Sociaal en Cultureel Planbureau (1998), *Sociaal en Cultureel Rapport 1998. 25 jaar sociale verandering.* Rijswijk: Sociaal en Cultureel Planbureau.
Sociaal en Cultureel Planbureau (1999), *Sociale en Culturele Verkenningen 1999.* Rijswijk: Sociaal en Cultureel Planbureau.
Thomassen, Jacques J.A., and Jan W. van Deth (1989), 'How new is Dutch politics?', in: Hans Daalder and Galen A. Irwin (eds.), *Politics in The Netherlands. How Much Change?* London: Cass.
Topf, Richard (1995), 'Beyond electoral participation', in: Hans-Dieter Klingemann and Dieter Fuchs (eds.), *Citizens and the State.* Oxford: Oxford University Press.
Vaus, David de, and Ian McAllister (1989), 'The changing politics of women: gender and political alignment in 11 nations', *European Journal of Political Research* 17, pp. 241-62.
Verba, Sidney, Kay Lehman Schlozman and Henry E. Brody (1995), *Voice and Equality. Civic Voluntarism in American Politics.* Cambridge, Massachusetts: Harvard University Press.
Vinken, Henk (1997), *Political Values and Youth Centrism. Theoretical and Empirical Perspectives on the Political Value Distinctiveness of Dutch Youth Centrists.* Tilburg: Tilburg University Press.
Visscher, Gerard (1995), *Kiezersonderzoek op een dwaalspoor. De in politiek geïnteresseerde burger als selffulfilling prophecy.* Den Haag: SDU Uitgeverij.
Vollebergh, W.A.M., J. Iedema and W. Meeus (1999), 'The emerging gender gap: cultural and economic conservatism in the Netherlands 1970-1992', *Political Psychology* 20(2), pp. 291-321.
Welch, Susan (1977), 'Women as political animals? A test of some explanations for male-female political participation differences', *American Journal of Political Science* 21, pp. 712-730.
Wirls, Daniel (1986), 'Reinterpreting the gender gap', *Public Opinion Quaterly* 50, pp. 316-30.
Wittebrood, Karin (1995), *Politieke socialisatie in Nederland. Een onderzoek naar de verwerving en ontwikkeling van politieke houdingen van havo- en vwo-leerlingen.* Amsterdam: Thesis Publishers.
Zaller, John R. (1992), *The Nature and Origins of Mass Opinion.* Cambridge: Cambridge University Press.

Cohen and the Basic Structure Objection

Christopher Woodard
University of Warwick

Abstract

G.A. Cohen's discussion of the incentives argument for inequality has made an important contribution to our understanding of the normative theory of justice. The incentives argument is particularly difficult for egalitarians to rebut, yet Cohen seeks to show how egalitarians can mount a general defence against it. This paper argues that Cohen's critique has so far been construed too narrowly, and that this has resulted in the mistaken impression that his critique stands or falls with the refutation of the so-called basic structure objection. I explain Cohen's argument and the objection, and I explain why I think his critique is invulnerable to this objection if it is construed in a different way. I also point out that the critique, if construed this way, has much wider implications than is usually thought, applying not only to the incentives argument, or even to arguments about justice, but to most arguments in ideal theory.

1 Introduction

This paper examines G.A. Cohen's recent critique of the so-called 'incentives argument' for inequality (Cohen 1995a; Cohen 1995b; Cohen 1997). Its first aim is to show that Cohen's critique has been construed too narrowly, so that one particular kind of objection to it, the 'basic structure objection', has wrongly been thought to be decisive. Contrary to common opinion, the success of Cohen's critique of the incentives argument does not depend on the failure of the basic structure objection. At least one strand of his critique is immune to that objection.

Why should we be interested in these arguments? Cohen has played an important role in recent discussions on the normative theory of distributive justice, and especially in discussions about the issues raised by egalitarianism (for example, Cohen 1989; Cohen 1993; Cohen 1995c). His critique of the incentives argument is, in turn, an important part of his distinctive 'socialist-egalitarian' (as opposed to 'left-wing liberal') view of justice.[1] That gives us one reason for being interested in it. But, perhaps a better one is that it raises important and very general issues about the nature of normative political

argument in general – issues which extend beyond a concern with egalitarianism, and even beyond a concern with distributive justice. The second aim of this paper is to show exactly why his arguments have this general importance. The interpretation of his critique that I shall propose thus claims two advantages: it escapes the basic structure objection, and it reveals more clearly the very general nature of the issues Cohen raises.

I shall begin, in section 2, by trying to explain the place of the incentives argument and Cohen's critique in recent developments in egalitarian theories of justice. Section 3 explains the nature of the incentives argument and of Cohen's critique in more detail. Section 4 contains the interpretative argument, claiming that there are at least three separate strands of his critique, only two of which are commonly recognized. Section 5 traces the general issues raised by the third strand, and shows how they apply to a general class of normative political arguments, not just to arguments about incentives or even about justice. Section 6 identifies some possible objections to the preceding argument and seeks to rebut them.

2 The incentives argument and recent developments in egalitarianism

Cohen's critique of the incentives argument has contributed to the increasing sophistication of egalitarian views of justice. The issues it tackles became clearer as theorists of justice investigated the nature of egalitarianism, and Cohen's critique is an important further stage in this process. In this section I shall try to explain its place in this development, and hence its significance, although it is, of course, not possible to give a comprehensive review of the literature.

Normative theories of distributive justice attempt to specify the considerations that make distributions of goods (benefits and burdens) in society more or less just. Roughly speaking, we can divide recent developments in *egalitarian* theories of justice into two kinds: those concerned with the *metric of advantage* in considering distributions, and those concerned with the *pattern of advantage* in considering distributions.[2] In 1979 Amartya Sen focused attention on the first issue with his famous question, *Equality of What?* (Sen 1995). Obviously egalitarians do not aim to make people's lives equal in every respect. So what is it that egalitarians do aim, or should aim, to equalize? Candidate answers include the following: welfare, opportunity for welfare, resources, capability to function, and (Cohen's own answer) access to advantage.[3]

The second issue, which is certainly related to the first but can be distinguished from it, can be indicated by the following question: do, or should, egalitarians really aim for *strict* equality in these respects, and if not,

which inequalities do, or should, they allow and why? Here, the focus is on the exact pattern of distribution of advantage that is aimed at, rather than on which metric of advantage is relevant.[4] The importance of Cohen's critique of the incentives argument lies mainly in this second area, in which the justifiability of departures from strict equality is at issue.

To be worthy of their name, surely 'egalitarians' ought to be in favour of *strict equality* as the appropriate pattern of advantage, whatever metric of advantage they favour? Possibly, but many 'egalitarians' are not in favour of strict equality, and it seems inappropriate, at least in some contexts, to insist on this simple linguistic inference. Rawls, for example, is often considered to be a 'liberal-egalitarian', but of course his famous difference principle seeks to tell us which inequalities are permitted (Rawls 1972: 60-65).[5] In fact it has become clear that strict equality reflects only one of a number of concerns that might influence an egalitarian's choice of pattern (Frankfurt 1987; Nagel 1979; Parfit 1995; Raz 1986: Chapter 9; Temkin 1993).

One kind of concern is with *relative levels of advantage*.[6] How does one person's, or one group's, level of advantage compare with that of another person or group, or with the average level? A different kind of concern is with *absolute levels of advantage*: how well off, absolutely, is a particular person or group? Obviously people can be equally advantaged at low absolute levels, or at high absolute levels, or at any level in between, and the same goes for inequality: the gap between the most and least advantaged may be the same in two societies, in which the absolute levels of the least well off are quite different. A third kind of consideration is with the numbers of people at each level of advantage: could an improvement for a large number of moderately well-off people outweigh a loss for a much smaller number of less well-off people, for an egalitarian? Fourthly, if cardinal information is available, the size of improvements, or intensity of levels of advantage, may be important too.[7]

A principle requiring strict equality reflects concern with only one of these candidate factors, that of relative levels of advantage. More complex egalitarian patterns may balance or combine two or more factors, in different ways – and, in doing so, depart from strict equality. Indeed, it is plausible to think that considerations of absolute levels of advantage, or numbers, or intensity, can matter morally, and can matter with respect to justice in particular. Of course, egalitarians could relegate such considerations to the meagre role of breaking ties between strictly equal distributions, giving the consideration of relative levels of advantage strict priority. But, if they attribute any greater importance to these other concerns they must face the possibility that their principles favour unequal distributions of advantage, at least in some circumstances.

