

*Incentives, Inequality, and Community*

*G. A. COHEN*

THE TANNER LECTURES ON HUMAN VALUES

Delivered at

Stanford University

May 21, 23, 1991

G. A. COHEN was educated at McGill and Oxford universities where he obtained, respectively, the degrees of B.A. in Philosophy and Politics and B. Phil in Philosophy in 1963. For twenty-two years he was a Lecturer and then a Reader in Philosophy at University College, London. In 1985 he became Chichele Professor of Social and Political Theory and a Fellow of All Souls, Oxford. Professor Cohen is the author of *Karl Marx's Theory of History: A Defence* (1978) and *History, Labour and Freedom* (1988). He was made a Fellow of the British Academy in 1985.

. . . the rulers of mankind . . . maintain side by side two standards of social ethics, without the risk of their colliding. Keeping one set of values for use, and another for display, they combine, without conscious insincerity, the moral satisfaction of idealistic principles with the material advantages of realistic practice.

—R. H. Tawney, *Equality*

## I. THE INCENTIVE ARGUMENT, THE INTERPERSONAL TEST, AND COMMUNITY

### 1.

In March of 1988, Nigel Lawson, who was then Margaret Thatcher's chancellor of the exchequer, brought the top rate of income tax in Britain down, from 60 to 40 percent. That cut enlarged the net incomes of those whose incomes were already large, in comparison with the British average, and, of course, in com-

Many friends and colleagues commented helpfully, and in some cases at magnanimous length, on earlier versions of the material forming these lectures. For perceptive admonitions at various Oxford meetings I thank Ronald Dworkin, Susan Hurley, Thomas Nagel, Derek Parfit, Thomas Scanlon, Samuel Scheffler, and Joseph Raz. For (often voluminous) letters of criticism, I thank Richard Arneson, John Baker, Annette Barnes, Gerald Barnes, Christopher Bertram, Akeel Bilgrami, Giacomo Bonanno, Joshua Cohen, Gerald Dworkin, Jon Elster, Keith Graham, Daniel Hausman, Ted Honderich, Will Kymlicka, Andrew Levine, Kasper Lippert-Rasmussen, Murray MacBeath, John McMurtry, David Miller, Michael Otsuka, Derek Parfit, Philip Pettit, Thomas Pogge, Janet Radcliffe-Richards, John Roemer, Amelie Rorty, Miles Sabin, Robert Shaver, William Shaw, Seana Shiffrin, Hillel Steiner, Joseph Stiglitz, Robert Ware, Martin Wilkinson, Alan Wertheimer, Andrew Williams, Joseph Wolff, and Erik Wright. My greatest debt is to Arnold Zuboff, who devoted countless hours to arguing with, and correcting, me, and I am particularly grateful to Samuel Scheffler for his incisive criticism at the seminar following the lectures. I should add that some of the argumentation of the lectures is anticipated in these articles about Rawls: Thomas Grey, "The First Virtue," *Stanford Law Review* 25 (January 1973); Jan Narveson, "A Puzzle about Economic Justice in Rawls' Theory," *Social Theory and Practice* 4, no. 1 (1976); and Joseph Carens, "Rights and Duties in an Egalitarian Society," *Political Theory* 14, no. 1 (February 1986).

parison with the income of Britain's poor people. Socialists hated the tax cut, and a recent policy document says that the Labour party would, effectively, restore the pre-1988 rate.<sup>1</sup>

How might the Lawson tax cut be defended? Well, economic inequality is no new thing in capitalist society, so there has been plenty of time for a lot of arguments to accumulate in favor of it. We hear, from the political right, that rich people are entitled to their wealth: to part of it because they produced it themselves — but for them, it would not have existed — and to the rest of it because it was transferred to them voluntarily by others who were themselves entitled to it because they produced it, or because they received it as a gift or in voluntary trade from others who were themselves entitled to it because . . . (and so on). (Some who hold that view also think that it is because it establishes moral desert that production justifies title, while others find the entitlement story compelling even when the idea of desert plays no role in it.) And then there is the utilitarian proposition, affirmed not only on the right but in the center, that inequality is justified because, through dynamizing the economy, it expands the gross national product and thereby causes an increase in the sum of human happiness.

Left-wing liberals, whose chief representative in philosophy is John Rawls, reject these arguments for inequality: they do not accept the principles (entitlement, desert, and general utility) which figure in their major premises.<sup>2</sup> But the right and center

<sup>1</sup> Strictly speaking, the top tax would be raised to 50 percent, but the ceiling on National Insurance contributions would be removed, and the effect of the two measures would be the same as that of raising the income tax to 59 percent and leaving National Insurance alone.

<sup>2</sup> To be more precise, they reject those principles *at the relevant fundamental level*. The qualification is necessary because left-wing liberals recognize desert and entitlement as (derivative) rules of legitimate reward in schemes of contribution and compensation which are not *grounded* in notions of desert and entitlement. (See John Rawls, *A Theory of Justice* [Cambridge: Harvard University Press, 1971], pp. 103, 310–15; and Thomas Scanlon, “The Significance of Choice,” in *The Tanner Lectures on Human Values*, Vol. 8 [Salt Lake City: University of Utah Press, 1988], pp. 188, 203. For a recent statement of nuanced views on desert and entitlement,

sometimes offer an additional argument for inequality, to the major premise of which the liberals are friendly. That major premise is the principle that inequalities are justified when they render badly off people as well off as it is possible for such people to be.<sup>3</sup> In one version of this argument for inequality — and this version of it is the topic of these lectures — their high levels of income cause unusually productive people to produce more than they otherwise would; and, as a result of the incentives enjoyed by those at the top, the people who end up near the bottom are better off than they would be in a more equal society. This was one of the most politically effective justifications of the unequalizing policy of Thatcher Conservatism. We were ceaselessly told that movement contrary to that policy, in a socialist egalitarian direction, would be bad for badly off people, by advocates of a regime which seems itself to have brought about the very effect against which its apologists insistently warned.<sup>4</sup>

Left-wing liberals deny the factual claim that the vast inequalities in Britain or America actually do benefit the badly off, but they tend to agree that if they did, they would be justified, and they defend inequalities that really are justified, in their view, by the incentive consideration. That is a major theme in John Rawls's work. For Rawls, some people are, mainly as a matter of genetic and other luck, capable of producing more than others are, and it is right for them to be richer than others if the less fortunate are caused to be better off as a result.<sup>5</sup> The policy is warranted by what Rawls calls the difference principle, which endorses all and

---

see John Rawls, *Justice as Fairness: A Briefer Restatement* [Cambridge: Harvard University, 1989, manuscript], pp. 54, sec. 2, and 57 n. 34; I do not understand the doctrine presented in the latter place).

<sup>3</sup>For extensive use of this principle, see F. A. Hayek, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1960), chap. 3, and esp. pp. 44–49.

<sup>4</sup>Strong support for that charge comes from *Punishing the Poor: Poverty under Thatcher*, by Kay Andrews and John Jacobs (London: Macmillan, 1990).

<sup>5</sup>See Rawls, *A Theory of Justice*, pp. 15, 102, 151, 179, 546; Rawls, *Justice as Fairness*, pp. 57, 89.

only those social and economic inequalities that are good for the worst off, or, more generously, those inequalities that either make the worst off better off or do not make them worse off: in this matter there is a certain ambiguity of formulation in Rawls, and in what follows I shall take the difference principle in its more generous form, in which it allows inequalities that do not help but also do not hurt the worst off.<sup>6</sup>

Back now to the socialist egalitarians, who did not like the Lawson tax cut. Being to the left of left-wing liberals, socialist egalitarians are also unimpressed by the desert, entitlement, and utility justifications of inequality. But it is not so easy for them to set aside the Rawlsian justification of inequality. They cannot just dismiss it, without lending to their own advocacy of equality a fanatical hue which they could not themselves on reflection find attractive.

Socialist egalitarians say that they believe in equality. We might well think that they count as egalitarians because equality

<sup>6</sup> Statements of the difference principle display ambiguity along two dimensions. There is the ambiguity remarked in the text above, between inequalities that *do not harm* and inequalities that *help* the badly off, and there is the further ambiguity between *mandated* and *permitted* inequalities. These distinctions generate the following matrix:

	Mandated		Permitted
Helping ones are	1	→	2
Non-harming ones are	3	→	4

Since what is mandated is permitted, and what helps does not harm, there exist the implications among possible interpretations of the principle indicated by the arrows above, and there are five logically possible positions about which inequalities are mandated and which allowed: all are mandated (1,2,3,4); helping ones are mandated, and others forbidden (1,2); none are mandated and only helping ones are permitted (2); none are mandated and all non-harming ones are permitted (2,4); helping ones are mandated and all non-harming ones are permitted (1,2,4). Rationales can be provided for each of these five points of view, and I believe that there are traces of all of them in the letter and/or spirit of various Rawlsian texts. (Although, as I have said, I take the difference principle in a form in which it allows *all* non-harming inequalities, my critique of Rawls is consistent with his holding any of the positions distinguished above: it depends only on his allowing helping inequalities and forbidding harming ones, and that stance is a constituent in each of the five positions).

is their premise, But the structure of that premise is too simple to accommodate the thought that gets them going politically, which is: why should some people be badly off, when other people are so *well* off? That is not the same as the colorless question, Why should some people be better off than others? for in that question there is no reference to absolute levels of condition, hence no reference to anyone being badly off, as opposed to just *less* well off than other people are. Maybe some egalitarians would maintain their zeal in a world of millionaires and billionaires in which no one's life is hard, but the politically engaged socialist egalitarians that I have in mind have no strong opinion about inequality at millionaire/billionaire levels. What they find wrong is that there is, so they think, unnecessary hardship at the lower end of the scale. There are people who are badly off and who, they believe, would be better off under an equalizing redistribution. The practically crucial feature of the situation is that the badly off are worse off than anyone needs to be, since an equalizing redistribution would enhance their lives.

For these egalitarians, equality would be a good thing because it would make the badly off better off. They do not think it a good thing about equality that it would make the well off worse off. And when their critics charge them with being willing, for the sake of equality, to grind everyone down to the level of the worst off, or even lower, they do not say, in response: well, yes, let us grind down if necessary, but let us achieve equality on a higher plane if that is possible. Instead, what they say is somewhat evasive, at the level of principle; they just deny that it is necessary, for the sake of achieving equality, to move to a condition in which some are worse off and none are better off than now. Were they more reflective, they might add that, if leveling down were necessary, then equality would lose its appeal. Either it would make the badly off worse off still, in frustration of the original egalitarian purpose, or it would make the badly off no better off, while others are made worse off to no evident purpose. Relative to their initial

inspiration, which is a concern about badly off people, an inequality is mandatory if it really is needed to improve the condition of the badly off, and it is permissible if it does not improve but also does not worsen their condition.

Accordingly, these egalitarians lose sight of their goal, their position becomes incoherent or untrue to itself, if, in a world with badly off people, they reject the difference principle and cleave to an egalitarianism of strict equality. (Given the priorities and emphases that I have attributed to them, they should, strictly speaking, affirm as fundamental neither equality nor the difference principle but this complex maxim: make the badly off well off, or, if that is not possible, make them as well off as possible. But, on a modestly demanding interpretation of what it means to be well off, and on a realistic view of the world's foreseeable resource prospects, the practical consequences of the complex maxim are those of the difference principle.) We might conclude that the socialist egalitarians that I have in mind should not be called "egalitarians," since (if I am right) equality is not their real premise. But that conclusion would be hasty, and I shall say more about the property of the name "egalitarian" in a moment.

For my part, I accept the difference principle, in its generous interpretation (see above), but I question its application in defense of special money incentives to talented people. Rawlsians think that inequalities associated with such incentives satisfy the principle. But I believe that the idea that an inequality is justified if, through the familiar incentive mechanism, it benefits the badly off is more problematic than Rawlsians suppose; that, at least when the incentive consideration is isolated from all reference to desert or entitlement, it generates an argument for inequality that requires a model of society in breach of an elementary condition of community. The difference principle can be used to justify paying incentives that induce inequalities only when the attitude of talented people runs counter to the spirit of the difference principle itself: they would not need special incentives if

they were themselves unambivalently committed to the principle. Accordingly, they must be thought of as outside the community upholding the principle when it is used to justify incentive payments to them.<sup>7</sup>

Speaking more generally, and somewhat beyond the limited brief of these lectures, I want to record here my doubt that the difference principle justifies *any* significant inequality, in an unqualified way. The principle allows an inequality only if the worst off could not benefit from its removal. And I believe that it is in general more difficult than liberals suppose to show that the worst off could not benefit from removal of an inequality, and hence in general more difficult than liberals think it is to justify an inequality at the bar of the difference principle. The worst off benefit from incentive inequality in particular only because the better off would, in effect, go on strike if unequalizing incentives were withdrawn. This inequality benefits the badly off only within the constraint set by the egalitarian attitude, and the consequent behavior, of the well off, a constraint that they could remove. And an inequality can also benefit the badly off within a constraint set, not by egalitarian attitudes *per se*, but by preexisting unequal structure. Thus, in a country with state medical provision, the inequality of treatment that comes from allocating a portion of hospital resources to high-fee-paying patients who get superior care benefits the badly off when some of the revenue is used to raise standards throughout the service. The unequal medical provision helps poor people, but only against the background of a prior income inequality (which no doubt itself reflects further structural inequality

<sup>7</sup> Although I shall press against left-wing liberals the thought that community cannot tolerate the inequalities that they endorse, I need not deny that enormous inequalities coexisted with community in premarket societies. For, if that was indeed true, then the coexistence was possible because of general acceptance, and, more particularly, because of acceptance by the less well off of ideologies of destiny and place which left-wing liberals do not countenance. That community can go with inequality when people believe things that liberals regard as false does not show that they can go together in a society possessed of a modern consciousness.

and inegalitarian attitude) that has not, within this argument, itself been shown to benefit them.

