly is agency-implicating also explains why many political theorists are wary of the concept, and would prefer, if possible, to construct theories of justice in which it plays no serious explanatory role. For reliance in explanation on agency-implicating concepts brings with it vulnerability to the deep philosophical problems to which our understanding of agency is subject.¹⁹ As a result, there seem to be theoretical benefits in avoiding such reliance.

The same points apply to the concept of desert. In its usual interpretation, *desert* is an agency-implicating concept. But we can interpret desert differently, such that to deserve something is merely to be justly entitled to it.²⁰ Hence we can give a non-agency-implicating interpretation of desert, according to which

it is defined, partly, in terms of justice. Again, we avoid problems of agency but forsake the possibility of explaining what justice requires in terms of what people deserve.

We can call any theory which seeks to avoid relying on agency-implicating concepts, such as responsibility or desert, or to minimise such reliance, *naturalist*.²¹ Naturalism in this sense is characteristic of the natural sciences, which typically eschew explanations in terms of agency-implicating concepts. And it characterises also some approaches to theorising about justice. Theories which find no connection between justice and responsibility or desert are naturalist in those respects. But so too are theories which find only tautological connections of the kind just described – since the normally agency-implicating concepts of desert and responsibility appear only in the revised, non-agency-implicating form.

Naturalism is attractive because it aims to leave theories invulnerable to worries about agency. It is attractive in any domain, including theorising about justice, for the same kind of reason. But those attractions are not arguments for naturalism. They give us, at most, reasons for hoping that the underlying idea is true, not reasons for believing it to be so. It may just turn out, for example, that justice depends on responsibility. If so, the attractions of naturalism remain, and it's just too bad for our understanding of justice that it inherits the problems associated with responsibility, as an agency-implicating concept.

We can explain why it's plausible to think that justice depends on responsibility, and we can see the advantages which would follow if it did not. But how can we decide whether it does or not? Is justice an essentially agency-implicating concept? I've argued that this is the really important issue for theorists of justice, but it looks as if it is too abstract for us to do more than invoke theoretical prejudices in either direction. We are likely to have such prejudices. Some people think it's just *obvious* that justice is an essentially agency-implicating concept, and hence that any attempt to explain what justice requires without relying on our understanding of persons as agents, is misguided. But there is little in the way of an argument for this view, beyond the inconclusive one I gave at the start of this section. On the other side are the pragmatic considerations, which carry more weight with some people than with others.

There is an impasse here which exacerbates the obscurity of the issue at hand. The question of the link between egalitarian views about justice, and responsibility or desert, is undercut by the ambiguity in those two concepts, between standard and naturalist interpretations. But recognition of that as the important issue is hindered by the fact that the two leading trains of thought about naturalism don't engage with each other. Those who are impressed with

the possibility of naturalism are impressed, chiefly, with pragmatic considerations, whilst those who aren't impressed think it's *obviously* not a genuine possibility. As yet, there is little genuine debate.

In sections 5-7 below, I shall examine the role of judgements of desert and responsibility in Rawls's arguments about justice. There are two aims in doing this. One is to use Rawls's remarks to illustrate some of the points I've made in a general way about the relationship between justice, desert, and responsibility. Rawls's remarks about desert in particular are open to several interpretations, some of them arguably naturalist. The second aim is to search for evidence on the issue about which anti-naturalists tend to be confident, about the feasibility of naturalism. Rawls's theory is well-developed, and arguably naturalist. Hence, if we search for the most sympathetic, most successfully naturalist interpretation of his theory, and find that it is weak or incoherent, we may have some evidence that naturalism itself is a non-starter. On the other hand, of course, we may find the opposite.

5. Rawls's Remarks on Desert

Rawls makes two broad claims about desert, which seem on the face of it to be in tension with each other.²² His first claim (R1) is that reflection on desert favours his conception of justice over some of its rivals. This claim is made in the context of a discussion about the possible interpretations of his second principle of justice. In its initial formulation, the second principle claims that inequalities must be attached to positions equally open to all, and must be to everyone's advantage, in order to be just.²³ Rawls discusses two possible interpretations of each of these conditions, which can be combined in a total of four ways. He then claims that reflecting on desert favours one of these four interpretations over the others.