The incentives argument for inequality exploits this fact in a very powerful way. It says that inequalities can be justified because of their effects; in particular, by their effects via incentive mechanisms on the absolute levels of

advantage enjoyed by the least well off. The idea is that the least well off might have a higher absolute level of advantage in an inequality-permitting society than they would in a strictly equal society, because the former society can provide incentives for potentially productive persons to work hard and in the right areas, thus encouraging them to realize their productive potential, which increases the total stock of goods.[8]

Supposing that egalitarians do indeed have some concern with absolute levels of advantage, and not just with relative levels of advantage, this particular argument for inequality is very difficult for them to refute. Consider the two parts of the argument: first, an ethical claim, which specifies the particular good effects (on absolute levels of advantage) that would serve to justify inequalities; and second, some empirical claims to the effect that a certain set of inequalities does indeed have those justifying effects. An egalitarian may not accept a particular version of the ethical premise – Rawls's difference principle, say – but we have already supposed that he has some concern with absolute levels of advantage, so he cannot reject all possible versions of the ethical premise.[9] And, challenging the empirical claims can provide at best only a contingent refutation of the argument, which holds in some cases but not others. In some cases, the egalitarian should accept that inequalities might have the effects attributed to them, and that, in such cases, must accept that these inequalities are indeed justified.

The reason why the incentives argument is difficult for egalitarians to handle is that it plays off one of their concerns, namely absolute advantage, against another of their concerns, relative advantage or strict equality. In that sense it uses 'egalitarian' premises to argue for the justice of inequalities. In this respect, it can be contrasted with other arguments for inequality, which rely on premises the egalitarian is unlikely to accept. Consider the pointed issue of whether or not the possession of talents can justify receipt of high rewards. Egalitarians are unlikely to accept libertarian arguments about the entitlement of the talented to keep what they can earn with those talents (Nozick 1974). And, even if they embrace a form of egalitarianism that is hospitable to the concept of desert (Arneson 1997), they are unlikely to accept the view that talented individuals deserve all or most of the fruits of their talents. However, they might (have to) accept that justice requires that the talented receive more, if it can be shown that this benefits the least well off.

For egalitarians, then, the incentives argument is an unusually powerful argument for inequality. Not all versions of its ethical premise can be rejected by most egalitarians, and its empirical premises cannot be rejected for all circumstances without lapsing into dogmatism. Finally, we can explain the importance of Cohen's critique. Cohen seeks to show how egalitarians can reject the incentives argument as applied to a wide range of cases – and he seeks to show this without challenging either the ethical premise or the empirical

claims about the incentive-seeking behaviour of the potentially productive. Many commentators, possibly including Cohen himself, think that the success of his argument depends on the failure of the so-called 'basic structure objection' to his arguments (Estlund 1998; Murphy 1999; Williams 1998). This will be explained in the next section. However, my argument will be that at least one strand of Cohen's critique is invulnerable to the basic structure objection. Moreover, the issues it raises at the most fundamental level are issues about the character of normative political argument, and so have very wide application – they are not restricted to discussions of incentives, equality, or even justice. So far his critique has been construed much too narrowly.

3 Cohen's critique and the 'basic structure objection'

The incentives argument for inequality says that inequalities are justified because of their positive effects – in particular, their effects via incentive mechanisms on production. In its clearest form, the incentives argument claims that an unequal, incentive-permitting economic regime benefits people more than an equal economic regime without incentives. There are at least three variables in this argument:

a. which particular unequal regime is compared with which particular equal regime;
b. which particular empirical claims are made about the effects of incentives; and
c. which particular kinds of benefit are said to justify incentives?

One could imagine a range of different versions of the incentives argument, which combine different possible claims under each of these three headings.

Cohen focuses his critique on one particular version of the incentives argument, but, as I shall explain, his criticisms are meant to apply to such arguments in general. He considers a version that compares a regime in which the top rate of income tax is 60% (the 'equal regime', as I shall call it), with a regime in which the top rate is 40% (the 'unequal regime'), and in which the specification of the benefits that would justify incentives is given by Rawls's difference principle:[10]

> Economic inequalities are justified when they make the worst off people materially better off . . . When the top rate of tax is 40 percent, (a) the talented rich produce more than they do when it is 60 percent, and (b) the worst off are, as a result, materially better off . . . Therefore, the top tax should not be raised from 40 percent to 60 percent (Cohen 1995a: 339).

As already mentioned, Cohen's aim is to discredit the incentives argument in general, not merely to pick holes in this particular version. So he accepts both premises for the sake of argument (Cohen 1995a: 340). How then can he deny the conclusion, and hope to cast doubt on other versions of the same form of argument?

He tries to do this in two ways, both of which start from a single observation but develop in different directions. The observation is that we can distinguish between cases in which the reason the talented work less hard without incentives is that they are *unable* to work just as hard as they would with them, and cases in which they are instead *unwilling* to work just as hard without them (Cohen 1995a: 356; Cohen 1997: 8).[11] That is, we can distinguish two explanations for the behaviour reported in part (a) of the minor premise. On the face of it, it seems to be morally relevant which of these explanations is true. But is that really so? After all, the minor premise merely reports certain behaviour; it does not stake a claim about the explanation of this behaviour, so wondering about the true explanation does not seem to be suited to casting doubt on the premise. It is here that Cohen develops the observation in two directions.

One direction casts the critique of the incentives argument as an internal critique of Rawlsian employment of it. In this view, whether or not the talented are able to work just as hard without incentives is morally relevant, because Rawls argues that ideal societies are characterized by full compliance, where all members of such societies understand and try to act on the principles of justice (Cohen 1995a: 385-388; Cohen 1997: 9; Rawls 1972: 8, 145). If the talented are able to work just as hard without incentives, but refuse to do so, Cohen argues, they cannot be said to understand and be trying to act on the difference principle – because in doing so they bring about inequalities which are not necessary to benefit the worst-off.[12] So, on the assumption that they are able to work just as hard without incentives, the incentives argument cannot be used by Rawlsians to argue that inequalities are features of ideally just societies.[13] If that assumption is correct, such Rawlsian attempts are internally inconsistent.

The other direction in which Cohen develops his observation has more general application. He suggests that we recognize a special sense of justiication in claims about ideals, which he calls *comprehensive justification*. For an argument to be comprehensively justified, it must be the case not only that (a) its premises are justified, and that (b) the argument is valid, but also that (c) the *behaviour* mentioned in the premises is itself justified (Cohen 1995a: 347-353). For example, the conclusion that one should pay a kidnapper the ransom money demanded might well be justified in the ordinary sense, using a valid argument with plausible premises about avoiding harm to the hostage. But the conclusion is not comprehensively justified, because the

premises mention prospective behaviour on the part of the kidnapper – harming the hostage if he does not receive payment – that is not justified. So, if we recognize comprehensive justification as a requirement of claims about ideal societies, we will think it relevant whether or not the talented can justify their incentive-seeking behaviour, regardless of whether we also accept Rawls's views about full-compliance. In this way, Cohen's second development of his starting observation is more general than the first, since it relies on a claim about the requirements of ideal theory in general, not the requirements of Rawlsian ideal theory in particular.