The farther back one goes, temporally and causally, in the construction of the feasible set, the more one encounters open possibilities that were closed by human choice, and the harder it is to identify inequalities that do not harm the badly off. Bringing the two cases distinguished above together, I conjecture that social inequalities will appear beneficial to or neutral toward the interest of those at the bottom only when we take as given unequal structures and/or inequality-endorsing attitudes that no one who affirms the difference principle should unprotestingly accept.<sup>8</sup>

Now if all that is right, then we might, in the end, in a round-about way, vindicate the application of the term “egalitarian” to the socialists that I have had in mind, provided that they are willing to tolerate a formulation of their position along lines just foreshadowed. For we might say that a person is an egalitarian if he applies the difference principle in circumstances in which there exist badly off (as opposed to just less well off) people *and* he believes that what the principle demands, in those circumstances, is equality itself, if, that is, he believes that in the long run, and prescindng from rooted inegalitarian attitudes and practices, there *are*, in such circumstances, no social inequalities that do not harm the worst off. Equality appears, at first, to be a premise. It is then rejected, *as* a premise, when the ream for wanting equality is clarified: it is rejected in favor of the difference principle (or, strictly, the more complex maxim stated at p. 268 above). But, now grounded in (something like) the difference principle, it reasserts itself as a conclusion, for our world, in these times, and for the foreseeable future.

<sup>8</sup> We can also say that inequalities are necessary to improve the condition of the badly off when we take for granted, not, as above, causal, but moral imperatives. Thus incentives can indeed be judged necessary to raise the condition of the badly off when elements of the desert and entitlement rationales that left-wing liberals reject are affirmed.

## 2.

I return to Rawls and the difference principle in Part III of these lectures. Right now I want to focus on Nigel Lawson's tax cut, and on the incentive case against canceling it, the case, that is, for maintaining rewards to productive people at the existing high level. And I shall consider that case only with respect to those who, so it is thought, produce a lot by exercising skill and talent, rather than by investing capital. Accordingly, the argument I shall examine applies not only to capitalist economies but also to economies without private ownership of capital, such as certain forms of market socialism. Of course, there also exists an incentive argument for high returns to capital investment, but I am not going to address that argument in these lectures.

Proponents of the incentive argument say that when productive people take home modest pay, they produce less than they otherwise do, and, as a result, relatively poor and badly off people are worse off than they are when the exercise of talent is well rewarded. Applied against a restoration of the top tax to 60 percent, the argument runs as follows:

Economic inequalities are justified when they make the worst off people materially better off. [Major, normative premise]

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. [Minor, factual premise]

Therefore, the top tax should not be raised from 40 percent to 60 percent.

It is immaterial to present concerns how the circumstance alleged to obtain in part *a* of the minor premise of the argument is supposed to occasion the result described in part *b*. One possibility is that the rich work so much harder when the tax rate goes down that the tax take goes up, and more is available for redis-

tribution. Another is that, when the rich work harder, they produce, among other things, (better) employment opportunities for badly off people.

I am going to comment negatively on the incentive argument, but my criticism of it will take a particular form. For I shall focus not, directly, on the argument as such, but on the character of certain utterances of it. Accordingly, I shall not raise questions about the validity of the argument, or about the truth of its premises, save insofar as they arise (and they do) within the special focus just described. And I shall not, in particular, pursue possible doubts about the minor, factual, premise of the argument. I shall question neither claim *a*, that the supposedly talented rich are more productive when they are more generously rewarded, nor claim *b*, that the badly off benefit from the greater productivity of the well off affirmed in *a*. I do not aim to show that the minor premise of the incentive argument is false.

The critique that follows is not of everything that could be called an incentive, but only of incentives that produce inequality and which are said to be justified because they make badly off people better off. I raise no objection against incentives designed to eliminate a poverty trap, or to induce people to undertake particularly unpleasant jobs. It is not constitutive of those incentives that they produce inequality. My target is incentives conferring high rewards on people of talent who would otherwise not perform as those rewards induce them to do. I believe that the familiar liberal case for incentives of that kind has not been thoroughly thought through.

### 3.

I said that I would criticize the incentive argument by focusing on certain utterances of it. For I believe that, although the argument may sound reasonable when it is presented, as it usually is, and as it was above, in blandly impersonal form, it does not sound so good when we fix on a presentation of it in which a talented

rich person pronounces it to a badly off person. And the fact that the argument undergoes this devaluation when it occurs in that interpersonal setting should affect our assessment of the nature of the society that the incentive justification by implication recommends.

A normative argument will often wear a particular aspect because of who is offering it and/or to whom it is being addressed. When reasons are given for performing an action or endorsing a policy or adopting an attitude, the appropriate response by the person(s) asked so to act or approve or feel, and the reaction of variously placed observers of the interchange, may depend on who is speaking and who is listening. The form, and the explanation, of that dependence vary considerably across different kinds of case. But the general point is that there are many ways, some more interesting than others, in which an argument's persuasive value can be speaker-audience-relative, and there are many reasons of, once again, different degrees of interest, why that should be so.

Before describing a form of dependence (of response on who is addressing whom) that operates in the case of the incentive argument, and in order to induce a mood in which we think of arguments in their contexts of delivery, I list a few examples of the general phenomenon:

(a) I can argue that the driver over there should not be blamed for just now making a right turn on a red light, since he does not know that the rules are different outside California. But he cannot, at the moment, make that very argument, entirely sound though it may be.

(b) You want the fishing rod for recreation, and I need it to get my next meal. I know that you are so unstoical that you will be more upset if you do not get to fish than I will be if I do not get to eat. So I let you have the rod, and I cite your hypersensitivity to disappointment as my reason. It would be a lot less good for you to give that as a reason why you should have the rod.

(c) I might persuade my fellow middle class friend that, because my car is being repaired, and I consequently have to spend hours on the buses these days, I have a right to be grumpy. The same conclusion, on the same basis, sounds feeble when the audience is not my friend but a carless fellow bus passenger who is forced to endure these slow journeys every day.

(d) As designers of advertisements for charitable causes know, our ordinary self-serving reasons for not giving much (we need a new roof, I'm saving for my holiday, I'm not actually *very* rich) sound remarkably lame when we imagine them being presented to those for whom our lack of charity means misery and death.<sup>9</sup>

(e) And such quotidian reasons also sound feeble when they are presented to people whose sacrifice for the cause is much larger than the one the speaker is excusing himself from offering.”

<sup>9</sup> “How do you tell a person dying of hunger that there's nothing you can do?” (*ActionAid* leaflet, 1990).

<sup>10</sup> An exploitation of (inter alia) this particular relativity occurred in an advertisement of 1943 whose purpose was to promote the purchase of war bonds. In March 1944 the advertisement won a prize for its contribution to the war effort. The top third of the ad's space pictures an American prisoner of war in a bleak cell. Below the picture, we find the following text:

WILL YOU WRITE A LETTER to a Prisoner of War . . . tonight?

Maybe he's one of Jimmie Doolittle's boys. Perhaps he was left behind when Bataan fell. Anyway, he's an American, and he hasn't had a letter in a long, long time.

And when you sit down to write, tell *him* why you didn't buy your share of War Bonds last pay day.

“Dear Joe,” you might say, “the old topcoat was getting kind of threadbare, so I . . .”

No, cross it out. Joe might not understand about the topcoat, especially if he's shivering in a damp Japanese cell.

Let's try again. “Dear Joe, I've been working pretty hard and haven't had a vacation in over a year, so . . .”

Hell, better cross that out, too. They don't even get vacations where Joe's staying.

Well, what are you waiting for? Go ahead, write the letter to Joe. *Try* to write it, anyhow.

But mister, if somehow you find you can't finish that letter, will you, at least, do this for Joe? Will you up the amount of money you're putting into War Bonds and keep buying your share from here on in? (1945 *Britannica Book of the Year [A Record of the March of Events of 1944]* [Chicago: Encyclopaedia Britannica, 1945], p. 22).

A word about the form of this ad, and about the sources of its motivating power (if it did the motivating it should have done to deserve the prize it won).

(f) Since the pot should not call the kettle black, an employee may be unimpressed when a routinely tax-evading well-heeled superior dresses him down because of his modest appropriations from petty cash.

The examples show that arguments vary in their capacity to satisfy because of variations in people's epistemic (a) or moral

The ad is directed not, of course, at one person but at a large set of people, all the people in the condition of material life and personal intention of the civilian that the ad sketches. Yet the ad speaks as though to one person and it has that single person address a single member of the set of POWs. The content of the ad implies that civilians as such have some kind of obligation to POWs as such. But the ad aims to convey the obligation falling on many by selecting one individual from each of the two groups and figuring forth an encounter between them. Notice, moreover, that the ad would have sacrificed little or nothing of its purpose and power if its personal references had been pluralized, if, that is, the civilian had spoken of *our* threadbare coats, run-down sheds, and lacks of vacation, to an imagined *assembly* of POWs. (Compare first-person plural presentation of the incentive justification, by a rich person, or by all of them in unison, to all the poor people),

The ad makers thought that they could expose the insufficiency of the reasons civilians give themselves for not buying bonds by portraying a civilian offering them to Joe. And they were right that it is easier to face yourself when you decide for the stated reasons not to buy bonds if you do not have to face Joe at the same time.

The power of the ad to move the reader is multiply determined, mingling elements that go into types *c*, *d*, and *e* above. The ad simulates an immediacy between the civilian and Joe, such immediacy being one rhetorical effect of casting an argument in interpersonal form. And then, immediacy having been secured, there are two or three separable things, mixed here in a powerful cocktail, on which the ad relies: that Joe and I are members of the same community, and he is suffering; that Joe and I are coparticipants in an immensely important enterprise in which at least the *quality* of my life and that of the members of my family is at stake; and that Joe is a moral hero — look what he has given, for the sake of the mentioned enterprise, compared with the modest thing that I resist giving. These considerations combine to make me feel answerable to Joe. The ad says that although it sounds quite reasonable for a person to choose a new coat before buying more bonds, the burden of wearing a threadbare coat carries no justificatory weight when it is compared with the burden Joe carries: that, so the ad implies, explains the shame a civilian would feel in telling Joe that his threadbare coat was a good reason for not buying more bonds.

Finally, a comment on the role of immediacy, which, so I noted, is one source of the advertisement's power. Immediacy can contribute to persuasion in cases where what is rendered immediate is not a person (or a group) that is addressed. We do not speak to animals, but arguments justifying their use in certain experiments might be hard to deliver in the lab while those experiments are in train. We also do not speak to trees, but it might be harder to justify the size of the Sunday edition of the *New York Times* when one is standing in a majestic forest. So: having to face a person when uttering an argument is a special case of immediacy, not part of its general form, and it is perhaps not crucial to the ad's power that the POW is *addressed*, as opposed to just on scene when the argument is presented.

(*e, f*) or social (*c*) position, or because of issues of tact and embarrassment (*c, d, e*) and immediacy (*d*), or because being generous is more attractive than being grabby (*b*). I shall not here attempt a systematic taxonomy of ways that arguments subside in different sorts of interpersonal delivery. Instead, I pass to a type of case which is of special interest here, since the incentive argument belongs to it.

4.

In this type of case, an argument changes its aspect when its presenter is the person, or one of the people, whose choice, or choices, make one or more of the argument's premises true. By contrast with other presenters of the same argument, a person who makes, or helps to make, one of its premises true can be asked to justify the fact that it is true.<sup>11</sup> And sometimes he will be unable to provide a satisfying justification.

For a dramatic example of this structure, consider the argument for paying a kidnapper where the child will be freed only if the kidnapper is paid. There are various reasons for not paying. Some concern further consequences: maybe, for example, more kidnapping would be encouraged. And paying could be thought wrong not only in some of its consequences but in its nature: paying is acceding to a vile threat. You will nevertheless agree that, because so much is at stake, paying kidnappers is often justified. And the argument for paying a particular kidnapper, shorn of qualifications needed to neutralize the countervailing reasons mentioned a moment ago, might run as follows:

Children should be with their parents.

Unless they pay him, this kidnapper will not return this child to its parents.

*So*, this child's parents should pay this kidnapper.

<sup>11</sup> As opposed to the claim that it is true, which every presenter of the argument can be asked to justify.

Now, that form of the argument is entirely third-personal: in that form of it, anyone (save, perhaps, someone mentioned in the argument) might be presenting it to anyone. But let us now imagine the kidnapper himself presenting the argument, to, for example, the child's parents. (What will matter here is that he is doing the talking, rather than that they are doing the listening: the latter circumstance achieves prominence in section 11 below.) The argument that follows is the same as that given above, by an unimpeachable criterion of identity for arguments: its major premise states the same principle and its minor premise carries the same factual claim:

Children should be with their parents.

I shall not return your child unless you pay me.

So, you should pay me.

Notice, now, that despite what we can assume to be the truth of its premises and the validity of its inference, discredit attaches to anyone who utters this argument in the foregoing interpersonal setting, even though uttering the same argument in impersonal form is, in most cases,<sup>12</sup> an innocent procedure. And there is, of course, no mystery about why the argument's presenter attracts discredit in the exhibited interpersonal case. He does so because the fact to which he appeals, which is that you will get your child back only if you pay, is one that he deliberately causes to obtain: he makes that true, and to make that true is morally vile.

When he presents the argument, the kidnapper shows himself to be awful, but it is hardly necessary for us to reflect on his utterance of the argument to convince ourselves that he merits disap-

<sup>12</sup> I express myself in that cautious way because, apart from the case, if you want to allow it, in which the kidnapper himself uses the impersonal form of the argument, referring to himself as "he," there is the case of a person who puts it forth and conveys (for example, by his tone) that he is quite insensitive to the countervailing (if properly overridden) considerations, and/or that he sees nothing untoward in the kidnapper's threat, and/or that he sees human dealings on the model of interaction of impersonal forces.

proval. Independently of any such reflection, we amply realize that the kidnapper's conduct is wrong, and we need not be particularly scandalized by his frank avowal of it. Indeed, in certain instances a kidnapper's presentation of the argument will be a service to the parents, because sometimes his utterance of the argument's minor premise will, for the first time, put them in the picture about how to get their child back. One can even imagine a maybe slightly schizoid kidnapper suddenly thinking, "Omigod, I've forgotten to tell the kid's parents!" and experiencing some concern for them, and for the child, in the course of that thought.