These claims seem to be an important part of Rawls's argument.²⁴ They are important, first, because they provide the official explanation of why Rawls favours his interpretation of the second principle over more meritocratic alternatives. As he recognises, this interpretative exercise fixes a good deal of the content of the second principle: hence, the rationale of the interpretation is an important part of the official account of how we move from thinking about justice in terms of the original position, to determinate conclusions about what justice requires. But the remarks are important in a second way too. They provide the materials for Rawlsians to respond to the polemically important charge that a Rawlsian society would be unjust because persons' shares would not reflect their differential desert.

Bearing this in mind, Rawls's second claim about desert is difficult to reconcile with the first. He claims (R2) that justice does not depend on desert:

A just scheme . . . 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 . . . do not mention moral desert, and there is no tendency for distributive shares to correspond to it.²⁵

Rawls here distinguishes between two different senses in which we may say that someone deserves something. The first sense (that of 'legitimate expectations') is the sense in which someone is entitled to something if they behave in a certain way against a certain institutional background. If I play a lottery and pick the winning ticket, that behaviour, against that institutional background, generates an entitlement to the winnings. If the institutions of the lottery are tolerably just, it would be *unjust* for me not to get the winnings. I have a legitimate expectation of them – and, in one sense of the word, I deserve them.²⁶

Note that this sense of desert makes desert conceptually dependent on justice, however. To be deserving, in the sense of having a legitimate expectation, is to have an entitlement generated by certain behaviour within tolerably just institutions.²⁷ Our understanding of justice grounds our understanding of which institutions are tolerably just, and so which expectations are legitimate.²⁸ This sense of desert is an example of the tautological connection of justice and desert, which we mentioned in section 4. It does not offer to illuminate justice by reference to desert, since desert in this sense is defined in terms of justice.²⁹ Rawls is keen that we distinguish this sense of desert from the second, which he calls 'moral desert'. The point of these remarks is to show how one can accept the tautological connection between justice and desert, whilst also rejecting a possible connection between justice and desert in its ordinary, agency-implicating sense.

Rawls is right about that distinction, which we examined in the text above. But R1 and R2 seem to be in tension with one another. Rawls first claims that reflection on desert favours his conception of justice, and then that justice does not depend on desert (in its usual, agency-implicating sense). If justice has nothing to do with desert, how can reflection on the nature of desert lead us to favour one conception of justice over another?

6. The Interpreter's Dilemma

Interpreters of Rawls's remarks on desert seem to be faced with a dilemma. The problem is to see how (R1) Rawls's conception of justice is favoured by thoughts about desert, whilst at the same time (R2) justice has nothing to do with desert. If R2 is taken seriously, it's difficult to see how getting clearer about the real bases of desert, or in any other way coming to understand desert better, can work to recommend a conception of justice. Rawls's interpreters seem bound to suppress either R1 or R2. They can't say, simply, that according to Rawls justice has nothing to do with *moral* desert, but that reflection on desert *in the sense of legitimate expectations*, favours his conception of justice. Such a view makes no sense since, as we've observed, the idea of legitimate expectations is dependent on a prior understanding of what justice requires, and so cannot be used as grounds for favouring one conception of justice over another. Instead, it seems, we can at best preserve one of Rawls's doctrines about desert, and we must choose between them.

Brian Barry presents a reconstructed version of Rawls's argument, which retains R1 but drops R2.³⁰ This solves the problem of seeing how reflections on desert can favour Rawls's conception of justice, by showing how that conception can be interpreted as incorporating a desert-tracking principle, the so-called *principle of redress*. The principle of redress requires undeserved inequalities to be compensated. Barry interprets Rawls as holding that inequalities of character, ability, and social circumstance are undeserved. If this claim about what people deserve is added to the principle of redress, a strong default preference for equality in distributive shares is, apparently, established.³¹ Rawls's conception of justice appears not to be desert-tracking only because it adds to these considerations some overwhelming considerations about the possible good effects of inequalities. On Barry's interpretation, the desert-tracking principle of redress is submerged but nevertheless present in Rawls's theory.