Cohen's critique of the incentives argument begins, then, with the observation that it seems to matter whether or not the talented could work just as hard without incentives; and it continues by trying to explain the significance of this question in two ways, in terms of specifically Rawlsian assumptions about full-compliance, or instead in terms of related, but not specifically Rawlsian, views about justification in ideal theory. Cohen himself anticipates an objection to the first of these criticisms. A Rawlsian might say that the difference principle applies only to the basic structure of society, so that it is not incoherent for members of an ideal Rawlsian society to seek incentives (Cohen 1997: 10-11; Rawls 1972: 7). According to this *basic structure objection*, as it is known, the claim that Rawlsian defence of inequalities using the incentives argument is internally inconsistent rests on a confusion about the difference principle. Rawls distinguishes between the basic structure of society, to which the principles of justice are intended to apply, and the behaviour of individuals, which is not subject to the principles of justice.[14] Thus, talented individuals can hold out for incentives without violating the difference principle.

Against this objection, Cohen argues that no such clear distinction can be made between the basic structure of society and the behaviour of individuals – at least, not if Rawls's rationale for describing the basic structure as the subject of justice is remembered. The formal or coercive structure of society can be distinguished from the day-to-day behaviour of individuals. But, as Cohen notes,

> Rawls says that 'the basic structure is the primary subject of justice because its effects are so profound and present from the start.' Nor is that further characterization of the basic structure optional: it is needed to explain why it is primary, as far as justice is concerned. Yet it is false that only the *coercive* structure causes profound effects, as the example of the family . . . reminds us (Cohen 1997: 21, emphasis in the original).[15]

According to Cohen the basic structure objection faces a dilemma: either the rationale for claiming that the basic structure is primary is retained, in which case that structure must extend beyond the coercive structure of society to include such informal structures as the family, and day-to-day choices must be

counted as part of the basic structure; or the restriction of the basic structure to coercive structures is retained, in which case day-to-day behaviour is not part of that structure, but the rationale for claiming it is primary is lost. The narrow characterization of the basic structure according to which it can be fairly distinguished from day-to-day behaviour is in tension with the rationale for singling out the basic structure as the subject of justice (Cohen 1997: 17-23).

It is clear that the basic structure objection is pertinent to the strand of Cohen's critique that focuses on Rawlsian attempts to justify inequalities using the incentives argument. If the basic structure objection is sound, Cohen's claim that such attempts are internally inconsistent is refuted. But, is the objection pertinent also to the other strand of Cohen's critique, which explains the significance of the explanation for the behaviour of the talented by invoking the idea of comprehensive justification? Not if the version of the incentives argument under criticism is not at all Rawlsian – for then it is irrelevant whether Rawlsians can coherently claim that the difference principle does not apply to day-to-day behaviour. However, if the idea of comprehensive justification is used to criticize a Rawlsian version of the incentives argument, then that question is of course still relevant. The basic structure objection is pertinent whether Rawlsian versions are attacked on grounds of internal inconsistency or in the light of the external idea of comprehensive justification.[16]

So the basic structure objection is pertinent to Cohen's critique, as applied to Rawlsian versions of the incentives argument, whether that critique takes the internal form relying on the requirement of full compliance, or the external form relying on the requirement of comprehensive justification. Recently Cohen's reply to the basic structure objection has come under critical scrutiny, notably from Andrew Williams. Williams argues that an account of the basic structure is available which (a) retains the rationale for describing it as the subject of justice, and (b) excludes incentive-seeking behaviour. He claims that Cohen's dilemma presents two alternatives, which together do not exhaust the possibilities for accounts of the basic structure. It is possible to describe the basic structure in a way that goes beyond the coercive structures of society, but which does not expand so far as to include incentive-seeking behaviour (Williams 1998: 232-235). Williams emphasizes the importance of publicity for Rawlsian principles of justice, and he argues:

> Rawlsian principles . . . are inapplicable to certain types of decision. For some choices, although they may be profoundly influential, cannot be regarded as according with, or violating, public rules. Consequently the nonpublic strategies and maxims that individuals employ in making those choices need not be assessed as just or unjust by means of Rawlsian principles (Williams 1998: 235).

Moreover, he claims, decisions about whether to seek incentives fall into this category of non-public strategies, and so are not subject to the difference principle. Hence, he argues, the basic structure objection is not refuted by Cohen (Williams 1998: 236-247).

This debate about the basic structure objection is of obvious importance. At stake is whether justice requires only the realization of certain institutions, or also the establishment of a certain 'ethos' governing individual behaviour (Carens 1986; Cohen 1997; Williams 1998; Wolff 1998). The latter view makes justice seem much more demanding, since it requires each of us to act in ways that promote justice, even within just institutions.[17] However, I now want to argue that there is a third, neglected strand of Cohen's critique, which wholly escapes the basic structure objection, even when it is a Rawlsian defence of inequalities that is under scrutiny. In considering Cohen's critique of the incentives argument, we should not suppose that it stands or falls with the refutation of the basic structure objection.

4 The three strands of Cohen's critique distinguished

Recall Cohen's observation that it seems to matter whether the talented are unable to work just as hard without incentives or instead are unwilling to do so. The key question about this observation is how exactly we explain why it matters, if indeed we can find an explanation. For the incentives argument seems to need only to *report* certain behaviour, not to *explain* it; hence we need some account of why some explanations might unseat that argument. We have seen two ways in which Cohen answers this question, which constitute two distinguishable strands of his critique. According to one strand, it matters for Rawlsians whether the behaviour is explained by unwillingness or inability, because Rawlsians assume full compliance with the principles of justice in their arguments about ideal societies. If the explanation is unwillingness, the question of compliance arises; if the explanation is inability, it does not. According to the other strand, it matters whether the behaviour is explained by unwillingness or inability for anyone who accepts the requirement of comprehensive justification in arguments about ideal societies. Again, if the explanation is unwillingness, we can raise the question of whether the behaviour is justified; but if the explanation is inability, we cannot.

But there is a third possible explanation, which is implied by what Cohen says, although it is not as explicit as the other two. According to this explanation, arguments in ideal theory have a special nature, such that, where they rest on behavioural assumptions, those assumptions are not justified merely by the fact that they are *realistic* – that is, they are not justified merely because they report accurately the state of things in our actual, non-ideal,

world. So, if we ask whether incentives are features of ideally just societies, we should not try to justify any behavioural assumptions; we need to answer that question on the grounds that they accurately report behaviour in our actual world. On the other hand, a behavioural assumption would be unjustified if the behaviour it describes is impossible. So, the reason why we should be interested in whether incentive-seeking behaviour is due to unwillingness or inability, is not because we wonder whether it could be justified, but because we wonder about the limits of possible behaviour.[18]

This is quite a different explanation of why it should matter whether incentive seeking behaviour is due to unwillingness or inability. The root of the first two explanations we considered is a concern with the justifiability of behaviour mentioned in the premises of the incentives argument: both the Rawlsian assumption of full-compliance, and Cohen's idea of comprehensive justification, make the justifiability of institutional schemes dependent on the justifiability of behaviour mentioned in the arguments for those schemes. These ideas embed an extra level of justification within the justification of institutional schemes.

The third explanation, in contrast, points to a concern with what is possible. We ask whether the talented are able to work just as hard without incentives not because we are interested in calling them to account, not even hypothetically, but because we want to know what kinds of behaviour are possible. Here we do not have two levels of justification – one focused on the justifiability of institutional schemes, another focused on the justifiability of behaviour mentioned in the premises of arguments for the justifiability of institutional schemes – but only one. It is just that our attitude to behavioural assumptions is conditioned by the fact that they are to be used in ideal theory. Realism, understood as accurately reporting the features of our actual world, is not an adequate guide to which assumptions we should make in ideal theory, because our actual world is not ideal.

This alternative explanation of the significance of the explanation of the behaviour reported in the incentives argument is a good deal simpler than the other two. It does not rely on special assumptions about the nature of justification in ideal theory, according to which the requirement of full compliance, or alternatively of comprehensive justification, gives arguments in ideal theory a two-level character, whereby to be justified their behavioural claims must be true and the behaviour mentioned in them must itself be justifiable. It portrays justification in ideal theory to be just like justification elsewhere: as a matter of the justification of the assumptions used, plus the validity of the arguments constructed from the assumptions. Its distinctive claim is about what is needed to justify behavioural assumptions used in ideal theory. It says that realism, the accurate portrayal of the actual world, is insufficient.