Yet although what is (mainly) bad about the kidnapper is not his voicing the argument, but his making its minor premise true, he should still be ashamed to voice the argument, just because he makes that premise true. The fact that in some cases he would do further ill not to voice the argument does not falsify the claim that in all cases he reveals himself to be ghastly when he does voice it.

In the kidnapper argument, there are two groups of agents, the kidnapper and the parents, both referred to in the third person in the initial presentation of the argument, and referred to in the first and second persons in its revised presentation. Consider any argument that refers to distinct groups of people, A and B. There are many different ways in which such an argument might be presented. It might be uttered by members of A or of B or of neither group, and it might be addressed to members of either group or of neither. And all of that applies to the incentive argument, with the groups being talented rich people on the one hand and the worst off on the other. In my treatment of the incentive argument I shall mainly be interested in the case where a talented rich person puts it forward, sometimes no matter to whom and sometimes where it matters that poor people are his audience; and at one point I shall consider the opposite case, where a poor person addresses the argument to a talented rich one.

The incentive argument has something in common with the kidnapper argument, even though there are major differences be-

tween withholding a hostage and withholding labor until one gets the money one desires. But before looking more carefully at similarities and contrasts between the kidnapper and incentive arguments, I want to explain why the word “community” appears in the title of these lectures.

## 5.

In its familiar use, “community” covers a multitude of conditions, and I shall introduce the particular condition that I have in mind by relating it to the concept of a *comprehensive justification*.

Most policy arguments contain premises about how people will act when the policy is, and is not, in force. Schemes for housing, health, education, and the economy typically operate by altering agents’ feasible sets, and their justifications usually say what agents facing those sets can be expected to choose to do.

Consider, then, a policy,  $P$ , and an argument purportedly justifying it, one of whose premises says that a subset,  $S$ , of the population will act in a certain fashion when  $P$  is in force. We engage in what might be called *comprehensive assessment* of the preferred justification of  $P$  when we ask whether the projected behavior of the members of  $S$  is itself justified. And *comprehensive justification* of  $P$  obtains only if that behavior is indeed justified.<sup>13</sup>

“We should do  $A$  because they will do  $B$ ” may justify our doing  $A$ , but it does not justify it comprehensively if they are not justified in doing  $B$ , and we do not provide a comprehensive justification of our doing  $A$  if we set aside as irrelevant the question whether they are justified in doing  $B$ . Thus, insofar as we are expected to treat the incentive argument as though no question arises about the justification of the behavior of the talented rich that its minor premise describes, what we are offered may be a justification, but it is not a comprehensive justification, of the incentives policy.

<sup>13</sup> It follows, harmlessly, that penal policies adopted to reduce the incidence of crime lack comprehensive justification. The very fact that such a policy is justified shows that all is not well with society.

Now, a policy argument provides a comprehensive justification only if it passes what I shall call the *interpersonal test*. This tests how robust a policy argument is, by subjecting it to variation with respect to who is speaking and/or who is listening when the argument is presented. The test asks whether the argument could serve as a justification of a mooted policy when uttered by any member of society to any other member. So, to carry out the test, we hypothesize an utterance of the argument by a specified individual, or, more commonly, by a member of a specified group, to another individual, or to a member of another, or, indeed, the same, group. If, *because* of who is presenting it, and/or to whom it is presented, the argument cannot serve as a justification of the policy, then whether or not it passes as such under other dialogical conditions, it fails (*tout court*) to provide a comprehensive justification of the policy.

A salient way that arguments fail, when put to this test, and the only mode of test failure that will henceforth figure in these lectures, is that the speaker cannot fulfill a demand for justification that does not arise when the argument is presented by and/or to others. So, to anticipate what I shall try to show, the incentive argument does not serve as a justification of inequality on the lips of the talented rich, because they cannot answer a demand for justification that naturally arises when they present the argument, namely, *why* would you work less hard if income tax were put back up to 60 percent? The rich will find that question difficult no matter who puts it to them, but I shall often focus on the case where their interlocutors are badly off people, because in that setting the question, and the difficulty the rich have with it, may lead to further dialogical development that carries further illumination.

When the justification of policies that mention groups of people is presented in the usual way, with exclusively third-person reference to groups and their members, the propriety of the question why various people are disposed to act as they do is not always apparent. It becomes evident when we picture the relevant

people themselves rehearsing the argument, and sometimes more so when the audience is a strategically selected one. The test of interpersonal presentation makes vivid that the justification of policy characteristically depends on circumstances that are not exogenous with respect to human agency.

And so to community. I began by observing that there is more than one kind of community, and I must now specify the kind that is relevant to present concerns. First, though, a few points about the semantics of the word “community.”

Like “friendship,” “community” functions both as a count noun and as a mass noun. It is a count noun when it denotes sets of people variously bound or connected (the European community, London’s Italian community, our community) and it is a mass noun when we speak of how much community there is in a certain society, when we say that some action enhances or reduces, or some attitude honors or violates, community, and so on.

A community, one could say, is a set of people among whom there is community: that is how the count-notion and the mass-notion are linked. “Community” is in this respect like “friendship”: a friendship is a relationship in which friendship obtains. Notice that friends can do and feel things that are inconsistent with friendship without thereby dissolving their friendship. There can be a lapse of friendship in a friendship without that friendship ceasing to be. But there cannot (enduringly) be *no* friendship in a friendship. And all that is also true of community: there can be violations and lapses of community in a community, but there cannot be no community in a community.

In addition to community in the adjectivally unqualified sense where it is analogous not only in form but also in content to friendship, there are specific types of community, some of which do, while others do not, contribute to community in the just denoted sense. And types of community (mass-wise) distinguish types of community (count-wise). Linguistic community, or community of language, constitutes a linguistic community as such;

community of nationality establishes a national community; and community of interest in stamps binds the philatelic community.

The form of community that concerns me here, which I shall call *justificatory community*, prevails in justificatory communities. And justificatory community, though something of a concocted notion, contributes to community *tout court*, that is, to community in the full (adjectivally unqualified) sense sketched a moment ago. A justificatory community is a set of people among whom there prevails a norm (which need not always be satisfied) of comprehensive justification. If what certain people are disposed to do when a policy is in force is part of the justification of that policy, it is considered appropriate to ask them to justify the relevant behavior, and it detracts from justificatory community when they cannot do so. It follows that an argument for a policy satisfies the requirement of justificatory community, with respect to the people it mentions, only if it passes the interpersonal test. And if all arguments for the policy fail that test, then the policy itself evinces lack of justificatory community, whatever else might nevertheless be said in its favor.

Now, an argument fails the interpersonal test, and is therefore inconsistent with community, if relevant agents *could* not justify the behavior the argument ascribes to them. What if the agents are actually asked to justify their stance and, for one reason or another, they refuse to do so? Then the argument in question does not necessarily fail the test, for it might be that they could justify their stance. But if their reason for refusing to justify it is that they do not think themselves accountable to their interrogators, that they do not think that they *need* provide a justification, then they are forswearing community with the rest of us in respect of the policy issue in question. They are asking us to treat them like a set of Martians in the light of whose predictable aggressive, or even benign, behavior it is wise for us to take certain steps, but whom we should not expect to engage in justificatory dialogue.

To employ the interpersonal test and to regard its failure as indicative of a lack of community is to presuppose nothing about which particular collections of people constitute communities in the relevant sense. Some may think that there is no reason why there should be community between rich and poor in a society, and they may therefore regard failure of the test as uninteresting, or, if interesting, then not because it shows lapse of community. Others, by contrast, might think that community ought to obtain among all human beings, so that it would stain a policy argument advanced by rich countries in North-South dialogue if it could not pass muster in explicit I-thou form.<sup>14</sup> The thesis associated with the interpersonal test is that, if a policy justification fails it, then anyone proposing that justification in effect represents the people it mentions as *pro tanto* out of community with one another. Whether they should be in community with one another is a separate question. That depends on a doctrine, not to be articulated here, about what the proper boundaries of a community are. In my own (here undefended) view, it diminishes the democratic character of a society if it is not a community in the present sense, since we do not make policy *together* if we make it in the light of what some of us do that cannot be justified to others.

It is often said that it is unrealistic to expect a modern society to be a community, and it is no doubt inconceivable that there should be a standing disposition of warm mutual identification between any pair of citizens in a large and heterogeneous polity. But community here is not some soggy mega-*Gemeinschaftlichkeit*. Instead, my claim about the incentive justification is that, to appropriate a phrase of Rawls's, it does not supply "a public basis

<sup>14</sup>In *Justice as Fairness* (p. 152 n. 28) Rawls expresses a view which has a bearing on how wide community can be: "the allegiance to, or the motivational support needed, for the difference principle to be effective presupposes a degree of homogeneity among peoples and a sense of social cohesion and closeness that cannot be expected in a society of states." This implies that there is sufficient such closeness domestically. (Three further contrasts between the single- and multi-society cases that Rawls sketches in the footnote seem to me to fail, but none of them matter here.)

in the light of which citizens can justify to one another their common institutions” and that the justification is therefore incompatible with what Rawls calls “ties of civic friendship.”<sup>15</sup>

Now some examples of the battery of concepts introduced above.

Under the premiership of Harold Wilson, some economic policies were justified by reference to the intentions of the so-called “gnomes of Zurich,” the international bankers who, it was said, would react punitively to various government decisions. It was a mark of their *foreign* status that economic policy had to *placate* those bankers, and although it might have been thought that they should behave differently, it would not have been considered appropriate for the British government to call upon them to do so. But such a call would surely be appropriate in the case of people conceived as belonging to our own community. Nor should members of our own community need to be *placated* by our community’s policies: when justified, their demands should be satisfied, but that is a different matter.

An example that for some readers may be close to home: the policy argument that rates of pay to British academics should be raised, since otherwise they will succumb to the lure of high foreign salaries. We can suppose that academics are indeed disposed to leave the country because of current salary levels. The issue of whether, nevertheless, they should emigrate is pertinent to the policy argument when they are regarded as fellow members of community who owe the rest a justification for decisions that affect the welfare of the country. And many British academics with an inclination to leave who put the stated policy argument contrive to avoid that issue by casting the minor premise of the argument in the third person. They say: “Academics will go abroad,” not: “We’ll go abroad.”

The connection between sharing community membership and being open to requests for justification comes out nicely in an ex-

<sup>15</sup> John Rawls, “Kantian Constructivism in Moral Theory,” *Journal of Philosophy* 77, no. 9 (September 1980): 561; Rawls, *A Theory of Justice*, p. 536.

ample of current interest. The Moscow generals might address the Lithuanian independence movement leaders as follows: "Widespread bloodshed is to be avoided. If you persist in your drive for independence, we shall intervene forcefully, and there will be widespread bloodshed as a result. You should therefore abandon your drive for independence." The Lithuanian leaders might now ask the generals to justify their conditional intention to intervene forcefully. If the generals brush that question aside, they forswear justificatory community with the Lithuanians.

The Lithuanian leaders might produce a parallel argument: "Widespread bloodshed is to be avoided. If you intervene forcefully, we shall nevertheless persist in our drive for independence, and there will be widespread bloodshed as a result. You should therefore abandon your plan to intervene forcefully." And the Lithuanians, too, might feel no obligation to justify their intentions to the generals. If, on the other hand, both sides labor under such a sense of obligation, they will enter a justificatory exchange in which each tries to show that the other's minor premise, whether true or not, should be false.

## 6.

The interpersonal test focuses on an utterance of an argument, but what it tests, through examination of that utterance, is the argument itself. If lack of community is displayed when the rich present the incentive argument, then the argument itself (irrespective of who affirms it) represents relations between rich and poor as at variance with community. It follows, if I am right, that the incentive argument can justify inequality only in a society where interpersonal relations lack a communal character, in the specified sense.

Sometimes, as, for example, in the kidnapper case, the interpersonal test will be a roundabout way of proving an already evident point (in the kidnapper case, that there is significant lack of community between the kidnapper and the parents). But in other

cases the test will illuminate, and I believe that the incentive argument is one of them. The argument is generally presented in thoroughly third-personal terms and, relatedly, as though no question arises about the attitudes and choices of the rich people it mentions. When, by contrast, we imagine a talented rich person himself affirming the argument, then background issues of equality and obligation come clearly into view, and, if I am right, the rich are revealed to be out of community with the poor in respect of the economic dimension of their lives. So we see more deeply into the character of the incentive argument when we cast it in the selected I-thou terms.

Now, an important qualification. I say that the incentive argument shows itself to be repugnant to community when it is offered *on its own* by well-off people. I insert that phrase because the present case against the argument lapses when the argument appears in combination with claims about desert, and/or with Nozick-like claims about a person's entitlements to the reward his or her labor would command on an unfettered market. I do not myself accept that sort of compound justification of incentive inequality, but I do not here contend that it fails the interpersonal test. My target here is the unadorned or naked use of the incentive justification. It is often used nakedly, and with plenty of emphasis that it is being used nakedly. That emphasis occurs when advocates say it is an advantageous feature of the incentive justification that it employs no controversial moral premises about desert or entitlement. (Notice that, since John Rawls rejects use of desert and entitlement to justify inequalities, the Rawlsian endorsement of incentives takes what I call a naked form.)

The sequence of claims that I make goes as follows: The talented rich cannot justify the fact that the minor premise of the (naked) incentive argument is true. If they cannot justify the truth of its minor premise, then they cannot use the argument as a justification of inequality. If they cannot use it as a justification of inequality, then it cannot be used as a justification within com-

munity. If it cannot be used as a justification within community, then anyone who uses it (in effect) represents society as at variance with community when he does so.

## II. TESTING THE INCENTIVE ARGUMENT

### 7.

The kidnapper argument discredits its advocate when the kidnapper puts it forward himself because, as I said (see sec. 4), he *makes* it true that the parent gets their child back only if they pay, and to make that true is morally vile.