Barry's interpretation explains how reflections about desert could be thought to favour Rawls's conception. In reconstructing Rawls's theory as containing a desert-tracking principle, Barry also shows how the intuitive appeal which many have found in Rawls's remarks about desert can be confidently retained. But Barry's suggestion is not attractive for our purposes, for two reasons. First, it deviates too radically from what Rawls actually says: amounting, in fact, to a decidedly non-Rawlsian argument for Rawlsian conclusions.³² Secondly and more importantly, it makes Rawls's argument definitely non-naturalist, since it gives a central place to judgements of desert, in the ordinary agency-implicating

sense, in the justification of his conception of justice. We're looking for the most sympathetic, naturalist interpretation of Rawls's arguments.

Should we, in that case, suppress R1 instead of R2? We can call this the *High Rawlsian* approach, which retains the idea that justice has nothing to do with desert, but rejects the idea that reflecting on desert favours Rawls's conception of justice. The implication of such an approach is that no support for Rawls's conception can be garnered from thoughts about desert.

The High Rawlsian response to the dilemma is as unattractive as Barry's for our purposes. It looks promising as a naturalist version of Rawls's theory since it does not give the considerations of desert any role at all in Rawls's theory, but it pays too high a price, firstly in failing to account for the favoured interpretation of the second principle of justice, and secondly in forsaking all intuitive support from such considerations.³³ What we would like, if possible, is a version of Rawls's theory which retains the intuitive appeal of the considerations of desert, whilst at the same time preserving the hope of naturalism.

The way to do this is to retain the remarks on desert, but to rely on them only as *rebuttals* to rival desert-tracking conceptions of justice. According to such a view, Rawls's claims about the moral arbitrariness of natural and social advantages are to be understood as providing no support for his conception of justice, *except* in the sense that they undermine some of its rivals. This is admittedly, a weak sense in which the remarks on desert can be said to favour Rawls's conception, for the support provided depends obviously on the particular list of rivals which is considered. But it seems to be the strongest sense which remains compatible with naturalism, understood as the attempt to avoid relying on agency-implicating concepts.

7. Incoherence Versus Falsity

The central issue, if we adopt this third possible interpretation of Rawls's composite position on desert, is how exactly we construe the rebuttal of rival conceptions of justice. The rival conceptions which are vulnerable to this kind of rebuttal are those which claim (a) that justice depends on desert, and claim (b) that, given what people deserve, a Rawlsian society would be unjust. Such conceptions can be rebutted in three ways, broadly speaking. One could deny (a), which is the High Rawlsian approach we've just discussed. Or one could ignore (a), but say that the claims about what people deserve which are made in (b) are *false*. Finally, one could ignore (a), but say that the claims about what people deserve which are made in (b) are *incoherent*.

A lot seems to turn on whether the rebuttals are construed as claims of falsity or claims of incoherence. If we say some judgement 'Jones deserves *X*' is *false*, we are logically committed to asserting its negation, 'Jones does not deserve *X*'. Thus, we are committed, on this line of rebuttal, to some judgements about what people deserve. If we say instead that some collection of judgements of desert (say, 'Jones deserves *X*' and 'Smith does not deserve *X*') are jointly *incoherent*, we may escape any positive assertion other than that two or more statements are logically incompatible. So the incoherence rebuttal seems better suited to naturalism than does the falsity rebuttal.

One can interpret Rawls as claiming that the rival interpretations of the second principle of justice include, as their rationale, judgements of desert which are jointly incoherent. One can interpret him as claiming that the rivals all incoherently treat some but not all matters of luck, such as luck in one's natural abilities or social class, as bases of desert. In order to present this as incoherence, one must say something like this: since luck in natural talents and luck in social class are both kinds of luck, and since each of the rival conceptions treats one but not both as a basis of desert, the rival conceptions are committed both to asserting and to denying that luck is a basis of desert. Thus, without taking any stand whatsoever on what people deserve, one can claim that the rivals to Rawls's conception of justice are incoherent. The best naturalist interpretation of Rawls seems to attribute such a claim to him.