This third strand of the critique is not only simpler than the other two: it is also untouched by the basic structure objection. The basic structure objection raises the issue whether incentive-seeking behaviour is justified by the lights of the difference principle. But, since the justification of incentive-seeking behaviour is not directly at issue in the third strand of the critique, which asks instead about the justification of behavioural premises, the basic structure objection seems to pass by this strand. It is true that the justification of behavioural premises in ideal theory is not entirely independent of the justification of behaviour mentioned in them, as I shall explain in Section 5. But the former is not wholly a matter of the latter.

It is puzzling, therefore, that the third strand of Cohen's critique has not been considered as a significant criticism of the incentives argument in its own right. It is more straightforward than the other two strands and it avoids the basic structure objection. If it is noticed at all, then perhaps there is suspicion that it is not really independent of the other two strands; that ultimately we must bolster the observation that a behavioural assumption in ideal theory is not justified by being realistic with the claim that the behaviour in question is also unjustified. But, we should not assume that the justification of behavioural premises in ideal theory turns on the permissibility of the behaviour reported in those premises.

5 Realism, acquiescence and ideal theory

The third strand of Cohen's critique is generated by taking a certain stance on the justification of behavioural premises in ideal theory. Such premises, it says, are not justified merely by their being realistic, in the sense that they accurately portray the features of our actual world. I should emphasize that this is a criticism of the incentives *argument*, not a putative refutation of that argument's *conclusion*. It casts doubt on the behavioural premise of the incentives argument when that argument is presented in ideal theory. As such, it does not show that the premise is certainly unjustified, and *a fortiori* it does not show that the conclusion of the incentives argument is certainly unjustified. (And it certainly does not show what talented people should earn.) Nonetheless, it is a real and powerful attack on the argument, because it calls attention to what is insufficient to justify its behavioural premise, which might otherwise be taken to be sufficient. We should not accept that premise on the grounds that talented people are incentive-seekers, even if we come to accept it on other grounds.

To get a clearer view of this third strand we need some diagnosis of why we are asked to reject realism as the appropriate criterion of justification of behavioural premises in ideal theory. Let us first consider why we might need

behavioural premises in ideal theory at all. Why do we need to make assumptions about individuals' behaviour in ideal theory? Can we not just say which institutions are required by justice?

The answer is that we need to make behavioural assumptions so long as we try to justify institutional schemes with any reference to their allegedly good effects. If we make any reference to good effects, we need to make assumptions about the behaviour of persons. Without such assumptions we cannot hope to characterize the effects of institutions. The effects of institutions, like the effects of other items, tend to vary greatly with changes in background conditions – which, in the case of institutions, include the behaviour of persons pre-eminently.[19] This variability of effects is just what Cohen calls our attention to with respect to tax schemes. But, it is easy to think of other examples too. Whether or not a scheme of parental leave, in which mothers and fathers have identical entitlements to leave, would have good effects, for example, depends on the behaviour of the mothers, fathers, and employers. Whether or not schemes of unconditional welfare payments would have good effects depends on how many free-riders there would be, and so on.

Some arguments in ideal theory do not refer to the effects of the institutional schemes under discussion, but most do. The incentives argument is certainly among those that do. Where effects are mentioned, behavioural assumptions must be made if we wish to characterize those effects accurately. And we may need such assumptions in an additional capacity, if we wish not only to say what the effects of an institutional scheme would be, but also what the effects of the relevant alternatives to it would be, for the purposes of comparative evaluation. (Behavioural assumptions might be needed not just for each characterization of effects, but also to determine the range of relevant alternatives.) So if we make arguments in ideal theory that rely even in part on claims about the effects of institutions, we need to rely on behavioural assumptions in those arguments.

Of course, it is possible to make some distinctions within the category of behavioural assumptions.[20] We can distinguish between people's occupational behaviour – their choice of how hard to work, and in what occupation – and the causes of that behaviour. With some simplification we might divide the latter into people's motivations, which may include a mix of self-interested and beneficent or justice-related motives, and the occupational context, which consists of the environment in which people make occupational choices, guided by their motivations, and may include such things as the tax system and the employment market. It is quite clear, even with this simplified model, that the 'effects of institutions' cannot be assigned wholly to what I have called the occupational context. At the very least, they are the result of the interaction between this context and people's occupational behaviour.

I take these to be the two main causes of the effects of institutions. But we

might notice a subtle influence of people's motivations on those effects too. Of course, people's motivations are amongst the causes of occupational behaviour, and so, for that reason, they indirectly cause institutional effects. But we might add to this a direct influence – if not on the raw effects, on their character. For the burdens of labour are part of the sum of justice-relevant advantages and disadvantages of persons, and these depend not only on how hard someone works, and in what job, but also on his or her attitude towards that work. Crudely, if someone views that work as something which others are entitled to expect from them, they will find it less of a burden than if they view it only as something which others are in a position to demand from them. Thus the justice-relevant character of the effects of institutions depends subtly on people's motivations to work. I shall return to this point later. For the moment the point is the simpler one that the 'effects of institutions' depend on the behaviour of persons. Hence, in order to characterize and evaluate those effects, we must make assumptions about that behaviour.

Now in ordinary deliberative contexts, where we are considering what action some particular agent, in some particular circumstances, should take, the grounds for making behavioural assumptions are fairly clear. Those assumptions should be true to the circumstances at hand; they should be realistic, in the sense that they accurately report the features of the circumstances under consideration. In these ordinary cases, the third strand of Cohen's critique has no application. If it is true that the talented will behave a certain way, a government considering raising the top rate of income tax from 40 per cent to 60 per cent should assume that that is how the talented will behave. It should not make unrealistic assumptions. In ordinary deliberative problems, then, the criterion of behavioural assumptions is realism.[21]

In ordinary deliberation, realism about behaviour can generate dilemmas of acquiescence.[22] In such cases, we believe that others will behave wrongly, and in doing so will make our preferred outcome unavailable. For example, we might be deciding how to respond to a kidnapper's demands (Cohen 1995a: 344-347). We are sure that he intends to make unavailable our preferred outcome – regaining the hostage without harm and without making any payment – and that this intention is wrongful. We may be unsure how exactly to treat the evaluation of our options in the light of this projected wrongful behaviour. But we should not be wantonly unrealistic, assuming that he will behave reasonably when all the evidence points the other way. To do that would be reckless.

Now consider ideal theory. Ideal theory is, perhaps, a kind of deliberation, but it is unlike ordinary deliberation because it is not addressed to any particular agent in any particular circumstances. It aims to describe the features of ideal societies, not to tell a particular agent how to respond to a particular problem.[23] For this reason, we cannot use the criterion of realism, understood

as the accurate portrayal of our actual world, to guide our behavioural assumptions in ideal theory (except in a limiting sense which I shall explain shortly). That criterion gets its point and its application from the orientation of ordinary deliberation to particular circumstances. Since ideal theory does not share that orientation, it does not share the criterion.

For this reason, dilemmas of acquiescence cannot arise in ideal theory. We get a dilemma of acquiescence, as I said, when we believe that others will behave wrongly and this affects the evaluation of our options. But we have no basis for this kind of prediction in ideal theory, because we cannot make sense of the idea of antecedent circumstances in a type of theory that is not addressed to any circumstances in particular. In ordinary deliberation, we know which circumstances to consider, and the criterion of realism can get to work, and in some cases can generate dilemmas of acquiescence. In ideal theory, in contrast, none of this can happen. One way of expressing the third strand of Cohen's critique, then, is to say that to accept the justice of incentives on the grounds given by the incentives argument is mistakenly to acquiesce in the behaviour of the talented,[24] despite the fact that dilemmas of acquiescence cannot arise in ideal theory. We think they can only if we mistake the nature of ideal theory, and import behavioural assumptions appropriate to ordinary deliberation that are out of place when considering ideals.