Accordingly, to discredit first-person affirmation of the incentive argument in a parallel way, I must defend two claims. First, that in a sufficiently similar sense, the rich *make* it true that they will not work as hard at 60 percent tax as they do at 40 percent: I have to show that the minor premise<sup>16</sup> of the incentive argument owes its truth to their decisions and intentions. (I say *sufficiently* similar, because there undoubtedly are some significant differences here, consequent on the fact that the rich are not an individual but a group, and a group with shifting membership: at the end of this section I address some of the complication which that fact generates.) And it also needs to be shown that, deprived as they here are of recourse to the considerations of desert and entitlement that are set aside in a naked (see sec. 6) use of the incentive argument, the rich cannot justify making the stated proposition true. I am not, of course, obliged to maintain, even then, that their making it true puts them on a moral par with kidnappers, but just that, if their posture is defensible, then its defense rests on grounds of the sort that a naked user of the incentive argument forgoes.

I turn to my first task, which is to show that the talented rich do make the factual premise of the argument true. Let us ask: if

<sup>16</sup> Or, strictly, part *a* of that premise: part *b* is true only if others — for example, the government — act in certain required ways. But for simplicity I shall continue to speak of the rich making the factual minor premise (*tout court*) true.

that premise is true, then why is it true? Is it true because the rich are *unable* to work at 60 percent as hard as they do at 40? Or is it true because they are *unwilling* to work that hard at 60 percent? If the truth of the premise reflects inability, then we cannot say that, in the relevant sense, the rich *make* the premise true. An inability explanation of the truth of the premise means that the rich could not, by choosing differently, make the premise false.

There are two forms that an inability claim might take. In the first form of the claim, the rich cannot work hard unless they consume things that cost a great deal of money.

Now, it might well be true that without enough money to buy superior relaxation some high-talent performances would be impossible: perhaps the massively self-driving executive does need, to be effective, more expensive leisure between one day's work and the next than he can get living in ordinary accommodation on an average wage. (When I say that he might *need* high-quality leisure, I refer not to his preference ordering or utility function but to what it is physically and/or psychologically possible for him to do. That kind of capacity limitation interacts causally with a person's utility function, but it is not identical with it or an aspect of it. But the income gap which that consideration would justify is surely only a fraction of the one that obtains even at 60 percent top tax. The extra money which executives (and so forth) get at 40 percent can hardly be required to finance whatever luxuries we might imagine that they strictly *need* to perform at a high level: they could afford those necessary luxuries with what they have left even when they pay at 60 percent tax.

In a different version of the claim that the rich could not work as hard at 60 percent tax as they do at 40 percent, what they are said to need is not the goods that only a lot of money will buy but the *prospect* of getting those goods or that money: the high reward is now said to be indispensable to *motivation*, or morale. (You eventually give the biscuit to the performing dog so that the same procedure will work again next time, and not because the

dog needs the calories it gets from the biscuit to enable it to go on performing.) This motivation story does not say that, unless they are handsomely paid, the rich will *choose* not to work very hard: the proposition that they have a real choice in the matter is just what the inability claim is designed to contradict. What is rather meant is that the allure of big bucks sustains, and is needed to sustain, the motivational drive required for heavy effort: the rich just cannot *get* themselves to work as hard when they expect to be taxed at 60 percent as they can get themselves to work when they expect to be taxed at 40 percent.

Now, in my opinion, there is not much truth in this contention: it represents people of talent as more feeble than, on the whole, they are. It is not likely to be lack of power to do otherwise that causes the rich to take longer holidays, to knock off at five instead of at six, or not to bother trying to get one more order, those being the things that they do when the income tax rises, if the minor premise of the incentive argument is true. The tax rise means that the rich face a new and less appealing schedule of the costs and benefits of alternative courses of action, and they will, of course, find it harder to raise up enthusiasm for choices that now promise smaller rewards. It does not follow that they cannot make, and effectively pursue, those choices.

Still, I say that there is not much, not no, truth in the contention mooted here. For I recognize that a perception that reward is "too low" can cause, at least somewhat independently of the will, a morose reluctance which operates as a drag on performance. But we should ask what brings about that disabling perception. And if two of its prominent causes were its only causes, then, as I shall now try to explain, the "motivation" version of the inability contention would be disqualified.

One thing that causes a dispiriting feeling that reward is too low is disappointed expectation. Socialized as they have been in a severely unequal society, the talented rich of course anticipate a handsome return for their exertions. They will therefore be down-

cast when such return is not forthcoming, even when they do not judge that they deserve or are otherwise entitled to it. But it is not unlikely that they also do make judgments like that. They think that they have a right to golden rewards if they work hard, and so powerful is that belief that it can act as a further cause of low morale: it can make the thought of working hard at 60 percent tax fill them with a truly disabling dismay.

Now, an inability to work hard at 60 percent tax (in people who, *ex hypothesi*, routinely work that hard at 40 percent) that reflects habituated expectation, or judgment of entitlement, or both, cannot count here, in rebuttal of the claim that optional decisions of the talented rich make the minor premise of the incentive argument true. Consider, first, the habituation factor. We are here engaged in a ground-level investigation of a certain justification of inequality. It is therefore inappropriate, by way of contribution to that justification, to cite mere habituation to unequal rewards. Habits can change,<sup>17</sup> and they are therefore beside the point in a fundamental inquiry. And the causal force of belief in the rightness of high reward (which helps to sustain the habitual expectation) must also be ignored here. For we are here envisaging the talented rich uttering the incentive argument in its naked form, in which invocation of entitlement is pointedly eschewed. There would, accordingly, be a kind of pragmatic inconsistency if the rich had to cite their own belief in entitlement when rejecting the claim that the truth of the minor premise of the argument

<sup>17</sup> If not always at the level of the individual, then certainly at the social level, through reformed structures of education. And even if the relevant habits could not change, that would have more implications for the practice than for the theory of justice. As Rawls says, "We do not consider the strains of commitment that might result from some people having to move from a favored position in an unjust society to a less favored position (either absolutely or relatively or both) in this just society. . . . The strains of commitment test applied to cases of hypothetical transition from unjust societies is irrelevant" ("Reply to Alexander and Musgrave," *Quarterly Journal of Economics*, 1974, p. 653, and see Rawls, *Justice as Fairness*, p. 44, on the role of education in sustaining a just society: the relevant strains of commitment are those that survive a socialization process that instills egalitarian principles in the young).

reflects what they are themselves willing and unwilling to do.<sup>18</sup>

If the “motivation” variant of the inability claim depended entirely on habit and normative belief, we could safely set it aside. We could say that if it is true, it is compromised in the present context by what its truth rests on, that it does not furnish an appropriate reason for saying that talented rich people could not work as hard at 60 percent tax as they do at 40. The claim might help to silence moralistic charges against the present generation of talented rich people, but it could not contribute to a robust vindication of inequality in human society.

Now I firmly believe that such truth as the inability claim possesses does depend, entirely, on factors of habit and ideology that, for the stated reasons, must here be ruled out. I think it hard to believe otherwise, when one focuses on the inability claim proper, as opposed to the claim, with which it is readily confused, that the talented rich have a *right* not to work as hard at 60 percent tax as they do at 40 percent. Nevertheless, I have not shown that there exist no relevant deeper restrictions on motivation, and, in the seminar following these lectures, I was rightly taken to task on this score by Samuel Scheffler, who did not reject my conclusion, but who emphasized that it had not been demonstrated, not, at any rate, in the general case, where the issue is not whether these particular people could keep their shoulders to the wheel under the contemplated tax rise, but whether *some* significant inequalities are required, in general, for optimal economic motivation.

For all that I had shown, so Scheffler said, incentives might elicit motives that could not “be summoned at will,” that nothing

<sup>18</sup> That particular inconsistency would not attach to naked use of the argument by a third party who cites (without endorsing) the belief of the rich in their entitlements as what happens to explain the truth of the argument’s minor premise. But reference to that belief would nevertheless be unacceptable when the argument for inequality is pitched at a fundamental level. If the rich are unable to work as hard at 60 percent tax as they do at 40 because they believe that they should be paid more if they work harder, then the stated incapacity cannot, without bizarre circularity, figure in an argument which would justify the proposition that it is *fundamentally right* that they be paid more for working harder.

else would induce, and that would enable agents to perform better than they otherwise could. To illustrate the form of his objection, he cited the “runner who needs competition to achieve his fastest times,” and people who work best under the pressure, adversity, or challenge. The compelling examples warn against being simple-minded about psychological feasibility. They show that what people are able to do depends on the reasons they have for doing it: with different reasons, the adrenalin flows to different extents. And Scheffler concluded that a fully adequate reply to the inability claim would have to include “at least the rudiments of a serious psychology of egalitarianism . . . a realistic account of the human motivational resources and mechanisms that egalitarian . . . institutions would expect to engage.”

I accept this criticism, which calls for a program of work that manifestly cannot be accomplished here. It needs to be shown that a society of people who believe in equality and act accordingly is reproducible, that it is not fated to collapse under disintegrative strains. Such societies seem to be possible on a small scale, and we need to explore what constraints of human nature and organization make them difficult—as they undoubtedly are—on a larger scale, and whether those difficulties approach impossibility. As a practical proposal, normative egalitarianism indeed requires a corresponding psychology. If the research program to which the Scheffler objection points were to deliver negative results, equality might still be a tenable value, but it could not, unmodified, represent a policy goal.

In pursuing such a program, in the search for possible equality-supporting “human motivational resources and mechanisms,” it is not inappropriate to reflect on the other (nonincentive) examples that illustrate the form of Scheffler’s objection. For they all involve a drive to perform well, whether as an end in itself or as a way of impressing others, and/or oneself. The motivation in question contrasts with the search for gains which, like money, are quite external to the performance itself. The examples remind us that the

desire to achieve, to shine, and, yes, to outshine, can elicit enormous effort even in the absence of pecuniary motivation. Of course, many would say that such nonpecuniary mechanisms just replace money inequality by status inequality, and that is yet another large challenge to which I cannot respond fully here.<sup>19</sup> Notice, though, that the notion of “replacement” is somewhat unapt, since money inequality itself generates status inequality. Status is not, moreover, redistributable in the same way that material resources are, and it therefore does not raise the same issues as money inequality does for an egalitarianism whose inspiration (see sec. 1 above) is that some people lead unnecessarily hard lives.

And there is another consideration to be borne in mind: in estimating what it would be like for a person to accept a salary that is much lower than what full exercise of market power would provide, the strain to think about is the one he would feel when, *ex hypothesi*, people like him are accepting similarly modest salaries. We are talking about an egalitarian society, not about a population of talented people each of whom is a unique moral hero.

That is the best I can do, right now, by way of facing up to the prodigious task Scheffler set me. So, realizing that some of the required case has not been proven, I nevertheless now set the motivation claim aside, and, with all the relevant implied *caveats*, I conclude that the reason why the minor premise of the incentive argument is true (if it is true) is that the executive and his like are *willing* to work hard only at a 40 percent top tax rate.

But, before we ask whether that choice is justified, let me address the complication that, even if each talented individual chooses not to work hard at a 60 percent tax, no such individual makes the minor premise of the incentive argument true, since its truth requires that many such individuals make similar choices.

<sup>19</sup>For an ingenious attempt to meet it, see Joseph Carens, *Equality, Moral Incentives, and the Market* (Chicago: University of Chicago Press, 1981). There are substantial flaws in Carens's book, but it is, in my view, a profound and pioneering work.

Here, then, is a disanalogy with the case of the kidnapper, since he makes the minor premise of his argument true all by himself.

In response to this important point, I shall say only two things here. First, notice that an individual talented rich person is relevantly analogous to a member of a large band of kidnappers, who could also truthfully say: it will make no, or not much, difference if I change my choice. Yet, if a member of such a band puts the kidnapper argument in the first-person plural, if he says, "Giving us the money is the only way you will get your child back," then the fact that he is only (a dispensable) one of the "us" who together ensure that the child is held captive does not make his posture justifiable. And it is similarly true that if what the rich together cause could not be justified if one rich person caused it, then being only one rich person and not all of them would not suffice to make one's behavior justifiable. One might not be *as* responsible as when one achieves something without assistance, but one also could not say that the result had nothing to do with one's actions.<sup>20</sup>

And whatever the complex truth may be about individual responsibility for a collectively produced result, I am not here primarily interested in commenting on the moral character of rich people. My primary interest is in an argument which, I claim, fails the interpersonal test. Rich people may benefit from a practice on which they have little occasion to reflect. If we here (counterfactually) imagine them trying to justify that practice by recourse to the incentive argument, it is in order to investigate not, in the first instance, how blameworthy they are, but how that argument fares in the light of a norm of justificatory community.

<sup>20</sup> For a case which bears on the issue dealt with in the foregoing paragraph, see Derek Parfit's "harmless torturers" at p. 80 of his *Reasons and Persons* (Oxford: Clarendon Press, 1984). If someone objects that the talented rich are unlike the just-imagined kidnappers in not being an organized group, then, so I believe, reflection on Parfit's case shows that they need not be one for my purposes. And one could also put forward a persuasive case of relatively unorganized kidnappers, where all that is essential to the analogy is restored, but I shall spare you the rococo detail.

## 8.

In its standard presentation, the incentive argument is put forward as though it is irrelevant to its assessment whether the rich are justified in making its minor premise true, and as though it would be inappropriate to put that question to them. I have protested that the question can be considered inappropriate only if the rich are conceived as inaccessible third persons who do not belong to the society for which the incentive policy is proposed. It does not follow that what the rich do could not be justified, that the neglected question, having been raised, could not be answered satisfactorily. In this section I explore possible answers to it.

The relevant part of the premise (that is, part *a*) says that, if the top tax rises to 60 percent, the talented rich will work less hard than they do now, when the top tax is 40 percent. And, so we have concluded, that is because they will then *choose* to work less hard. As a result of that choice, the badly off will be worse off than they were before (by the truth of part *b* of the minor premise of the incentive argument), and, a fortiori, worse off than they would be if the talented rich maintained at 60 percent tax the effort they put in at 40 percent. On the factual assumptions behind the minor premise of the argument, the ordering of benefit to the badly off from the three work/tax packages just mentioned is as follows:

- 1 The talented rich work  $w$  at 60 percent tax
- 2 The talented rich work  $w$  at 40 percent tax
- 3 The talented rich work  $w-x$  at 60 percent tax,

where  $w$  is the amount the rich choose to work at 40 percent and  $x$  the amount by which they reduce their input if the tax rises to 60 percent.