But in practice it is difficult to maintain a strong contrast between rebuttal on grounds of incoherence and rebuttal on grounds of falsity. The simplest response to a rebuttal on grounds of incoherence is to introduce a distinction between cases which removes the alleged inconsistency. Thus, for example, the proponent of a desert-tracking rival to Rawls's theory may say that there is a relevant distinction to be made between some kinds of luck and others, and hence that the real rationale for her view does not involve both asserting and denying that luck is a basis of desert, but instead asserting that some kinds of luck are a basis for desert, whilst quite consistently denying that some other kinds of luck are a basis for desert. The onus is then on the person who makes a charge of incoherence to explain why the proposed distinction is invalid, and *that* exercise typically involves claiming that some judgement or other is *false*, not simply inconsistent with other claims.

Ultimately, then, even the most sympathetic attempt to interpret Rawls's argument naturalistically seems forced to concede that it relies on some judgements about what people deserve; in order, if nothing else, to make charges of incoherence on the part of rival theories stick. We could take the High Rawlsian route, but that seems to pay too high a price. And we can't plausibly construe the rebuttal of rivals as based solely on charges of incoherence, since

in general such charges depend on background assumptions about which distinctions are eligible, and which are not, and these in turn depend on substantive views about the issues at hand.³⁴ We end up attributing something like this to Rawls: the rival conceptions of justice rely on judgements of desert which are inconsistent *with any sane view*. One could describe this as quasi-naturalist, since it involves reliance on some judgements of desert: it's just that the judgements in question are said to be obviously correct.

It is worth mentioning briefly a further respect in which Rawls's theory seems not to be naturalist. Rawls responds to an objection from Arrow, that primary goods are an unsatisfactory conception of advantage because they do not take into account relative cost of preferences, with the claim that persons are to be held responsible for revising their preferences in the light of such differences of cost.³⁵ It could be argued that the sense of 'responsible' here is a naturalist one, such that Rawls's claim means only that it is just to let persons bear the costs of their expensive tastes. But there is a price to such a move: interpreting the claim this way robs it of explanatory power, since Rawls's reply becomes simply the counter-assertion that it is, contrary to what Arrow suggests, fair to let persons bear these costs. If, on the other hand, we interpret the claim non-naturalistically, we have at least an explanation of why it might be fair to let them bear those costs, which is that people are typically able to revise their ends. Hence a naturalist interpretation of Rawls's views about responsibility for ends requires an alternative explanation of why it is fair to let persons bear the costs in question, which Rawls does not provide.

8. Conclusion

Egalitarians have often framed their views about justice in terms of responsibility, and sometimes in terms of desert. It is tempting to think that a conceptual connection between justice, on one hand, and responsibility or desert, on the other, is distinctive of egalitarian views, or even definitive of them. But I've argued that the important issue is not whether there is a conceptual connection between justice and responsibility or desert, but whether either of the latter concepts can be used to explain what justice requires. For we can always define responsibility or desert *in terms of justice* (thus establishing a conceptual connection between them), if we say that 'responsibility' means 'just cost or benefit' and 'desert' means 'just entitlement'. But these naturalist senses of desert and responsibility can't shed any light on what justice requires.

The important issue is whether we try to explain what justice requires in terms of standard, agency-implicating senses of desert and responsibility. Two kinds of consideration bear on this issue. The purely theoretical consideration is

whether or not what justice requires does depend on what people deserve or are responsible for, in a standard agency-implicating sense. Some people are anti-naturalist because they think it's obvious that justice does so depend. I offered a rudimentary argument for this view, but suggested that it was inconclusive. As I see it, the theoretical issue isn't settled.

There are also pragmatic considerations at issue, about the pros and cons of setting out to explain what justice requires in terms of persistently controversial, problem-ridden ideas such as desert or responsibility in their standard interpretations. One could think this route mistaken for pragmatic reasons even if one thought that justice depends on these concepts. But, by their nature, pragmatic considerations don't settle the theoretical issue.

Debate about what I've claimed is the important issue is hampered in two ways. It is hampered first by a tendency to concentrate instead on whether justice, responsibility, and desert are conceptually connected; and second by a lack of engagement between those who are immediately certain that naturalism is mistaken, for theoretical reasons, and those who think that somehow or other it has to be made to work, for pragmatic reasons. I suggested that one way out of this impasse was to examine a well-developed, arguably naturalist theory of justice, such as Rawls's, to look for evidence of the coherence of naturalism. If a coherent naturalist version of Rawls's theory is available, naturalism can't be generally incoherent. On the other hand, if all naturalist interpretations of Rawls's theory are incoherent, naturalism may be the fault. Examining Rawls's theory might help settle the important issue about justice and agency-implicating concepts.