What then *should* guide the behavioural assumptions we must make in ideal theory if we refer to the effects of institutions?[25] There seem to be two uncontroversial kinds of ground, although even when added together they do not give as much guidance as we might like. One good reason we have for assuming, in ideal theory, that people behave a certain way, is that doing so is *morally required*. (Equally, we can rule out behaviour if it is morally forbidden.) Of course, the argument that such behaviour is morally required must be independent of the argument for which we need the behavioural premise, if we are to use this kind of reason without circularity. But sometimes we may have such independent arguments, and these can be used to ground behavioural assumptions in ideal theory.

The other kind of ground is given by the limits of what is humanly possible. There is a limiting sense, which I mentioned above, in which ideal theory is addressed to specific agents in specific circumstances: it is addressed to *us humans in our circumstances*. The agents and circumstances mentioned here are not at all particular, and so the constraints on behavioural assumptions that they generate are very weak; but there must be some such constraints. We should not assume that people in ideal societies behave in ways in which it is impossible for humans to behave. Nor should we imagine institutions or mechanisms that contravene the general constraints of social organisation, whatever those are.

This explains the significance of the explanation of incentive-seeking

behaviour. The incentives argument *works* if it is the case that the talented could not possibly work as hard without incentives as they could with them,[26] but not if they are only unwilling to do so, since unwillingness does not imply impossibility. Admittedly, the notion of 'possibility' at work here is not straightforward: it is broader than what is possible for people *as they are now*, whilst not being so broad as to rule out nothing (Nagel 1991: Chapter 3). Political disagreements sometimes turn on disagreement about exactly how broad this notion should be. But we may agree sometimes that certain patterns of behaviour are impossible given the kind of creatures humans are, and the kind of circumstances they inhabit.

The nature of ideal theory is such, then, that the criterion of realism cannot be used to guide our behavioural assumptions (except in the limiting sense that we can discuss the most general features of us and our circumstances, where 'us' means 'us humans'). There are two fairly uncontroversial grounds we can use to guide those assumptions – what is morally required, and what is possible in the broad and problematic sense just discussed. But, unfortunately, these grounds are unlikely to select a *single* pattern of behaviour that can be used in ideal theory to characterize and evaluate the effects of an institution. In many contexts there will not be just one kind of behaviour that is possible and permitted. I propose, therefore, that we add to the constraints of permissibility and possibility a third constraint, so that we may discriminate amongst the various patterns of behaviour that are both permissible and possible. This third constraint, of *optimality*, is embodied in the following necessary condition of justification by reference to good effects in ideal theory:

> An institution is justified in ideal theory by its effects only if, combined with some possible and permissible pattern of behaviour, it has effects which are better (in the relevant respect or respects) than the effects of any alternative institution which is combined with any possible and permissible pattern of behaviour.

The incentives argument violates this constraint in apparently trading on the misplaced criterion of realism, and characterising the effects of an equal regime *only* on the assumption that the talented would behave in the sub-optimal way that they actually do behave. Because of this narrow-mindedness it fails to justify its conclusion.

In most cases, then, the constraints of possibility and permissibility do not select a single pattern of behaviour; hence we must consider a very wide range of possible and permitted behaviour when we make claims in ideal theory about the effects of institutions. This makes the task very onerous. We must compare the effects of each institution, characterized in many different ways on many different behavioural assumptions, with the effects of each relevant alternative institution, again characterized in many different ways on many

different behavioural assumptions. Subject to other constraints we may impose, the best pair (possible institution plus possible and permitted pattern of behaviour) wins.

6 Optimality and ideal theory

The condition I have just proposed, and indeed the claim that there is a distinct third strand to Cohen's argument, rests on a view of ideal theory in which the considerations of permissibility and possibility do not exhaust the grounds we have for rejecting behavioural assumptions. This has an important effect on our attitude to arguments (such as the incentives argument) in which the behavioural premises describe actual behaviour. Since all actual behaviour is necessarily possible, the only wholly uncontroversial ground for dismissing such a premise is to claim that the behaviour in question is impermissible. This is what the first two versions of Cohen's critique do. But this concedes too much to the proponent of the incentives argument, and opens the door to the basic structure objection – which, in seeking to show that incentive-seeking behaviour falls outside the scope of the difference principle, attempts to head-off the claim that such behaviour is impermissible.

In contrast, the third strand of Cohen's critique does not depend on the claim that incentive-seeking behaviour is impermissible. Instead, it supposes that such behaviour may be sub-optimal. What is wrong with the incentives argument, in this view, is that it does not show convincingly that inequality-permitting institutions have the benefits claimed for them, because it characterizes and evaluates their effects on the basis of unduly narrow behavioural assumptions. The incentive-seeking behaviour that is presupposed by the incentives argument is alleged to be inappropriate not because it is impermissible (although that would suffice), but because it is probably sub-optimal behaviour. Hence, this version of the critique, if sound, is not unseated by the basic structure objection. But it should be admitted that the credibility of this third strand of the critique depends on our view of the criterion of optimality.

I want to try to forestall two possible objections to my use of this criterion, and my claim that Cohen's critique has this third strand. According to the first objection, talk of optimality in this context reflects conceptual confusion.[27] My claim is that we should select behavioural assumptions in ideal theory – in this case about persons' occupational behaviour – by considering three things: first, whether the behaviour in question is possible in the broad sense; second, whether this behaviour is morally permitted by the lights of principles that are independent of the argument under consideration; and third, whether this behaviour is optimal by the lights of the principles or values that are at stake

in the argument under consideration. Applied to the incentives argument, this general view about behavioural assumptions in ideal theory claims that incentive-seeking behaviour, while no doubt possible, and perhaps morally permitted by the lights of independent principles, is probably not optimal. But, it could be questioned whether this idea of optimality make sense.

There might be two reasons for doubting whether it does. The first, bad reason, is that it might be thought that we do not know which values or principles are at stake in judging whether some behaviour is 'optimal'. But this is wrong; the relevant values or principles are those in use in the argument of ideal theory that refers to good institutional effects. In the particular version of the incentives argument discussed by Cohen, for example, the relevant consideration is the difference principle. We are to consider whether incentive-seeking behaviour is optimal by the lights of this principle – which is to say that we are to consider whether it maximizes the level of primary goods obtained by the least well off. We are not to consider whether this behaviour is optimal in a wider sense, taking into account the other obligations, rights or entitlements people have. Thus there is no uncertainty about which considerations are to be used to judge optimality.

The second reason is more substantial. It casts doubt on the idea of 'optimal behaviour' by making essentially the same point as I made earlier about the effects of institutions. It might be said that behaviour by itself is ill-suited to being judged either optimal or sub-optimal, since its effects vary with institutional background. We might add to this the point mentioned earlier that the character of the raw effects generated by institutions plus behaviour varies with people's motivations (since these effects will be more or less burdensome depending on those motivations); hence we might conclude that talk of 'optimal behaviour' is nonsensical. But I think this is an exaggeration. I have stressed throughout that we should evaluate institutions in the context of specific behaviour, and I made the same point in the other direction when I stated the condition on justification at the end of Section 5: we should evaluate the effects of behaviour when taken together with a specific institutional context. We can add to this that the effects of either member of this pair, and of the pair as a whole, vary also with people's motivations. But this does not rob the expression 'optimal behaviour' of all sense. Optimal behaviour is that behaviour which features in the best combination of possible and permissible behaviour, institution and motivation, judged in the light of whatever values or principles are at stake in the particular argument of ideal theory we are considering.