We must now ask whether the choices of rich people, which make 3 rather than 1 true if the tax rises, and thereby make the badly off worse off than when the tax is low, can be justified, when

notions of desert and entitlement are not allowed to figure in justifications.

In certain cases, where working just as hard at 60 percent tax as one did at 40 percent would mean an oppressive existence, the choice that the rich make is undoubtedly justified. Think of those harried and haggard Yuppies, or overworked surgeons, who really would lead miserable lives if the massive amount of work that they do were not compensated by the massive amount of income that leads them to choose to work that hard. We can set such "special burden" cases aside, not because they do not exist, but because of the nature of the justification of the talented rich person's choice in this sort of case.

Let me explain. In the present exercise, the incentive argument is supposed to justify inequality. But when special burden is invoked, what we get is not a justification of an inequality, all things considered, that incentives produce, but a denial that they do produce an inequality, all things considered. That is so because, when we compare people's material situations, we must take into account not only the income they get but also what they have to do to get it. Accordingly, if the talented rich could plausibly claim special burden, the move to the 40 percent tax which induced them to work harder might also be required for the sake of equality: where work is specially arduous, or stressful, higher remuneration is a counterbalancing equalizer, on a sensible view of how to judge whether or not things are equal. Since I oppose only those incentives that induce unambiguous inequalities, my opposition retires in face of the special burden case, and I acknowledge that, where special burden holds, the rich have a persuasive answer to the question why they make the minor premise of the incentive argument true.

My primary target, as a philosopher, is a pattern of justification, from which the incentive argument deviates when special burden holds. But, as a politically engaged person, I also have another target: the real-world inequality that is actually defended

on incentive grounds. And because I also have that second target, I have to claim that the special burden case is statistically uncommon. But I do not find that difficult to do, since I am confident that, if talented rich people were to provide, at 60 percent tax, the greater effort we are supposing them to supply at 40 percent, then a large majority of them would still have not only higher incomes but also more fulfilling jobs than ordinary people enjoy.<sup>21</sup>

Since I propose to cast no doubt on the truth of the minor premise of the incentive argument, I must now set aside another case, that in which well-paid talented people so enjoy their work or are so dedicated to making money that they would actually work no less hard after a tax rise. Such people are bluffing if, in the hope of inducing a political effect, they announce that a tax rise would lead them to work less. But in their case, and, a fortiori, in the case of talented people whose labor supply curve is in the relevant range not merely vertical but backward-bending, the minor premise of the incentive argument is false, since these people will *not* work less hard if the tax goes up, and this case is therefore out of bounds here.

Summarizing and extending the foregoing discussion, I now ask you to look at a table that depicts three positions that the talented rich person might be thought to be in. Of the three cases that appear in the table, two are, for different reasons, irrelevant to our purposes, the special burden case because it poses no problem for the egalitarian point of view (and is in any event not

<sup>21</sup> Anyone who dissents from that statistical assessment is invited to settle for the following more modest claim, which will suffice here: although it is difficult to tell how much any given individual enjoys or disenjoys his work, it is false that jobs demanding talent are, on the whole, less satisfying. Accordingly, the consideration of burden cannot justify the fact that on the whole they command much more pay.

It is an important point, for Rawls, that the talented are fortunate to be talented, and that is partly because the exercise of talent in work is satisfying. Accordingly, Rawlsians are not well placed to adduce the special burden consideration in support of the justice of incentives. As Robert Nozick remarks, "Rawls is *not* imagining that inequalities are needed to fill positions that everyone can do equally well, or that the most drudgery-filled positions that require the least skill will command the highest income" (*Anarchy, State, and Utopia* [New York: Basic Books, 1974], p. 188).

widely instantiated), and the case of bluff, because in that case the minor premise of the incentive argument is false. So, from now on, let us focus on what is called the standard case in the table.

In the table,  $w$  denotes the amount which the rich actually work at 40 percent, and  $w-x$  denotes some significantly smaller amount. In all three case, the rich prefer working  $w$  at 40 percent to working  $w-x$  at 60 percent. This preference may not be readily apparent, but we can demonstrate<sup>22</sup> that they have it. For they choose to work  $w$ , rather than  $w-x$ , when the tax is 40 percent,

Benefit to the (currently) badly off	Preference orderings of the rich across three work/tax packages
<i>The standard case</i>	
2	Work $w$ at 40%
3	Work $w-x$ at 60%
1	Work $w$ at 60% (and be much better off than others are)
<i>The bluff case</i>	
2	Work $w$ at 40%
1	Work $w$ at 60%
3	Work $w-x$ at 60% (and be much better off than others are)
<i>The special burden case</i>	
2	Work $w$ at 40%
3	Work $w-x$ at 60%
1	Work $w$ at 60% (and be worse off than others are as a result)

<sup>22</sup> On the usual economists' assumptions, which are innocent here, that choice tracks preference, and that wide choice is preferred to narrow.

and they must prefer  $w-x$  at 40 percent to  $w-x$  at 60, since work is the same and income is higher in the first package. It follows that the rich prefer working harder at 40 percent to working less hard at 60.

The preference orderings of the rich are identical in the standard and special burden cases. The difference between those cases (which is formulated in parentheses) lies in the comparison between the lot of the rich and that of other people when the rich are at the bottom of that preference ordering. This comparison reflects both income level and quality of work experience: were they to work as hard at 60 percent as they do at 40, the rich would in the special burden case be worse off than others are, but in the standard case they would still be much better off than others are. The ordering of benefit to the badly off from the various work-tax packages (which is given by the numbers in the column on the left, and which is the same in all three cases) is based on the assumption that part *b* of the minor premise of the incentive argument is true (so “ $w$  at 40 percent” ranks above “ $w-x$  at 60 percent”) and on the further assumption that, if the rich worked as hard at 60 as they do at 40, then that would bring still further benefit to the poor (so “ $w$  at 60 percent” ranks above “ $w$  at 40 percent”).

The interpersonal test has talented rich people *themselves* uttering the incentive argument. Now, for present purposes, the talented rich do not fall under the bluff case, in which the minor premise is false: they really will work less if the tax goes up. And, if we follow a distinction that has found favor with philosophers, the rich do not *threaten* anything if they utter the incentive argument, since, in the recommended distinction, you merely *warn* that you will do *A* when you are bent on doing *A* independently of the leverage you get from saying that you will do it. Notice that, in the recommended distinction, a kidnapper who likes children merely *warns* if he would actually prefer (for nonstrategic reasons) to keep the child if he is not paid: this shows that, under

the recommended distinction, nonthreatening warnings can be very unpleasant.

So imagine, now, a set of highly paid managers and professionals addressing poorly paid workers, unemployed people, and people indigent for various personal and situational reasons, who depend on state welfare. The managers are lobbying against a rise in tax from 40 to 60 percent, and this is what they say:

Public policy should make the worst off people (in this case, as it happens, you) better off.

If the top tax goes up to 60 percent, we shall work less hard, and, as a result, the position of the poor (your position) will be worse.

So, the top tax on our income should not be raised to 60 percent.

Although these argument-uttering rich may not, for one or other reason, count as *threatening* the poor, they remain people of superior income and form of life who could continue to work as now if the tax rose to 60 percent, and thereby bring more benefit to the poor, while still being much better off than they are, but who would refuse to do that. They say, in effect: we are unwilling to do what we could do to make you better off and yet still be much better off, ourselves, than you are. We realize that, at the present level of fuel allowance, many of you will be very cold this winter.<sup>23</sup> If the tax went up to 60 percent and we worked no less hard in response, revenue for fuel expenditure could rise, and some of you would be more comfortable. But in fact we would work less, and you would be worse off, following such a tax rise.

Having presented their argument, the rich are not well placed to answer a poor person who asks: "Given that you would still be much better off than we are if you worked as you do now at the 60 percent tax, what justifies your intention to work less if the tax

<sup>23</sup>According to Robin Cook, MP, Labour spokesman on health, in the severe winter of 1991 there were 4,000 more deaths of old people than are usual in such a period.

rises to that level?" For these rich people do not say that they deserve a lot because of their prodigious effort, or merit more because of their higher contribution to production. There is in their approach no appeal to such controversial moral premises, and many of them would think that, being free of such premises, their argument is consequently less vulnerable. And they cannot respond by saying that the money inequality which they defend is necessary to make the poor better off, since it is they who make it necessary, and the question put by the poor asks, in effect, what their justification is for making it necessary.

The incentive argument does furnish the poor with a reason to accept the inequality that it recommends. For the poor can take it as given that the rich are determined to sustain the intentions that make the argument work. But the argument cannot operate like that for the rich themselves: since they cannot treat their own choices as objective data, they cannot take it as given that the minor premise of the argument is true. Correspondingly, and unlike the poor, they need a justification not for accepting but for imposing the inequality that the argument defends.

But it might be said that the rich can indeed respond convincingly to the poor, and without advancing the controversial claims about desert and entitlement that are here ruled out. They can say: "Look, it simply would not be worth our while to work that hard if the tax rate were any higher, and if you were in our shoes you would feel the same way."<sup>24</sup> Would that not be a good answer to the question the poor pose?

As I shall presently allow, there is some power in this answer. But its rhetorical cast makes it seem more powerful than it is.

Notice, to begin with, that the first part ("Look . . . higher") of the quoted plea has no independent interest, no interest, that is, which is independent of the associated claim that the poor, if

<sup>24</sup> This piece of dialogue comes from Samuel Scheffler's seminar commentary on these lectures. Scheffler pressed the challenge to which the rest of this section is a response.

better placed, would feel (and act) as the rich now do. For it is a presupposition of the challenge the poor put to the rich that the latter do prefer, and intend, to work less hard if the tax goes up, and in speaking of what is “worth their while” the rich can only be reminding the poor of those preferences and intentions: they cannot mean, for example, that they are paid nothing, or paid badly, if they work hard at a 60 percent rate of tax.

So the burden of the rhetorically presented justificatory move is that a typical poor person would behave just as the rich, on the whole, do. But there is something that the poor person can say in reply. He can say: “Neither of us really knows how I would behave. Not all rich people market maximize as a matter of course, and I hope that, if I were rich, I would not belong to the vast majority that do, especially if I retained a lively sense of what it is like to be in the condition I am now actually in.” (A slave need not be impressed when a master says: “Had you been born into the slaveholder class, you too would have lived well and treated your slaves like slaves.” Such counterfactual predictions do not show that what people at a certain social level typically choose to do is justifiable.<sup>25</sup>

Suppose, now, that the rich abandon the vivid but problematic “you’d do the same in my shoes” style of justification. Suppose they just say (this being the content of the text to note 24, without its rhetorical cast) that, even when desert and entitlement are set aside, only an extreme moral rigorist could deny that *every person has a right to pursue self-interest to some reasonable extent* (even when that makes things worse than they need be for badly off people).

I do not wish to reject the italicized principle, which affirms what Samuel Scheffler has called an “agent-centered prerogative.”<sup>26</sup> But a modest right of self-interest seems insufficient to

<sup>25</sup> I have always thought that the right reply to a white South African who says, to an anti-Apartheid advocate, “You would see things differently if you were in my position,” is: “Quite: I’m sure it does blind one’s vision.”

<sup>26</sup> See his *Rejection of Consequentialism* (Oxford: Clarendon Press, 1982).

justify the range of inequality, the extremes of wealth and poverty, that actually obtain in the society under discussion. Entitlement or desert might justify vast differences between rich and poor: no limit to the inequality they might endorse is inscribed in them. This is particularly clear in the case of the entitlement principle that I am absolute owner of my own labor power. When my power to produce is conceived as fully private property, I may do with it as I will and demand what I may for its use. A proportionately greater attention to one's own interest, as opposed to that of others, is more limited in its justificatory reach, and it seems unlikely to justify the existing contrast of luxury and want.

Now, it might be objected that, in characterizing the position of the less well off as one of deprivation or want, I am unfairly tilting the balance against the incentive argument. To such an objection I have three replies.

First, I am in this part concerned with a real political use of the incentive argument. Reference to real circumstances is therefore entirely appropriate.

Second, the incentive argument is quite general. It should therefore apply no matter how badly off the badly off are, both absolutely and relatively to the well off. Accordingly, it is methodologically proper to focus on particularly dramatic cases of its application.

And it is precisely when the condition of the badly off is especially wretched that the *major* premise of the incentive argument can pass as compelling. Where the worst off are not too badly off, it looks more fanatical to assign absolute priority to their claims. But the stronger the case for ameliorating the situation of the badly off is, the more discreditable (if I am right) the incentive argument is on the lips of the rich. So the argument is most shameful where, at first sight, it is most apt.

Now, a world that implements John Rawls's *two* principles of justice will not display the degree of inequality that characterizes contemporary Britain. Accordingly, the foregoing attempt to neu-

tralize the agent-prerogative defense of the incentive argument in its common use will not serve to defeat it as a defense of the Rawlsian use of the argument, and more will be said about that later.

## 9.

The resolve of talented rich people to produce less if the tax rises makes the factual premise of the incentive argument true and ensures that the poor are poorer than they otherwise would be. I have argued that, within the restriction of naked use of the incentive argument, the rich cannot justify making its factual premise true. There is consequently an impression of incoherence when *they* employ the argument in defense of low taxes on top salaries: for the more disposed they are to affirm its normative premise, the less disposed they should be to make its factual premise true.<sup>27</sup> The argument stands up only because the agents mentioned in its minor premise do not act as one would expect people who put forward its major premise to act. If they did so act, they would not make the minor premise of the argument true, it would then not be true, and the argument would collapse.