Rawls's theory seems to be a good example of an attempt to formulate a naturalist theory of justice. He carefully distinguishes the idea of desert as we usually understand it from the naturalist idea of legitimate expectations, and he claims that justice should respect those expectations, but should not attempt to track desert. In these respects, his theory is classically naturalist, and it seems to have the benefits of naturalism, in avoiding the philosophical and other controversies which afflict agency-implicating concepts such as desert.

However, there are some respects in which Rawls's theory seems inescapably to rely on agency-implicating concepts. The most sympathetic interpretation of his remarks on desert construes them, as we've seen, as seeking to rebut rival conceptions of justice on the grounds that they are incompatible with any sane view about what people deserve. This seems to be a residual non-naturalist component of his theory, which is not easily removed. I argued that it is difficult to construe the rebuttal of rivals as wholly based on charges of their logical inconsistency, since such charges immediately invite a saving distinction, which must then be shown to be ineligible for the rebuttal to be successful. And

the High Rawlsian route, of dropping the rebuttal of rival conceptions, is unsatisfactory because it leaves the second principle of justice indeterminate, and because it forgoes the apparently important intuitive appeal of the claims about moral arbitrariness of natural and social advantages. Moreover, naturalist interpretations of Rawls's theory must explain why it is fair for persons with expensive tastes to bear the costs of those tastes, without resorting to standard interpretations of responsibility.

These elements of Rawls's otherwise naturalist theory, then, seem to resist naturalisation. If that diagnosis is correct, we may have some evidence that naturalism can't be carried through as a project; that justice is, in other words, itself an essentially agency-implicating concept. Any such inference must be treated with caution, of course, for the resistance to thoroughgoing naturalism which we seem to have identified in Rawls's theory may be the result of some unusual feature of that theory, rather than a symptom of the general non-viability of thoroughgoing naturalism. Nevertheless, we do have some reason at least to wonder whether that project is viable.

Notes

*I would like to thank Chris Bertram, G. A. Cohen, James Griffin, Susan Hurley, Philippe Van Parijs, Andrew Reeve, Hillel Steiner, Andrew Williams, and participants of a seminar at All Souls, Oxford, in May 1998 for helpful discussion and comments.

1. J. Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1972), section 12. References for relevant contributions to the debate about what egalitarians should try to equalise will be given below, in n. 8.
2. G. A. Cohen, "On the Currency of Egalitarian Justice", *Ethics*, 99 (1989), pp. 931-934 especially; R. J. Arneson, "Equality and Equal Opportunity for Welfare", *Philosophical Studies*, 56 (1989); J. E. Roemer, *Theories of Distributive Justice* (London: Harvard University Press, 1996), p. 276 and *passim*.
3. For an argument that egalitarianism should not be characterised in terms of the concept of responsibility, see S. L. Hurley, "Justice without Constitutive Luck", *Ethics*, ed. A. Phillips Griffiths (Cambridge: Cambridge University Press, 1993).
4. See S. Scheffler, "Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics", *Philosophy and Public Affairs*, 21 (1992). See also A. Ripstein, "Equality, Luck, and Responsibility", *Philosophy and Public Affairs*, 23 (1994); and S. Scheffler, "Individual Responsibility in a Global Age", *Social Philosophy and Policy*, 12 (1995).
5. For a discussion of the affinities between egalitarianism and the concept of desert, see R. J. Arneson, "Egalitarianism and the Undeserving Poor", *Journal of Political Philosophy*, 5 (1997). There (p. 328) Arneson writes: "any plausible theory of justice should be hospitable to norms of individual responsibility and desertness".
6. J. Rawls, *A Theory of Justice*, p. 72. See the discussion of Barry's interpretation of Rawls in the text below.
7. This seems to be Kymlicka's view of Rawls, in W. Kymlicka, *Contemporary Political Philosophy. An Introduction* (Oxford: Clarendon Press, 1990), Chapter 3.