Now in some cases, admittedly, the fact that 'optimal' makes sense as applied to a certain whole does not imply that it makes sense as applied to one or more of its parts. Consider the following example. It is a cold day, I choose to go for a walk, I should dress warmly. However, if I choose to go for a run, I should

dress lightly. On the basis of this information we can talk of optimal combinations of clothing and activity – walk/warm clothing; run/light clothing – but it does not make sense to talk of optimal clothing as such, if we do not know whether a walk or a run is planned (nor does it make sense, on this information alone, to talk of optimal activities). This is an example of optimality applied to a whole not transmitting to optimality of the parts taken separately, and we might worry that talking of 'optimal behaviour' is prey to the same difficulties. But this example does not provide an appropriate analogy.

What blocks talk of optimal clothing as such in our example is the (no doubt correct) assumption that the considerations that apply to the choice of clothing do not apply to the choice of whether to go for a run or for a walk. The two decisions are heterogeneous, answering to different considerations. Although choice of clothing follows from choice of activity in the light of certain considerations, the choice of activity is logically independent of choice of clothing (assuming both kinds of clothing are available), and it is not made on those same considerations. In contrast, the values and principles that are relevant to the specification of behaviour in ideal theory are the same as those that apply to the choice of institutions. The two decisions are homogeneous, and as a result it does indeed make sense to talk of optimal behaviour, not just of optimal behaviour given certain institutions.

Now let me turn to the second objection, which is that, even if the idea of optimality makes sense, it is inappropriate to ideal theory. A critic might suggest the following.[28] Suppose, for the sake of argument, the talented are morally permitted on independent principles to hold out for incentives. Perhaps respect for their liberty ensures this: they are not morally required, although they are morally permitted, to work just as hard without incentives as they do with them.[29] Suppose also that they are in fact unwilling to work just as hard without incentives. If both these assumptions are true, the critic claims, we should not calculate what justice requires on the assumption that the talented work just as hard without incentives. Although it might be the case that mere realism alone cannot justify behavioural assumptions in ideal theory, he suggests, realism together with the permissibility of the behaviour concerned justifies such assumptions. If people are actually disposed to seek incentives, and they are morally permitted to do so, we should evaluate institutions in ideal theory on the assumption that they seek incentives.

This critic claims that we should not go looking for other possible and permitted patterns of behaviour if there is a pattern that is actual and permissible. For example, we should not evaluate an equal tax regime on the assumption that the talented are willing to work just as hard without incentives, if the talented are morally permitted and actually disposed not to work as hard. After all, he says, we could not insist that they change their ways.

But is this objection valid? We can represent the implied argument as follows:

1. The talented are morally permitted not to work just as hard.
2. Therefore, it would be morally wrong to require them to work just as hard.
3. Therefore, we should not assume willingness to work just as hard.

Claim 1 is assumed to be true for the sake of argument. Claim 2 follows from claim 1, since if someone (or some group) is morally permitted not to do X, then it would be morally wrong to require them to do X. But claim 3 does not follow from claim 2 without the addition of the following claim, which we may dispute:

4. If we assume a pattern of behaviour in ideal theory, we are committed to requiring that pattern of behaviour.

This claim is needed to conclude that we should not assume willingness to work just as hard from the fact that it would be wrong to require willingness to work just as hard. But claim 4 is almost certainly false. Those who advocate assuming realistic attitudes to work in ideal theory should admit that it is false, since they presumably do not hold that we should require people to have *those* attitudes to work. In general, assuming a certain pattern of behaviour for the purposes of calculating what is ideal does not entail claims about that behaviour being morally required. It does entail claims about that behaviour being optimal, but not claims about it being required.

The critic might go on to question more directly whether the ideal theory of justice is concerned with optimality. According to such a view, the extent of our interest in people's behaviour in the ideal theory of justice is with whether or not they act wrongly. Where there is no wrongdoing, he might say, there is no injustice (at least, not so far as individual behaviour is concerned). So in this instance, he might continue, our focus ought to rest on the question of whether the behaviour of the talented is permissible; it ought not to move onto the question of what behaviour would be optimal on their part. Hence there is an important and controversial issue underlying the discussion about whether Cohen's critique really has a third strand with the quite general implications that I have discussed. The issue is whether the ideal theory of justice is properly concerned with optimizing benefits, or instead only with excluding wrong behaviour. If concern with optimality is out of place in ideal theory, the question about which behavioural assumptions to make would turn out to be the question of whether the actual behaviour of the talented is permissible. If that was correct, the third strand would after all be subject to the basic structure objection (as well as to an objection in terms of liberty, such as we have just been considering). But I hope to show that concern with

optimality is not only appropriate but inescapable in the type of argument in ideal theory we have been considering.

Views of the kind the critic advocates are appropriate to conceptions of justice that do not define justice, not even in part, in terms of good effects. Libertarians, for example, may adopt the view that the proper extent of our concern with people's behaviour in the ideal theory of justice is with whether or not they act wrongly. But such minimalism seems to be unavailable to those who define justice partly in terms of good effects.[30] They must make some behavioural assumptions, since institutions do not have effects just on their own. And they cannot hope to arrive at a single characterization of the effects of an institution by asking what behaviour is permissible, because of two interlocking facts. One is that, on the critic's own assumption, a range of different patterns of behaviour is permissible. The other is that the difference between these patterns of behaviour is almost certain to make a difference in the effects of institutions. It would be fortunate indeed, if the range of behaviour that is supposed to be permissible was such that the effects of the institutions we are interested in would be the same, whatever particular permitted behaviour we imagine. But Cohen's arguments show us that this is not the case, at least with respect to taxation, granting, for the sake of argument, that a range of different attitudes to work is permissible. The effects of a tax regime depend quite strongly on the different attitudes to work and remuneration that are imagined.

So those who define justice at least partly in terms of the good effects of institutions cannot hope to characterize the effects of tax regimes by imagining only that the behaviour of the talented is permissible. That leaves two candidates for a method of calculating effects: the critic's original suggestion, which is that we treat permissible actual behaviour as uniquely relevant, and the optimizing procedure I have suggested. Now, I have already claimed that treating permissible actual behaviour as uniquely relevant is arbitrarily narrow; but I cannot rely on that claim now, since we are considering a criticism of its presuppositions. But, there are two other possible arguments, which together defeat the critic's suggestion that optimality has no place in theories of justice.

One is that the alternative to optimality, treating permissible actual behaviour as uniquely relevant, has the effect of making the content of the ideal theory of justice strongly dependent on idiosyncratic features of our actual world. Of course, there is a sense in which the content of the ideal theory of justice ought to be dependent on features of our actual world. We want a theory of justice for us humans, not for radically different creatures. But this dependence is on only very general facts, about what kind of creatures humans are, what is possible for them, and so on – what Rawls calls "suitably general" facts (Rawls 1972: 155-161). It is not a dependence on fine-grained facts, about, for example, the attitudes that talented members of our society happen to have.

To think that what justice ideally requires depends on facts like those is odd; it runs against the notion of ideal theory itself, which gets its sense from a contrast with non-ideal theory, in which the features of particular circumstances are taken into account. (It might be said that as soon as we seek to specify what justice requires not only in terms of principles, but also in terms of institutional schemes with definite features, we have departed from ideal theory. Maybe ideal theory has to remain highly abstract. And yet if that is true, then the incentives argument cannot be made in ideal theory.)

The second argument is that the critic's proposed method, which is supposed to allow us to avoid a concern with optimality, is inapplicable where actual behaviour is impermissible, as must often be the case. We cannot select our behavioural assumptions to agree with actual permissible behaviour if actual behaviour is impermissible. What should we do in that case? Again, we cannot characterize the effects of institutions in terms of an undifferentiated notion of 'permissible behaviour', so long as more than one pattern of behaviour is permissible, and the differences between these patterns are significant. If we limit our selection criteria to possibility and permissibility we are bound to have seriously indeterminate results. It seems inescapable that we should ask which permissible behaviour, together with some institutional scheme, is *optimal*. Hence, I conclude, a concern with optimality is ineliminable from arguments in ideal theory about good effects. Therefore, the critic's charge, that to call treating actual permissible behaviour as uniquely relevant 'arbitrarily narrow' is wrongly to adopt a concern with optimality, is defeated. And the suspicion that the third strand is ultimately dependent on the other two, and is likewise vulnerable to the basic structure objection, is defeated with it.