For an analogy to the bad faith that comes when the rich themselves propound the argument, think of kidnappers who say that, since the safety of the hostage should be of paramount concern, its loved ones should pay for its release. That structurally similar—and risible—posture is portrayed in the film *Ruthless People*, in which frustrated kidnappers express outrage against the husband

<sup>27</sup> This claim, that affirmation of the major premise of the argument by the talented rich *does not cohere* with their disposition to make its minor premise true, is not identical with (though it is related to) the claim labored above, to wit, that the talented rich *cannot justify* making the minor premise true, when the incentive argument is used nakedly. The incoherence claim depends on what the major premise says, and goes with certain formulations of it only. If the kidnapper changes the major premise of his argument from “children should be with their parents” to “parents should pay to retrieve kidnapped children,” the claim about his minor premise survives, but the incoherence claim goes. Analogous results emerge if the rich employ as major premise: “the poor should vote for whatever enhances their interests.”

of the unwanted wife whom they hold hostage, when they become apprised of that husband's blithe lack of desire to pay for his wife's release. Or think of a crowd of recently munching strollers who complain about the failure of the city's street-cleaning service as they toss their Big Mac containers into the gutter. There is similar incongruity when talented rich people indignantly condemn parties of the Left for a supposed lack of concern for the poor supposedly shown by the Left's policy of taxing the rich heavily. They can say that such parties are stupid, in light of the terms of cooperation to which they are themselves resolved to stick, but, although that may be true, it is not a reason for *them* to display indignation.

## 10.

The incentive rationale might convince the poor that they should vote to maintain low taxes. But it does not show them why the rich have made it true that they might be well advised to vote like that. And the poor might refuse to vote that way. They might support Labour and press for higher taxation on high incomes, and not because they do not accept the minor, factual, premise of the incentive argument. On the contrary: they might believe that the minor premise is true, they might notice that its truth reflects the insistence of the rich on an unusually high standard of life and work, and they might want not to condone that insistence but to resist it, even at the cost of their own material self-interest. They would then reject the major premise of the incentive argument. They would say that inequalities that enhance their own position are not justified when the reason why they enhance it is the one that features in the minor premise of the argument.

The poor have *a* reason to respond as the rich suggest, since if they do they prosper better materially, and they might care enough about that to play ball with the rich. But it would not necessarily be irrational for the poor to reject what the rich suggest and forgo the promised material gain. It is not necessarily irrational (and it

is sometimes felt to be morally imperative) to refuse to deal with a person who wields power in an untoward way even if, should you accede to the proposal he makes, you would be materially better off. That is not necessarily irrational both because how well off you are is not a matter of your material situation alone, and because how well off you are is not the only matter it is rational for you to care about. (That low-income people sometimes care about other things, such as retaining their self-respect and not collaborating with what they think unfair, is shown in their frequent willingness to hold out on strike for higher wages beyond the point where that could be thought rational in income-maximizing terms.)

Still, the indignant poor might, as I said, care enough about prospective extra income to fall in with what the rich propose. They could think: we want to improve our modest lot, so it is entirely reasonable for us to accept the enhanced-incentives proposal. But they could not say to the rich: yes, your proposal is entirely reasonable. If the rich could claim that they *need* extra money to perform better or that without superior pay a superior performance would mean that they live bleak lives, then the poor could accede to the proposal of the rich in the dimension of I-thou interaction. In the case that I am envisaging (that is, in the standard case of the table), resisting their proposal is one way of treating the rich as a set of thous, rather than as a powerful opaque force. If the rich could be regarded as external things, like machines, or bits of nature, it would then be irrational for the poor not to accept their proposal. It is irrational to be angry with a lofty mountain, to think "I'm damned if I'm going to climb you or walk around you to get where I want to go," since, unless you are an animist, your relationship to a mountain properly takes an I-it form. But the poor know that the rich are persons, and they may regard them as fellow members of community who can be asked, face to face, for justification. And then rejection by the

poor of the proposal made by the rich is not necessarily irrational: uncooperative anger is one rational response to what the rich say.

## 11.

The incentive argument is not problematic (in the particular way that I say it is) when it is thought acceptable to view the rich as outside the community to which the poor belong. But sometimes, in Britain, anyway, many of the rich themselves are eager to invoke community, when, for example, they react with (real or fake) horror to militant agitation among the poor. (Maybe some of the rich think that “belong to the same community as” denotes a nonsymmetrical relation.)

Of course, particular talented people can affirm the incentive argument without difficulty, by declaring that they personally lack the disposition attributed to members of their class in the argument. But if the argument is going to pass muster as a justification of unequal reward within community, then putting it forward in the first person, and without such disavowal, should not be problematic.

In the third person, the minor premise of the argument just predicts how the rich will behave, and it can show misunderstanding of the speaker’s message to demand a justification of that behavior: the speaker is not responsible for it, and he might himself be disposed to condemn it. But to affirm the minor premise of the argument with full first-person force is to declare, or, what suffices for present purposes, to manifest, an intention, and a demand for justification is therefore in order. Observe the difference between these two interchanges, each of which follows assertion of the minor premise of the argument to a poor person, in the first case *by* a poor person, or by some third party.

Poor person: But they, the rich, should not demand so much.

Reply: That has nothing to do with me. The fact is that they do.

That is a valid reply to the poor person's lament. But now consider an analogous interchange following a first-person presentation of the premise:

Poor person: But you, the rich, should not demand so much.

Reply: That has nothing to do with me. The fact is that we (I, and the others) do.

Here the very incoherence of the reply confirms the aptness of the challenge against which it strains.

Finding it difficult to provide a convincing reply, the rich may represent their own optional attitudes and decisions *as* given facts. They might say to the poor, "Look, we all have to accept the reality of the situation." Yet it is not an exogenous reality which they are asking the poor to recognize. In this rhetoric of the rich, a declaration of intention masquerades as a description of something beyond choice: the rich present *themselves* in third-personal terms, in alienation from their own agency.<sup>28</sup>

For an analogous self-misrepresentation, consider how absurd it would be for the kidnapper to say: "Gee, I'm sorry, but the fact is that unless you pay I will not release your child." If he says that in factual style, and not as a piece of macabre humor, his remark expresses an estrangement from his own intention which means that he is crazy.

And I believe that there is also something weird going on when the will of a class is depicted by its members as *just* a sociological fact. The rich man sits in his living room, and he explains, in a detached style that says that *his* choices have nothing to do with the matter, why the poor should vote against higher taxes on the rich. Here, too, there is alienation, but, because it is less ob-

<sup>28</sup> This is not a rerun of the inability claim, which we left behind at the end of section 7. That claim acknowledges that the rich form and execute a set of intentions, but denies that they could form and/or execute certain alternative ones. In the motif of alienation, the very fact of intentional agency is concealed, or at least obscured.

vious than the alienation of the single kidnapper that I just portrayed, you do not have to be completely crazy to slip into it. It is easy to slip into *this* alienation because each rich person's individual choice lacks salience, lost as it is among the millions of similar choices typical of members of his class: he participates in a practice so familiar that it gets treated as part of, or on a par with, the course of nature. In a reflective moment he might be appalled by the situation of the badly off, but he reifies the intentions of rich people (his own included), which frustrate their claim to priority, into hard data which social policy must take as parametric. He is unalive to the fact that his own decisions contribute to the condition he describes, a condition which is the upshot of a vast number of personal choices, but which he describes in the impersonal discourse of sociology or economics.

Recall the crazy kidnapper, who says, "Gee, I'm sorry." The child's parents might display a corresponding craziness. They do so if they treat the kidnapper's intention as an objective fact not only for them but even for him. And then they think of his demand as just what they happen to have to pay to get their child back, and maybe one of them says to the kidnapper, as to a possibly sympathetic bystander: "Well, £5000 is a lot of money, as I'm sure you'll agree, but it's less, after all, than what it cost to have Sally's adenoids removed, and, as you've pointed out, it is her *life* that's at stake."

And these reflections also have a bearing on the incentive argument. I have said that the incapacity of that argument to serve as a justification of inequality when the rich present it to the poor shows that the argument presupposes a lack of community between them. And I have just now also said that when the rich deliver it in a certain cast or tone, they imply that they do not qualify as choosing human agents. In considering that second point, it may be instructive to contemplate a presentation of the incentive argument that we have not yet considered, one in which a poor person addresses a set of rich ones. Now the minor premise

will say: if the top tax rises to 60 percent, *you* will work less hard, and we shall consequently be worse off. If the poor speaker says that in an objective tone of voice, his rich listeners might, as a result, feel the weirdness that comes when someone predicts your behavior as though you have no control over it. Some of the listeners might even protest: “Hey, wait a minute. We would like at least to *try* not to work less if the tax rises.” And the poor speaker might counter: “You’re not likely to stick to that resolution. *Please* vote against the tax rise.” In his insistence on the truth of the incentive argument’s minor premise, this poor person would be setting his face against community, or against the capacity for agency of his listeners, or against both.

### III. INCENTIVES AND THE DIFFERENCE PRINCIPLE

#### 12.

I have thus far scrutinized a defense of the inequalities of an actually existing capitalist society (Great Britain) that occurs in ordinary political discourse. I now leave the vernacular context and turn to a text-based examination of John Rawls’s difference principle. It is certain that Rawls would not endorse the particular inequalities that prevail in Britain. But his own defense of inequality has significant elements in common with the case for Lawson’s tax cut, and much of my criticism of the latter also bears against Rawls’s views.

It is usually supposed, and it is evidently supposed by Rawls himself, that his affirmation of the difference principle is consistent with his endorsement of the inequalities that come with special incentives to people of talent. But I shall argue that, when true to itself, Rawlsian justice condemns such incentives, and that no society whose members are *themselves* unambivalently committed to the difference principle need use special incentives to motivate talented producers.

In these lectures I have been concerned to distinguish between inequalities that are necessary, apart from human choice, to make the worst off better off, and inequalities that are necessary to that end only given what some people's intentions are. And this distinction, between, as one might say, intention-relative and intention-independent necessities, generates a question about how we are to take the word "necessary" in John Rawls's difference principle. When he says that inequalities are just if they are necessary to improve the position of the worst off,<sup>29</sup> does he countenance only inequalities that are necessary (to achieve the stated end) apart from people's intentions, or also, and more liberally (in more than one sense of that term), inequalities such as those that are necessary when talented people lack a certain sort of commitment to equality and are set to act accordingly? We confront here 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. So, for example, if an inequality is needed to make the badly off better off but only *given* that talented producers operate as self-interested market maximizers, then that inequality is endorsed by the lax, but not by the strict, reading of the difference principle.

I shall argue that each of these incompatible readings of the principle is nourished by material in Rawls's writings, so that he has, in effect, two positions on the matter. His comments on the spirit in which people in a just society affirm the difference principle point to the strict, "intention-independent" reading of it: that reading goes with his remarks about "full compliance," the dignity of the badly off, and fraternity. Yet, by endorsing incentives, Rawls treats inequalities whose necessity is relative to the intentions of talented people as acceptable to the difference prin-

<sup>29</sup> That is one part of the difference principle. Another part says that inequalities are unjust if they worsen the position of the worst off, and, on the generous interpretation of the principle (see sec. 1 above), a third part says that they are (not un)just if they have no effect on the worst off.

ciple: he proceeds as though he affirms the principle in its lax interpretation.

## 13.

Before turning to Rawls's texts, I want to argue that the strict interpretation of the difference principle is mandatory *if* we suppose that the people in the society in which it is applied are themselves attached to the idea of justice that the principle articulates and are motivated by it in their daily lives. In other words: if we begin with an uninterpreted statement of the principle, where it is ambiguous across strict and lax interpretations, *and* we suppose that all of the people in the society it governs comply wholeheartedly with it, by which I mean that they are concerned to ensure that their own conduct is just in the sense defined by the principles, then what they comply with is the principle in its strict interpretation.

In such a society, the difference principle affects the motivation of citizens in economic life. It controls their expectations about remuneration, that is, what they will regard as acceptable pay for the posts they are invited to fill. It is generally thought that the difference principle would be used by government to modify the effect of choices which are not themselves influenced by the principle, but, so I claim, in a society of wholehearted commitment to the principle, there cannot be so stark a contrast between public and private choice. Instead, citizens want their own economic behavior to satisfy the principle, and they help to sustain a moral climate in which others want the same. I show in the next section that much of what Rawls says commits him to such an understanding of the difference principle, even though his approval of incentives embodies a rejection of that understanding, since approving of incentives means accepting the difference principle in its lax form, and in that form it can be satisfied in a society where it has no direct influence on economic motivation.

Suppose I am a doctor, contemplating a hospital post which I know I could obtain at, say, £40,000 a year. I also believe that,

if — and only if — I took something in the region of £15,000 for filling it, then any difference between my reward and what the less well paid get would be justified by what I strictly need to do the job, or by its special burdens. Then how can I say, with a straight face, that justice forbids inequalities that are detrimental to the badly off *and* be resolved to act justly in my own life, unless, should I indeed go for this particular job, I offer myself at £15,000 and thereby release £25,000 for socially beneficial use?<sup>30</sup>

I might say: “Look, I *am* concerned about the less well off, but I do not have to devote my whole life to them. It is right for government to serve their interests by taxing me, but I should also be allowed to pursue my own *self*-interest, and that is why I feel justified in taking the salary that hospitals have to offer to attract physicians like me.”

But this reply is not sustainable here.

First, notice that I cannot mean the reply in a spirit of apologetic self-criticism, for I am here, *ex hypothesi*, resolved to act justly. Under that hypothesis, I must show that my behavior is not unjust, not that it is an understandable compromise between justice and self-interest. I have to show that the inequalities caused by what I and other professionals choose to do are not unjust, even though they make the lot of the badly off worse than it needs to

<sup>30</sup> People who favor a lax interpretation of the difference principle have suggested that I could say that my giving up the £25,000 would deliver little benefit to any particular person. Yet that need not be so: if it be required that my sacrifice make a palpable difference to some particular person or particular people, then channels which do not fragment its impact could be devised. But the requirement is anyhow misconceived. For one could argue, by the same token, against those who support the difference principle in its lax interpretation, that it is pointless to collect income tax from one person in particular, since that too makes no significant difference to any individual.