8. This debate generated a very large number of publications. See, as a sample: A. Sen, 'Equality of What?', *Liberty, Equality, and Law*, ed. S. M. McMurrin (Salt Lake City: University of Utah Press, 1987); R. Dworkin, 'What is Equality? Part 1: Equality of Welfare', *Philosophy and Public Affairs*, 10 (1981); R. Dworkin, 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs*, 10 (1981); G. A. Cohen, 'On the Currency of Egalitarian Justice', *Ethics*, 99 (1989); R. J. Arneson, 'Equality and Equal Opportunity for Welfare', *Philosophical Studies*, 56 (1989); J. E. Roemer, 'A Pragmatic Theory of Responsibility for the Egalitarian Planner', *Philosophy and Public Affairs*, 22 (1993). Expensive tastes are those tastes which require an above-average amount of resources to satisfy.
9. For an example of the pragmatic argument, see J. Griffin, *Well-Being* (Oxford: Oxford University Press, 1986), p. 299.
10. See the works cited in n. 2 above.
11. G. A. Cohen, 'Incentives, Inequality, and Community', *Equal Freedom. Selected Tanner Lectures on Human Values*, ed. S. Darwall (Ann Arbor: University of Michigan Press, 1995). See also G. A. Cohen, 'Where the Action Is: On the Site of Distributive Justice', *Philosophy and Public Affairs*, 26 (1997).
12. Cohen, 'Incentives, Inequality, and Community', pp. 355-362.
13. J. E. Roemer, *Theories of Distributive Justice*, pp. 308-309.
14. Black's Law Dictionary defines strict liability as 'Liability without fault. Case is one of "strict liability" when neither care nor negligence, neither good nor bad faith, neither knowledge nor ignorance will save defendant'. *Black's Law Dictionary*, Revised Fourth Edition, H. Campbell Black (St. Paul, Minnesota: West Publishing Co., 1968), p. 1591.
15. I discuss briefly an interpretation of Rawls's doctrine of responsibility for preferences along these lines, in section 7 below.
16. Two things need to be noted here. First, I assume that it makes sense to divide concepts into those which are agency-implicating, and those which are not; and that *responsibility* is amongst the former in its usual interpretations, whilst *height*, say, is not; and that *travelled*, say, has ambiguous status, sometimes implicating our understanding of agency and sometimes not. It is of course easier to invoke this intuitive contrast than to give a satisfactory criterion for it, and I do not offer such a criterion here.
- Secondly, I make the assumption that, whenever responsibility (or desert) is used genuinely to illuminate what justice requires, it is being used in an agency-implicating sense. My only argument for this assumption, which may prove to be false, is that (a) it's at least possible to see how these concepts could illuminate justice, in their agency-implicating forms; and (b) the non-agency-implicating forms of these concepts seem to be just those which define them parasitically in terms of justice.
17. Strictly speaking, there are four possible views about the relationship between justice and responsibility. These are: (a) justice depends on responsibility, but not *vice versa*; (b) responsibility depends on justice, but not *vice versa*; (c) neither depends on the other; (d) they are interdependent. Interdependency views are interesting, and perhaps most plausible, but also more complex than the others, and I shall not discuss them here.
18. P. F. Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and other essays* (London: Methuen and Co., 1974).