7 Conclusion

The third strand of Cohen's critique of the incentives argument is genuinely independent of the other two, and it escapes the basic structure objection. It proceeds from an analysis of the grounds on which we should make behavioural assumptions in ideal theory, and charges the incentives argument with making arbitrarily narrow assumptions because of misplaced realism. Whilst realism is the appropriate criterion of behavioural assumptions in ordinary deliberation, it is not appropriate to ideal theory (except in the limiting sense discussed), because such theory is not addressed to any particular agent in any particular circumstances.

The puzzle about why the third strand has so far not been taken seriously can be explained as follows. Questions of justifiability are amongst the grounds of behavioural assumptions in ideal theory. In particular, if a certain pattern of behaviour is morally required, it is justified to assume that pattern of

behaviour as uniquely relevant in ideal theory – for any ideal society must contain it. Permissibility, too, makes for relevance, in the sense that a pattern of behaviour that is not permissible is not relevant. But this thought easily becomes the following thought, which is incorrect: permissibility plus realism makes for unique relevance, so that if a pattern of behaviour is a feature of our actual world and it is permissible, it should be treated as the only pattern of behaviour we should consider in ideal theory. This latter, incorrect, thought, supports the view that the third strand is not really independent of the other two, since questions about which behavioural assumptions to make would turn, in many cases, on questions about whether actual behaviour is permissible – about whether, for example, the talented can justify their own actual attitudes towards incentives. That is why it can seem as if we need to add the claim that the talented could not justify their behaviour to the observation that it is not the only possible pattern of behaviour. But in fact, questions about which behavioural assumptions we should make are not like that. Many different patterns of behaviour might be permissible; if so, we should consider all of them, not privilege those that happen to be actual.

Cohen's critique is thus significant in two ways. One important implication is local to egalitarian views of justice: it shows that one common argument for the justice of certain inequalities, which seems very difficult for egalitarians to rebut, can be rebutted in a way which is not dogmatic, and which is consistent with accepting the importance of absolute levels of advantage. It is important to emphasize that, at least in one of its forms, this rebuttal is not subject to the basic structure objection. Second, Cohen's critique has important implications for the epistemic burden of arguments in ideal theory that refer to the effects of institutions. This point is quite general: it is not restricted to arguments about justice in ideal theory. The behavioural premises of such arguments must be treated very carefully, since, as I have argued, they are not justified by realism, or by permissibility plus realism.

These conclusions depend on the assumption that ideal theory is legitimately concerned with optimizing benefits. Acknowledging the concern with optimality in ideal theory can raise the suspicion that we are being asked to design institutions for gods. But this kind of concern should be taken up in the discussion of what is possible, in the broad sense, for humans – which, I have claimed, is one of the three criteria on which we should select behavioural assumptions. The alternative to acknowledging a concern with optimality is to leave arguments of ideal theory seriously indeterminate in conditions where persons' actual behaviour is impermissible.

Acknowledgements

I am grateful for helpful comments on previous versions of this material to G.A. Cohen, James Griffin, Susan Hurley, Andrew Levine, Andrew Reeve, Hillel Steiner, Robert van der Veen, Andrew Williams, and several anonymous referees. I am grateful also to the Leverhulme Trust for support of the work in this paper as part of a project entitled *The Rationality of Acquiescence*.

Notes

1. For this distinction, see Cohen (1995a: 331-338).

2. This division is only rough, however, as it leaves out (at least) the following areas: our understanding of the rationale for egalitarianism (Dworkin 1995; Hurley 1993; Woodard 1998); and our understanding of intergenerational (Parfit 1984: Part 4) and international extensions of egalitarianism (Beitz 1979: Part 3).

3. Sen's question can be taken in at least two ways. It can be taken to ask what the appropriate conception of advantage is, for an egalitarian (what the 'metric' is), or instead what would be equalized in an egalitarian society. But for our purposes nothing turns on this distinction. Equality of welfare and equality of resources were famously discussed by Dworkin (1981a; Dworkin 1981b). Dworkin defended the resource view. Equality of opportunity for welfare is proposed by Richard Arneson (1989). Sen proposed equality of capability to function (Sen 1993; Sen 1995). Cohen reviews much of the literature, and defends his preferred metric of access to advantage (Cohen 1989). Rawls's metric, the idea of primary goods, may be counted as a variant of the resource view, though his formulation of the idea predates Sen's lecture and answers somewhat different theoretical concerns (Rawls 1972: 90-95).

4. The two issues are related because a single distribution of goods could be described by several equivalent pairs, where a 'pair' consists of: (a) a conception of advantage; and (b) a specified pattern of distribution. But theorists of justice have tended to operate with something like the distinction presented here, in order to avoid discussing all the issues at once. For example, see Temkin (1993: 10-11, n. 15).

5. A further complication, which is clear in Rawls's account and shared by some others, is that the difference principle is not meant to apply directly to the distributions of goods, but instead to the design of basic social institutions. These institutions form the background against which individual behaviour generates distributions of goods. But in what follows, for the sake of simplicity, I shall ignore this procedural aspect. See Rawls (1972: 83-90).

6. Note that, in this context, 'advantage' is not essentially comparative (although of course, 'relative levels of advantage' is). As we will see, theorists often discuss a single person's 'level of advantage', where this means how well-off that person is absolutely, rather than how well-off compared to others. To have a level of advantage, in this sense, is not necessarily to be better off than some other(s) in some respect(s).

7. Rawls's difference principle is sometimes criticized for being numbers-blind and

intensity-blind (Nagel 1979: 121-122; Sen 1995: 318-320).

8. A clear statement of the incentives argument can be found in Cohen (1995a: 339-340). I shall explain it in more detail in Section 3 below.

9. Again supposing that the concern with absolute levels of advantage is not limited to tie-breaking.

10. A regime in which the top rate of income tax is 60% is unlikely to be strictly equal, perhaps. But the simplification involved here does not affect the nature of Cohen's argument, or my diagnosis of it. It is sufficient that this society is more equal than one in which the top rate is 40%.

11. The observation comes at the 'start' not in the sense that it appears towards the beginning of the relevant texts, but in the sense that it is a common root of all three strands of Cohen's critique, as I shall explain in due course.

12. Rawls's difference principle states that "Social and economic inequalities are to be arranged so that they are . . . to the greatest benefit of the least advantaged..." (Rawls 1972: 302). Cohen's contention is that it is inconsistent to hold out for incentives whilst understanding and adhering to this principle.

13. Cohen argues that, leaving aside what he calls 'special burden cases' (in which the extra money received is, strictly, an equality-restoring compensation, rather than an inequality-generating incentive), potentially productive people could work just as hard without incentives as they do with incentives. Hence, incentives are necessary to benefit the badly-off only because of the attitudes and behaviour that productive people choose to adopt (Cohen 1995a: 355-362). It is obviously vitally important to these issues exactly what the talented are capable of doing. For example, Samuel Scheffler is reported by Cohen (1995a: 359-361) as querying whether incentives might, in some cases, be psychologically necessary for the talented to work as hard as they possibly can. But I shall not address that issue here, since it affects only the scope of Cohen's critique, not its force where it does apply.

14. Liam Murphy criticizes this 'dualist' conception of justice (Murphy 1999).

15. The phrase Cohen quotes is from Rawls (1972: 7).

16. It is perhaps arguable whether the idea of comprehensive justification is external to Rawls's theory – we might see it as a fuller account of the requirement of full compliance. But what is clear is that the idea of comprehensive justification can be extended to non-Rawlsian arguments – so that it can be used, for example, to criticize non-Rawlsian versions of the incentives argument, without fear of compromise by the basic structure objection. It is important to notice that that objection is not pertinent to Cohen's criticisms as applied to those versions of the incentives argument. I shall argue later that another strand of his critique is independent of the basic structure objection, even when applied to Rawlsian arguments.