And this is anyway not the central issue here. For the appropriate question here is not: what, irrespective of the character of the society in which he finds himself, is the moral obligation of a talented individual who believes in Rawlsian justice? The right question is: what would a *society* that is just by the lights of the difference principle be like? How, among other things, would talented people in *general* behave in such a society? If, as I am claiming, they would in general take jobs for modest salaries, then each could reflect that, together with others, he or she is making a massive difference to (what would otherwise be) badly off people.

be. Consequently, I am claiming that some inequality is just because it reflects legitimate pursuit of self-interest on the part of people with a fortunate endowment of talent. I am saying that justice is itself a compromise or balance between self-interest and the claims of equality.

As I indicated earlier (see sec. 8), I do not aim to impugn the integrity of a conception of justice which allows the agent a certain self-regarding prerogative. But the doctor's reply is meant not merely to articulate a defensible conception of justice, but to reconcile his claim to be wholeheartedly committed to the difference principle's idea of economic justice with his lax reading of that principle. We must ask whether his reply accomplishes that result.

Now, Rawls does not speak of distributive justice as a compromise of the contemplated sort,<sup>31</sup> but our question is whether he might, whether, that is, he could vindicate the lax difference principle along the lines of the doctor's reply. And I do not think that he could, since the reply turns on what is here the wrong distinction. The reply defends assistance to the badly off moderated by the pursuit of self-interest, as opposed to a more total devotion to them. But that is not the same as the distinction between benefiting the badly off by virtue of what I pay in taxes and through other by-products of self-seeking activity, and benefiting them in less contingent fashion. The government, pursuing the lax difference principle, might tax me more, or less, than whatever should count as a reasonable compromise between self-interest and service. If it taxed me more than that, I would, according to the reply, have reason of legitimate self-interest to object, and the reply therefore represents the difference principle as (at times) too demanding, even in its lax form. And if government taxed me less than what a reasonable compromise would dictate, then I could not say that the laxness in the principle I affirm was justified

<sup>31</sup> It would be a mistake to think that the priority of liberty over the difference principle makes for such a compromise. Among the reasons why it does not do so is that we are not here concerned with coercive restrictions, in the name of justice, on the doctor's liberty, but with what would count as a just use of his liberty.

on compromise grounds: for on those grounds it would then be too lax.<sup>32</sup>

In short, the compromise idea will not in general draw the same line as the lax difference principle does. Defended along compromise lines, the lax difference principle is at best an imperfect proxy for a just balance, and not, what it is supposed to be, a fundamental principle of justice. The compromise idea is, simply, different from the idea that inequalities are justified if they are necessary to benefit the badly off, given that agents are (or might be) self-regarding maximizers on the market. Accordingly, the lax difference principle cannot be what agents committed to difference-principle justice affirm: from their point of view, it draws an arbitrary line between serving oneself and serving others.

We are left with the strict difference principle,<sup>33</sup> which government cannot by itself implement. For the strict difference principle to prevail, there needs to be an ethos informed by the principle in society at large. Therefore, a society (as opposed to its government) does not qualify as committed to the difference principle unless it is indeed informed by a certain ethos, or culture of justice. Ethoses are, of course, beyond the immediate control of legislation, but I believe that a just society is normally impossible without one,<sup>34</sup> and Rawls himself requires that there be a nur-

<sup>32</sup> With a certain distribution of talent, the inequalities allowed by the lax difference principle could be quite large, of a size that is intuitively incongruent with the central Rawlsian idea that the gifted owe their special powers to mere good fortune. How can *they* be thoroughly just people, think themselves merely lucky to have the assets they do, and nevertheless take as much advantage of them as they can on the market? Of course, there are also distributions of talent under which the inequalities might be quite small. But they are not small as a matter of principle, and it is no defense of a supposed fundamental principle, when its consistency with certain consequences generates criticism, to show that it lacks those consequences in practice. So defended, the principle is not, as intended, fundamental, but warranted because, given the facts, it serves more fundamental aims.

<sup>33</sup> I do not mean that there is no other game in town, but just that there is no third way of playing the difference principle game. (A further alternative would be the strict difference principle constrained by an agent-centered prerogative. But the added constraint *modifies*— it does not interpret— the difference principle.)

<sup>34</sup> For example, because of problems of asymmetrical information and incentive compatibility that are familiar to economists, and that are crudely illustrated by the

turance and cultivation of appropriate attitude in the just society that he describes.

In a culture of justice shaped by the difference principle, talented people would not expect (what they usually have the power to obtain) the high salaries whose level reflects high demand for their talent (as opposed to the special needs or special burdens of their jobs). It follows that the difference principle in a society of just people would not induce the inequality it is usually thought (e.g., by Rawls) to produce, and it would not, in particular, justify incentive payments in the “standard” sense of that phrase (see the table above), that is, payments not to compensate for unusually arduous work, but to draw talent to jobs that are not in general especially grueling. In a just society, where justice is defined by the difference principle in its preinterpreted form, the difference principle will prevail in its strict interpretation.

(It is not true that, in the society I have in mind, a person would have to worry about unfortunate people every time he made an economic decision. Liberals would regard that as oppressive,<sup>35</sup> and, whether or not they are right, one function of the egalitarian ethos is to make conscious focus on the worst off unnecessary. What rather happens is that people internalize, and—in the normal case—they unreflectively live by, principles which restrain the pursuit of self-interest and whose point is that the less fortunate gain when conduct is directed by them.)

#### 14.

On the lax interpretation of what the difference principle demands, it is satisfied when everyone gets what he can through self-

---

propensity of the productive to withdraw labor when taxes rise too high. Under abnormal conditions, justice might be consistent with universal self-interested maximizing: if, for example, talents and utility functions are identical, then initial equality of tangible assets might be considered sufficient for justice. On a Dworkinian view, that would be so even with different utility functions.

<sup>35</sup> See, e.g., Thomas Nagel, “Libertarianism without Foundations,” in J. Paul, ed., *Reading Nozick* (Totowa, N.J.: Rowman and Littlefield, 1981), pp. 199–200.

seeking behavior in a market whose rewards are so structured by taxation and other regulation that the worst off are as well off as any scheme of taxed and regulated market rewards can make them. On my view of what it means for a society to institute the principle, people would mention norms of equality when asked to explain why they and those like them are willing to work for the pay they get. This strict interpretation conflicts with Rawls's endorsement of unequalizing incentives. Yet, as I now propose to show, the strict interpretation of the principle coheres with a number of significant general characterizations of justice to be found in Rawls's work.

It is very important in the present connection that Rawls's theory describes what he calls a well-ordered society, one, that is, whose citizens display full and willing compliance with the demands of justice. In a well-ordered society each person acts out of a sense of justice informed by the principles of justice not merely at the ballot box but as he goes about his daily business.

So much is clear from many passages in Rawls's writings. We are told not only that "everyone accepts, and knows that others likewise accept, the same first principles of right and justice," which might, by itself, be consistent with a ballot box view of their commitment, but also that the parties "in everyday life . . . affirm and act from [those] first principles of justice."<sup>36</sup> Full compliance with the principles means that they act *from* them, in everyday life, "in the course," as Rawls also puts it, "of their daily lives."<sup>37</sup> And their "full autonomy is achieved" partly through "acting *from* these principles as their sense of justice dictates."<sup>38</sup> Citizens are strongly committed to acting that way. They "have a highest-order desire, their sense of justice, to *act* from the principles of jus-

<sup>36</sup> Rawls, "Kantian Constructivism," p. 521.

<sup>37</sup> *Ibid.*, p. 528. Cf. Rawls, *A Theory of Justice*, p. 253: they "knowingly act on the principles of justice in the ordinary course of events."

<sup>38</sup> Rawls, "Kantian Constructivism," p. 528, emphasis added.

tice.”<sup>39</sup> They “have a desire to express their nature as free and equal moral persons, and this they do most adequately by acting *from* the principles that they would acknowledge in the original position. When all strive to comply with these principles and each succeeds, then individually and collectively their nature as moral persons is *most* fully realized, and with it their individual and collective good.”<sup>40</sup>

Now, such statements seem to me to imply that the economic motivation of Rawlsian citizens is influenced by the difference principle. How could they act like maximizing incentive seekers if in “*their daily lives*” they act “*from*” a principle which directs primary concern for the badly off? Can we say that they act *from* such a principle in their daily lives just because they support taxation which is shaped by the principle and which aims to modify the results of their acting *from* maximizing motives? Such support might show that you respect the claim of the principle against you, but it surely does not suffice as proof of your being inspired by it as part of a sense of justice on which you operate in your daily life.<sup>41</sup> How could your “nature as [a] moral person” count as “*most* fully realized” when you go for as much as you can get in your own market choices,<sup>42</sup> and merely endorse application of the principle by the government in imperfect moderation of the

<sup>39</sup> *Ibid.*, p. 532, emphasis added.

<sup>40</sup> Rawls, *A Theory of Justice*, p. 528, emphases added.

<sup>41</sup> Rawls says that “citizens have a normally effective sense of justice, that is, one that enables them to understand and to apply the principles of justice, and for the most part to act from them as their circumstances require (*Justice as Fairness*, p. 154). Why would they have to apply the principles themselves to their own circumstances if just behavior consisted in obeying laws designed to effect an implementation of those principles?”

<sup>42</sup> How does the economic behavior of a maximizer who is committed to the lax difference principle differ from that of a maximizer who is not? It might be said that, unlike the latter, the former is willing to maximize only when and because the principle is in force. But that is not necessarily true: people who believe that the lax difference principle should be instituted may have various views about what they should do in a society in which it is not. And even if our believer would indeed behave nonmaximizingly if the principle were not in force, that hardly shows that he “strive[s] to comply with” the difference principle in his “daily life.”

inequality which the choices of people like you tend to cause?

Consider this passage from *A Theory of Justice*: “by abstaining from the exploitation of the contingencies of nature and social circumstances within a framework of equal liberty, persons express their respect for one another in the very constitution of their society.”<sup>43</sup> If that is so, then it seems to me that in the Rawlsian society there will not be incentive seekers, since they do exploit their contingent talent and social advantages, and the passage says that people who do that show a lack of the respect for other people that the constitution of their society requires. If you deny that the passage has this implication, then you must make one or other of two implausible claims, You must claim either that (1) despite what the passage says, Rawlsianly just talented people might exploit the contingency of their superior talent, or that (2) contrary to what seems evident, talented market-maximizers do not engage in such exploitation.<sup>44</sup>

Think about it this way. On a Rawlsian view, there is no reason of basic principle why the talented should earn more than the untalented. It is merely that things (supposedly) fall out that way, when the difference principle is applied. So imagine that we address the talented rich people, and we ask them why they do not give the above-average parts of their incomes to people of below-average income, when, *ex hypothesi*, they would have compliantly accepted the resulting post-giveaway incomes had the difference principle happened to mandate them. What could they say? They

<sup>43</sup> Rawls, *A Theory of Justice*, p. 179, and see, generally, pp. 72–75.

<sup>44</sup> It has been claimed, against my interpretation of the quoted passage, that it speaks of an expression of mutual respect when persons choose their constitution and not, as I have supposed, when they act in the society it constitutes. But there is no scope of “abstaining from the exploitation of the contingencies of nature and social circumstance” at the stage of constitutional choice, since, at that point, no one knows what those contingencies are. This and other phrases (not quoted above) in the paragraph from which the passage is drawn establish that the objection is misguided, that Rawls is here commenting on people’s choices in real life rather than in the original position. He is speaking about how things are “when society follows these principles” (*ibid.*, p. 179).

certainly could not say that they were abstaining from exploitation of their talent advantages, and we could not say that they live under “a conception of justice that nullifies the accidents of natural endowment and the contingencies of social circumstance as counters in the quest for . . . economic advantage.”<sup>45</sup>

## 15.

Rawls believes that a just society, on the lax understanding of how it operates, honors the dignity of the worst off, since, so he says, they know that they are caused to be as well off as they could be. But that is an illusion. For they are as well off as they could be only *given* the self-seekingness of those who are better off, and maybe far better off, than they.

Joshua Cohen is a strong advocate of the difference principle. He draws a contrast between a society ruled by that principle and one whose rule is a basic minimum for all and then laissez-faire. Cohen disparages the basic minimum/laissez-faire arrangement because of how weak its “affirmation of [the] worth” of the worst-off individuals is. For if I am one of them in such a society, “then I know that I could do better if those who are better off were prepared to forgo some of their advantages. And I know that this loss of advantage to me is not just for a stretch of time but covers the course of my entire life. Others know this, and know that I know it, and so on. Still they accept the advantages.”<sup>46</sup> Yet Cohen fails to see that all those things can be said about the less gifted in a society ruled by the lax difference principle, where talented people demand, and get, incentive payments. In such a society, clear-thinking unfortunate people know that they “could do better if those who are better off were prepared to forgo some of their advantages.” Cohen describes a badly placed person in a Rawlsian society reflecting with satisfaction that “other citizens act from maximin” (that is, in this context, from the principle of putting

<sup>45</sup> Rawls, *A Theory of Justice*, p. 15.

<sup>46</sup> Joshua Cohen, “Democratic Equality,” *Ethics*, July 1989, p. 743.

the interests of the badly off first) and thereby “display a concern for my good and the good of those to whom I am attached.”<sup>47</sup> But the badly placed person can enjoy such a reflection only on my revisionary conception of the character of a Rawlsian society. When the difference principle in its standard, lax, interpretation prevails, it is not in general true that citizens “act from maximin,” and the inequalities that come as a result might challenge the sense of self-worth of those who are at the bottom. If they succeed in sustaining that sense, that will not be because of their perception of how the better off regard them.

Joshua Cohen’s remarks are in the spirit of the Rawls passage which says that “the least favored man,” here called B, “can accept A’s being better off since A’s advantages have been gained in ways that improve B’s prospects. If A were not allowed his better position, B would be even worse off than he is.”<sup>48</sup> The second sentence of this passage does not compel agreement with the first, since, with everything else equal, A could have refrained from seizing the full complement of the advantages he was able to seize, and then B would have been better off than he is. It indicates how little A cares about B’s lot that he refuses to improve B’s bad prospects without the advantages he gets in the course of doing so.