19. Cohen expresses mixed views about this. He writes, '... we may indeed be up to our necks in the free will problem . . .', but then adds that he thinks 'a plausible nuancing approach reduces the dependence of political philosophy on the metaphysics of mind'. Cohen, 'On the Currency of Egalitarian Justice', at p. 934. Roemer's 'pragmatic' theory, on the other hand, is a response to the perceived persistence of problems with agency. J. E. Roemer, 'A Pragmatic Theory of Responsibility for the Egalitarian Planner'.
20. See the discussion in the text below (section 5) of Rawls's idea of legitimate expectations.
21. Isn't it odd to say that a *normative* theory, such as a theory of justice, may be 'naturalist'? It may sound odd, but it isn't really: it's not the normative entity, the theory, which is naturalist: what's naturalist is the attempt to avoid relying on agency-implicating concepts. It would be more accurate, perhaps, but certainly more cumbersome, to speak of 'theories born of naturalism', rather than 'naturalist theories'. I'm grateful to G. A. Cohen for raising this point.
22. The discussion of Rawls's remarks on desert is based on what he says in *Theory of Justice*, sections 12, 13, 17, and 48.
23. Rawls, *A Theory of Justice*, p. 60.
24. He insists that they should not be considered to be an argument for his conception of justice, since they do not employ the device of the original position: Rawls, *A Theory of Justice*, pp. 75, 66. They do seem to fit the argumentative scheme of reflective equilibrium, however, according to which our considered judgements are to be brought into equilibrium with our theoretical account of justice. In any case, as I shall discuss in the text below, they have a polemically important role in Rawls's discussion, which cannot easily be dispensed with.
25. Rawls, *A Theory of Justice*, p. 311. See also p. 103.
26. Echoing Feinberg, Rawls notes that we sometimes say of a game that 'the losing side deserved to win'. Such claims presuppose a sense of desert distinct from that of legitimate expectations, since in making them we do not necessarily dispute that the winner is, nonetheless, entitled to the spoils. Rawls explains this as a result of imperfections in the rules governing games and institutions, such that, '[t]he claims that individuals actually acquire inevitably deviate more or less widely from those that the scheme is designed to allow for'. But he immediately makes it clear that the distinct sense of desert involved here is not the idea of moral worth, and that justice has nothing to do with desert in the sense of moral worth. See *A Theory of Justice*, p. 314.
27. Rawls writes in explanation of the idea of legitimate expectations: 'it is true that as persons and groups take part in just arrangements, they acquire claims on one another defined by the publicly recognized rules'. *Theory of Justice*, p. 311. It is interesting to ask how just institutions must be before they can generate legitimate expectations.
28. Not only that: Rawls says also that 'the concept of moral worth . . . cannot be introduced until after the principles of justice and of natural duty and obligation have been acknowledged. Once these principles are on hand, moral worth can be defined as having a sense of justice . . .'. Rawls, *A Theory of Justice*, pp. 312-313. This suggests that Rawls conceives of *both* senses of desert – legitimate expectations *and* moral worth – as derivative of justice. A less marked naturalism might say that the idea of legitimate expectations is the sense of desert relevant to justice, and is derivative of the latter concept; but that there is a distinct and independent sense of desert, which we may call 'moral

worth'. (I do not claim that the marked naturalism detected here is Rawls's settled position, only that it seems present at this point in his remarks.)

29. Note that desert in this sense may not be an agency-implicating concept. The idea of legitimate expectations refers essentially to someone's behaviour (as well as to the justice of institutions), but that may not be agency-implicating. We can describe persons' behaviour in non-agency-implicating ways, and without further argument we should not assume that these kinds of description are insufficient to explain how persons come to have entitlements based on legitimate expectations. Thus the idea of legitimate expectations seems to be consistent with naturalism.

30. Brian Barry, *Theories of Justice* (Hemel Hempstead: Harvester-Wheatsheaf, 1989), pp. 213-215.

31. A critique of this idea is found in S. L. Hurley, 'Justice without Constitutive Luck', p. 185; see also G. A. Cohen, 'The Pareto Argument for Inequality', *Social Philosophy and Policy* 12 (1995).

32. Rawls states explicitly that his account of justice does not contain a desert-tracking principle, even as 'a prima facie principle'. Rawls, *A Theory of Justice*, p. 310. He does say that the difference principle 'gives some weight to the considerations singled out by the principle of redress' (p. 100), but this seems to me to mean only that, in placing a check on inequalities, the difference principle does part of what the principle of redress requires, though for different reasons.

33. In rejecting R1, the High Rawlsian approach also looks weak on exegetical grounds. Why did Rawls include the remarks on desert, supposedly favouring his interpretation of the second principle, in the first place? But, once again, we have more than just exegetical reasons for resisting this response to the interpreter's dilemma.

34. Parallel points about the lack of a sharp contrast between charges of falsity and charges of inconsistency are made, in a different context, in S. L. Hurley, *Natural Reasons* (New York and Oxford: Oxford University Press, 1989), pp. 57-64.

35. See J. Rawls, 'Social Unity and Primary Goods', in A. Sen and B. Williams (eds), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 168-169.