17. There are obvious connections here with concerns over the demands made by morality more generally. On these, compare for example Kagan (1989) and Scheffler (1992).

18. I believe that what Cohen says implicitly contains this third strand. Consider for example what he says (1995a: 379) about 'strict' and 'lax' readings of the difference principle: "We confront . . . two readings of the difference principle: in its *strict* reading, it counts inequalities as necessary only when they are, strictly, necessary, necessary, that

is, apart from people's chosen intentions. In its *lax* reading, it countenances intention-relative necessities as well." This suggests, to me at least, that the fact (if it is one) that the talented actually have incentive-seeking intentions is not sufficient to justify assuming that they do for the purposes of ideal theory. This thought is naturally generalized to the argument I call the third strand. But Cohen himself tends to *combine* points about not accepting behavioural assumptions just because they are realistic with points about the justifiability of such behaviour. So when I refer to 'the third strand of Cohen's critique' I mean to refer to a criticism of the incentives argument which is certainly suggested by his writings, but which he may or may not personally accept, absent the further points about the *justifiability* of incentive-seeking behaviour. For further evidence of the plausibility of this interpretation, see Cohen's remarks about the inappropriateness of basing arguments in ideal theory on the habits of the talented (Cohen 1995a: 358-359).

19. Some causes have the interesting feature of producing more or less the same effect, more or less regardless of changes in background conditions (Lewis 1986: 184-188). But most institutions are not like this: their effects vary significantly with changes in the behaviour of individuals.

20. I am grateful to an anonymous reviewer for discussion of the issues raised in this paragraph and the next.

21. If this is not the whole truth about ordinary deliberative problems, it is close to it. I examine the grounds for behavioural assumptions in ordinary deliberation in *The Rationality of Acquiescence*, work in progress.

22. On the concept of acquiescence, see Woodard (2000).

23. For a similar claim, see Räikkä (1998: 30). We might try to *derive* prescriptions from ideal theory, by adding to it premises about the actual circumstances of particular agents, but it undermines the distinctiveness of ideal theory to say that it consists of such prescriptions. It does not, because the point of it is not to tell any particular agent in any particular circumstances what to do. (But it is not for that reason an attempt to occupy the view from nowhere. It is an attempt to get a view, from *here*, about what ideal societies are like. I am grateful to Andrew Levine for discussion of this point.)

24. As before, on the assumption that the incentives are not strictly necessary for the good effects.

25. For related discussion, see Bertram 1993; Räikkä 1998: 32.

26. As Cohen agrees (1995b: 172).

27. I am grateful to an anonymous reviewer for alerting me to this possible objection.

28. Robert van der Veen drew my attention to this objection, and I am grateful to him and to Andrew Williams for helpful discussion of the issues involved.

29. Cohen raises the issue of occupational choice at the end of his discussion of the 'Pareto argument' for inequality (Cohen 1995b: 184-185). Williams also briefly discusses the issue (Williams 1998: 228-229).

30. Rawls defines the justice of particular distributions in a purely procedural way; but the justice of the background conditions that constitute the procedure is defined, via the difference principle, in terms of good effects. He says of his conception of justice that "it contains a large element of pure procedural justice. No attempt is made to define the just distribution of goods and services on the basis of information about the

preferences and claims of particular individuals ... But if the notion of pure procedural justice is to succeed, it is necessary ... to set up and to administer impartially a just system of surrounding institutions. The reliance on pure procedural justice presupposes that the basic structure satisfies the two principles" (Rawls 1972: 304).

Bibliography

Arneson, R.J. (1989), 'Equality and equal opportunity for welfare', *Philosophical Studies* 56, pp. 77-93.

Arneson, R.J. (1997), 'Egalitarianism and the undeserving poor', *Journal of Political Philosophy* 5(4), pp. 327-350.

Beitz, C.S. (1979), *Political Theory and International Relations*. Princeton, NJ: Princeton University Press.

Bertram, C. (1993), 'Principles of distributive justice, counterfactuals and history', *Journal of Political Philosophy* 1(3), pp. 213-228.

Carens, J. (1986), 'Rights and duties in an egalitarian society', *Political Theory* 14(1), pp. 31-49.

Cohen, G.A. (1989), 'On the currency of egalitarian justice', *Ethics* 99(4), pp. 906-944.

Cohen, G.A. (1993), 'Equality of what? On welfare, goods, and capabilities', in: M. Nussbaum and A. Sen (eds.), *The Quality of Life*. Oxford: Clarendon Press.

Cohen, G.A. (1995a), 'Incentives, inequality, and community', in: S. Darwall (ed.), *Equal Freedom*. Ann Arbor: University of Michigan Press.

Cohen, G.A. (1995b), 'The Pareto Argument for inequality', *Social Philosophy and Policy* 12(1), pp. 160-185.

Cohen, G.A. (1995c), *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press.

Cohen, G.A. (1997), 'Where the action is: on the site of distributive justice', *Philosophy and Public Affairs* 26(1), pp. 3-30.

Dworkin, R. (1981a), 'What is equality? Part 1: equality of welfare'. *Philosophy and Public Affairs* 10(3): 185-246.

Dworkin, R. (1981b), 'What is equality? Part 2: equality of resources', *Philosophy and Public Affairs* 10(4), pp. 283-345.

Dworkin, R. (1995), 'Foundations of liberal equality', in: S. Darwall (ed.), *Equal Freedom*. Ann Arbor: University of Michigan Press.

Estlund, D. (1998), 'Liberty, equality and fraternity in Cohen's critique of Rawls', *Journal of Political Philosophy* 6(1), pp. 99-112.

Frankfurt, H. (1987), 'Equality as a moral ideal', *Ethics* 98(1), pp. 21-43.

Hurley, S.L. (1993), 'Justice without constitutive luck', in: A. Phillips Griffiths (ed.), *Ethics*. Cambridge: Cambridge University Press.

Kagan, S. (1989), *The Limits of Morality*. Oxford: Clarendon Press.

Lewis, D. (1986), 'Causation', in: D. Lewis, *Philosophical Papers Volume II*. Oxford: Oxford University Press.

Murphy, L.B. (1999), 'Institutions and the demands of justice', *Philosophy and Public Affairs* 27(4), pp. 264-269.

Nagel, T. (1979), 'Equality', in: T. Nagel, *Mortal Questions*. Cambridge: Cambridge University Press.

Nagel, T. (1991), *Equality and Partiality*. New York: Oxford University Press.

Nozick, R. (1974), *Anarchy, State, and Utopia*. Oxford: Basil Blackwell.

Parfit, D. (1984), *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (1995), *Equality or Priority?* The Lindley Lecture. Lawrence: University of Kansas.

Räikkä, J. (1998), 'The feasibility condition in political theory', *Journal of Political Philosophy* 6(1), pp. 27-40.

Rawls, J. (1972), *A Theory of Justice*. Oxford: Oxford University Press.

Raz, J. (1986), *The Morality of Freedom*. Oxford: Clarendon Press.

Scheffler, S. (1992), *Human Morality*. New York: Oxford University Press.

Sen, A. (1993), 'Capability and well-being', in: M. Nussbaum and A. Sen (eds.), *The Quality of Life*. Oxford: Clarendon Press.

Sen, A. (1995), 'Equality of what?', in: S. Darwall (ed.), *Equal Freedom*. Ann Arbor: University of Michigan Press.

Temkin, L.S. (1993), *Inequality*. New York: Oxford University Press.

Williams, A. (1998), 'Incentives, inequality, and publicity', *Philosophy and Public Affairs* 27(3), pp. 226-248.

Wolff, J. (1998), 'Fairness, respect, and the egalitarian ethos', *Philosophy and Public Affairs* 27(2), pp. 97-122.

Woodard, C. (1998), 'Justice, responsibility and desert', *Imprints* 3(1), pp. 25-48.

Woodard, C. (2000), 'The concept of acquiescence', forthcoming.