The mistake in Cohen’s comment on the difference principle shows that, given his own lax application of it, Rawls is wrong to represent it as a realization of the value of fraternity, which he glosses as “the idea of not wanting to have greater advantages unless this is to the benefit of others who are less well off. . . . Members of a family commonly do not wish to gain unless they can do so in ways that further the interests of the rest. Now wanting to act on the difference principle has precisely this consequence.”<sup>49</sup> But “wanting to act on the difference principle” has the stated consequence only if we interpret the principle strictly. For wanting

<sup>47</sup> Ibid., p. 746.

<sup>48</sup> Rawls, *A Theory of Justice*, p. 103.

<sup>49</sup> Ibid., p. 105.

not “to gain unless they can do so in ways that further the interests of the rest” is incompatible with the drive for enrichment motivating market maximizers.

We should note an ambiguity in the phrase “not wanting to have greater advantages unless this is to the benefit of others who are less well off.” A person of that description does not want to gain unless others thereby do. Does that mean: unless they gain something (no matter how little)? But that is not the maximin concept. Or does it mean: unless none of his gain means that theirs is less than it need be? But, provided that the feasible set is sufficiently ample, that means going for equality.<sup>50</sup>

A society of maximizers with taxation and regulation dictated by the lax difference principle is necessarily preferable from the point of view of the worst off to a laissez-faire society; but in neither society is the conduct of high fliers consistent with the essentially socialist value of fraternity or with motivation informed by the difference principle. Rawls must give up either his approval of incentives to the exercise of talent or his ideals of dignity, fraternity, and the full realization of persons’ moral natures. I think the ideals are worth keeping.<sup>51</sup>

<sup>50</sup> Let me illustrate this point. Two brothers, A and B, are at benefit levels 6 and 5, respectively, in New York, where they live. If they moved to Chicago, their levels would rise to 10 and 6. If they moved to Boston, they would rise to 8 and 7. Is fraternity, as Rawls means to characterize it, consistent with A proposing that they move to Chicago? If so, it is a thin thing. Or is Rawlsian fraternity strictly maximizing? In that case, Boston is the choice, and, in a feasible set with no bar to redistribution, equality is the result.

<sup>51</sup> It might be thought that, beyond his commitment to those ideals, Rawls has further reason to reject incentives and the lax difference principle, to wit, that the risk aversiveness which induces the parties in the original position to select the difference principle would also incline them to prefer its strict form. For two reasons, I have not used this argument. The first reason is that I seek to pursue my case against the part of Rawls to which I object by invoking Rawlsian ideas with which I agree, and I agree neither that principles chosen in the original position are ipso facto just nor that its parties would choose a maximin strategy and, therefore, if the foregoing suggestion is sound, the strict difference principle. The second reason is that the strict difference principle *might* be thought to imply a principle, or set of principles, “for individuals,” as Rawls uses that phrase, and it is difficult to say both whether it does and whether, if it does, the principle or principles are of the right type to be chosen in the original position: see Rawls, *A Theory of Justice*, pp. 108–10, 115, 135.

## 16.

At one point, Rawls comments on the view that “the greater expectations allowed to entrepreneurs encourages [*sic*] them to do things which raise the long-term prospects of the laboring class.”<sup>52</sup> He does not (quite) endorse that factual claim, but he says that, if, as he shows he believes, it is true, then the difference principle recommends the rewards generating those greater expectations and the “initial inequality in life prospects” associated with them.

There are other passages to relevantly similar effect,<sup>53</sup> and there is no point quoting them all here. But I do want to quote and comment on a remark by Rawls which might be read as an attempt to anticipate and deflect the line of criticism that I have developed.

Following one of his incentives-endorsing passages, Rawls says: “One might think that ideally individuals should want to serve one another. But since the parties are assumed not to take an interest in one another’s interests, their acceptance of those inequalities is only the acceptance of the relations in which men stand in the circumstances of justice. They have no grounds for complaining of one another’s motives.”<sup>54</sup> It might be said, on the basis of this passage, that my critique of Rawls displays misunderstanding of the role of principles of justice, as he conceives them. Those principles, it might be said, are rules observed by fair-minded people in their mutually advantageous interaction, fair-minded people who may or may not *care* about one another, but who qualify as just as long as they observe the rules. They go beyond justice if they do care about one another, and in demanding that the difference principle be strict I am demanding more than justice.

<sup>52</sup> *Ibid.*, p. 78.

<sup>53</sup> See *ibid.*, pp. 151, 157, 279; Rawls, *Justice as Fairness*, pp. 44, 46; Rawls, “Distributive Justice,” in P. Laslett and W. G. Runciman, eds., *Philosophy, Politics and Society*, 3d series (Oxford: Basil Blackwell, 1967), p. 67.

<sup>54</sup> Rawls, *A Theory of Justice*, p. 151. For more on “mutual disinterest,” see Rawls, *Justice as Fairness*, p. 62.

But that line of thought seems to me untenable. For it wrongly attributes to people in the achieved, just society the mutual indifference that characterizes the specially tailored persons of Rawls's original position, in which the principles that are to govern the just society are chosen. In the original position mutual indifference is assumed for methodological reasons, to derive justice from rational self-interest under a veil of ignorance constraint. But it does not follow that the principles chosen by the mutually indifferent parties of the original position are consistent with mutual indifference when they operate as rules of interaction in a functioning society. And to attribute mutual indifference to people in the realized society is, surely, to contradict the idea that their relations partake of fraternity, as Rawls describes that condition (see sec. 15 above).<sup>55</sup> People who, like "members of a family," "do not wish to gain unless they can do so in ways that further the interests of the rest" are not people who take no interest in one another's interests. How could a person who takes no interest in the interests of others want advantages for himself only if his enjoyment of them benefits the less well off (see sec. 15) ?

## 17.

Rawls says that "a person in the original position would concede the justice of [the] inequalities [required for incentives]. Indeed, it would be short-sighted of him not to do so."<sup>56</sup> Now, the phrasing of this contention is curious, since we normally think of short-sightedness as poor perception not of justice but of one's own interests.<sup>57</sup> And I point out this infelicity in the formulation

<sup>55</sup> Rawls himself distinguishes similarly between people's attitudes to one another in the original position and in society when he writes that "although the parties in the original position take no interest in each other's interests, they know that in society they need to be assured by the esteem of their associates. Their self-respect and their confidence in the value of their own system of ends cannot withstand the indifference . . . of others" (*A Theory of Justice*, p. 338).

<sup>56</sup> *Ibid.*, p. 151.

<sup>57</sup> A person in the original position does not, in any case, ask himself what is *just*. He asks himself what, given his ignorance, is the best choice from the point of view of his interests.

because I believe that it reflects an unresolved tension in the Rawlsian architectonic, one that underlies the difficulties exposed in these lectures. That underlying tension is between a *bargaining* conception and a *community* conception of social relationships. (There are conceptions which fall between those two, with elements of each, but, as I read Rawls, both of them appear in his work, in relatively pure forms.)

But let us ignore the infelicitous phrasing in the passage, and concentrate on the implied claim that it would be a mistake not to concede the justice of incentive inequalities. My reply to that claim, a reply that by now is entirely predictable, is that, *if* we are talking within the assumption of full compliance, then we need not and should not concede either that incentive inequalities are required to motivate performance or that they are just. Let us now, however, retire the heady assumptions of full compliance and a widespread sense of justice. Consider, instead, a society like the United States, where fortunate people learn to expect more than they would get when the difference principle prevails in a comprehensive way. In that case, we might agree that it would be a mistake not to concede incentive inequalities. If we need inequalities to “encourage effective performance”<sup>58</sup> then it might be folly not to have them, but it does not follow that having them is a requirement of basic justice, where a *basic* principle of justice is one that has application in a society where, as in Rawls’s, everyone always acts justly.

Although his primary topic is justice under full compliance, Rawls also treats his principles as standards for assessing actually existing society.<sup>59</sup> In my view, the difference principle, conceived as one that would govern a just society, condemns as unjust those existing inequalities which are necessary to benefit the worst off where that necessity reflects the intentions of the talented rich; but, given that the inequalities are necessary, albeit for the stated

<sup>58</sup> Rawls, *A Theory of Justice*, p. 151.

<sup>59</sup> See *ibid.*, pp. 245–46.

reason, to remove them would be reckless. Along with Nikolai Bukharin, I would have said to the kulaks: “Enrich yourselves!” without supposing (any more than Bukharin did) that I was thereby voicing a demand of justice. If we are concerned about the badly off, then we should sometimes concede incentives, just as we should sometimes satisfy even a kidnapper’s demands. We are not then acting on the difference principle in its strict interpretation, in which it is a principle of justice governing a society of just people who are inspired by it. We are acting on the lax version of the difference principle, which endorses incentives and which has application in societies of the familiar unjust kind. On the assumption that they are indeed unavoidable, incentive payments may be justified, but it does not follow that no injustice occurs when they are provided. (One might say, to a child’s guardian: the kidnapper is unjustly threatening the safety of the child, and justice to the child therefore demands that you pay him. And one might say, to legislators in a structurally unequal society: the talented are unjustly indifferent to the plight of the poor, and justice to the poor therefore demands that you do not impose very high taxation.)

The policy of paying productive people plenty to get them to produce so that badly off people will be better off is rational when productive people are resolved to serve only if they are richly rewarded. But their stance is then unjust by the very standard which the difference principle itself sets. Accordingly, on a strict view of Rawlsian justice, the difference principle in its lax interpretation, which does mandate the incentives policy, is not a basic principle of justice but a principle for handling people’s injustice. It is not a basic principle of justice, since it confers benefit on market maximizers who offend against justice. We might call it a principle of damage limitation in the field of justice.<sup>60</sup>

When doing so limits the damage, it is wise to run society on lax difference principle lines, but it is also wise to recognize that

<sup>60</sup> Or, a “principle for meeting injustice” (ibid., p. 246).

society is not then based on justice. A related and more general point is that one should not suppose that, as Rawls says in *A Theory of Justice*, “justice is *the* first virtue of social institutions,” where that means that “laws and institutions . . . must be reformed or abolished if they are unjust.”<sup>61</sup> For sometimes justice is unattainable, and we do well to settle for something else. When there is no way to get the child back without paying, when a just outcome is not to be had, then paying, which makes all (kidnapper, parents, child) better off than refusing to pay, is almost certainly preferable, although in some cases, with less at stake, we might prefer to forgo the Pareto improvement, in order not to accede to an unjust demand.

Similarly, and according to an ancient Marxist wisdom, justice is not the first virtue of institutions in conditions of scarcity. Under those conditions a just distribution may be impossible to achieve, since powerful people will block it. In that case striving for justice may make everyone worse off, and unjust laws and institutions should not be “reformed or abolished.” And scarcity in the Marxist sense is not poverty of supply, but the wider circumstance that, to secure what might be a quite reasonable supply, most people must spend most of their time engaged in labor that interferes with self-realization.

Under such a condition, and it is a huge and difficult question whether we are still in it, it might be right to tolerate, and even, sometimes, to nourish, incentive motivation, despite the fact that it contradicts justice. Sometimes, the difference principle, in its lax interpretation, can be recommended as a first virtue of social institutions, because we cannot get justice, and the injustice that goes with incentives is the best injustice we can get.

## 18.

My principal contention about Rawls is that (potential) high fliers would forgo incentives properly so-called in a full compli-

<sup>61</sup> *Ibid.*, p. 3, my emphasis.

ance society governed by the difference principle and characterized by fraternity and universal dignity. I have not rejected the difference principle in its lax reading as a principle of public policy: I do not doubt that there are contexts where it is right to apply it. What I have questioned is its description as a principle of (basic) justice, and I have deplored Rawls's willingness to describe those at the top end of a society governed by it as undergoing the fullest possible realization of their moral natures. My own socialist-egalitarian position was nicely articulated by John Stuart Mill in his *Principles of Political Economy*. Contrasting equal payment with incentive-style payment according to product ("work done"), Mill said that the first

appeals to a higher standard of justice, and is adapted to a much higher moral condition of human nature. The proportioning of remuneration to work done is really just, only in so far as the more or less of the work is a matter of choice; when it depends on natural difference of strength or capacity, this principle of remuneration is in itself an injustice: it is giving to those who have; assigning most to those who are already most favoured by nature. Considered, however, as a compromise with the selfish type of character formed by the present standard of morality, and fostered by the existing social institutions, it is highly expedient; and until education shall have been entirely regenerated, is far more likely to prove immediately successful, than an attempt at a higher ideal.<sup>62</sup>

Rawls's lax application of his different principle means "giving to those who have." He presents the incentive policy as a feature of the just society, whereas it is in fact, and as Mill says, just "highly expedient" in society as we know it, a sober "compromise

<sup>62</sup> John Stuart Mill, *Principles of Political Economy*, book 2, chap. 1, sec. 4, in J. M. Robson, ed., *Collected Works of John Stuart Mill*, vol. 2 (Toronto: University of Toronto Press, 1965), p. 210. In chapter 5 of his *Utilitarianism* Mill argues, at great length, that justice is a species of expediency. But here the self-same principle of remuneration is, under the stated conditions, both "highly expedient" and "an injustice." It is a nice question whether that conjunction of designations is compatible with everything that Mill says in *Utilitarianism*.

with the selfish type of character” formed by capitalism.<sup>63</sup> Philosophers in search of justice should not be content with an expedient compromise. To call expediency *justice* goes against the regeneration to which Mill looked forward at the end of this fine passage.

<sup>63</sup> For sapient criticism of Rawls along these lines, see Allen Buchanan, *Marx and Justice* (Totowa, N.J.: Rowman and Littlefield, 1982), pp. 127–28. According to Mill, “the deep-rooted selfishness which forms the general character of the existing state of society is so deeply rooted only because the whole course of the existing institutions tends to foster it” (*Autobiography* [New York: New American Library, 1965], pp. 168–69). See, for further pertinent references, Richard Ashcraft, “Class Conflict and Constitutionalism in J. S. Mill’s Thought,” in Nancy Rosenblum, ed., *Liberalism and the Moral Life* (Cambridge: Harvard University Press, 1989), pp. 117–18.