THREE PROBLEMS WITH CONTRACTARIAN-CONSEQUENTIALIST WAYS OF ASSESSING SOCIAL INSTITUTIONS*

BY THOMAS W. POGGE

We can look at the events of our social world in two ways. On the one hand, we can see them interactionally: as actions, and consequences of actions, performed by individual and collective agents. On the other hand, we can see them institutionally: as effects of how our social world is structured—effects of our ground rules, practices, or social institutions. These two ways of viewing events lead to different descriptions and explanations of social phenomena. I will here discuss how they can also lead to different moral assessments.

I. INSTITUTIONAL MORAL ANALYSIS

During the last few decades, political philosophy has drifted from interactional toward institutional viewing. We ask less frequently how governments and their agencies and officials (as well as other agents in the political arena) ought to conduct themselves, and more frequently how social institutions ought to be designed. In the English-speaking world, this shift is largely due to John Rawls, who has reconceived justice as “the first virtue of social institutions.”¹ It is worth noting, however, that the same shift shows up elsewhere as well—for example, in the German “critical theory” tradition. Having been an enthusiastic supporter of this shift, I will here examine some of its problems.

Talk of social institutions does not refer to organized collective agents such as General Motors or the Environmental Protection Agency; a focus on institutions in this sense would exemplify interactional moral analysis. Rather, institutions are a social system’s practices or “rules of the game,” which govern interactions among (individual and collective) agents as well as their access to material resources. Institutions define and regulate property, the division of labor, sexual and kinship relations, con-

* Work on this essay was supported by a Laurance S. Rockefeller fellowship at the Princeton University Center for Human Values. Many thanks also to Marko Ahtisaari, Richard Arneson, Christian Barry, Hilary Bok, Tony Coady, Peter de Marneffe, Amy Gutmann, Alan Houston, Hidé Ishiguro, Yasuo Iwata, Francis Kamm, George Kateb, Bonnie Kent, Andrew Koppelman, Arthur Kuflik, Sidney Morgenbesser, Liam Murphy, Thomas Nagel, Takeshi Ohba, Debra Satz, Ling Tong, and the editors of this volume for all their critical help.

control over and responsibility for children, as well as political and economic competition; and they govern how collective projects are undertaken and executed, how conflicts are settled, and how social institutions themselves are created, revised, interpreted, and enforced. The totality of the more fundamental and pervasive institutions of a social system may be called its basic structure (as Rawls calls it) or institutional scheme.

Later we will need the distinction between the rules of the game as it is supposed to be played, according to some ideal blueprint laid down perhaps in the law books, and the rules of the game as it is actually played. Of course, these two need not diverge, but they generally do—as when particular conduct is in theory to be punished, but in practice generally ignored by the relevant authorities. I will use the term "ground rules" to refer to the official rules of the game, and "practices" or "institutions" to refer to how such ground rules are actually understood, implemented, and followed.

The concern with social institutions is motivated by an increasing appreciation of their importance—and of their increasing importance—for human welfare. We have gradually come to understand how deeply the structure of our society shapes our conduct—not merely by determining in large part a person's menu of options and the various incentives and disincentives attached to them, but also by influencing rather profoundly what interests, desires, and abilities persons develop in the first place. Moreover, human lives are increasingly interdependent, affecting one another through highly complex networks of interrelations. This trend is well illustrated by our modern practices of economic cooperation, with their extended markets for goods, services, and capital, through which our decisions may intensely affect the livelihoods of unknown others over great distances. Because such effects result from the confluence of decisions made independently by a multitude of agents, it is unfeasible to hold such agents morally accountable for those effects; we cannot plausibly blame particular market participants for widespread homelessness and unemployment. Yet we also cannot plausibly view these as natural calamities on a par with the striking of a large meteorite. A straightforward solution to this dilemma is to view the participants in an institutional scheme as having not merely an individual moral responsibility for their conduct within it, but also a shared collective moral responsibility for its terms. If a laissez-faire market scheme tends to produce more severe and more widespread poverty than some feasible alternative ways of arranging the economy, and if this poverty is morally regrettable, then, other things equal, we should support institutional reforms toward such an alternative. It is our common moral task and duty as citizens to set and reset the institutional parameters of our society in light of their foreseeable morally relevant effects. Moreover, this is not part of a wide-ranging and weak positive duty to do something to improve the world, but a sharp and strong negative duty not to collaborate in the imposition of an
unjust institutional scheme without working toward institutional reform or at least helping the victims of injustice.

What this negative duty comes to more concretely depends on a variety of factors: Which institutional schemes can I be said to participate in—for example, am I a participant in a global scheme of trade and diplomacy, and thus co-responsible for its effects? How do we assess institutional alternatives in order to determine whether an existing scheme really is unjust? What reforms of any unjust institutional scheme I participate in are feasible? What can I do to promote such reforms, and what can I do to support the victims of those unjust institutional schemes?

Institutional viewing, though suggested by cases like that of complex economic interrelations which resist interactional moral analysis, is not confined to cases of that sort. If a legal requirement of regular safety-inspections of automobiles tends to reduce the incidence of serious traffic accidents, and if the burdens imposed by the requirement are of comparatively minor moral significance, then we ought to institute this requirement. Here it is not implausible, when the requirement is absent, to attribute moral responsibility to individuals: those who own or operate unsafe automobiles can quite plausibly be held morally accountable for any resulting accidents. One can nevertheless locate an additional, shared moral responsibility at the collective level. One can say that the citizenry at large is responsible for imposing a scheme that, without offering important compensating benefits, foreseeably produces an excessively and avoidably high rate of traffic casualties.

Saying this seems compelling: for why should it matter whether harm we could prevent by changing the rules can or cannot also be seen as the responsibility of agents? Compare a fair tax code that does not work well because it is too complex for people to follow correctly with another one that works equally poorly because it is difficult to monitor and thus leads to widespread noncompliance. Ought we not change each of these codes, and for the same basic reason?

II. THE CONTRACTARIAN-CONSEQUENTIALIST APPROACH

Once we view our social world institutionally, we will be drawn to a broadly consequentialist mode of assessment: an institutional scheme is just, or good, if it is on the whole better for the persons living under it than its alternatives would be. In the case of conduct, such a "recipient-based" mode of assessment is at least problematic, if not implausible:

2 Such formulations can be found not only in Rawls and in utilitarian authors, but also on the Continent—for example, in Jürgen Habermas, *Moralbewusstsein und kommunikatives Handeln* (Frankfurt: Suhrkamp, 1983), pp. 75f., where Habermas proposes that social norms should be assessed by reference to the interests of all those affected by them. (For an English translation, see Jürgen Habermas, *Moral Consciousness and Communicative Action*, trans. C. Lenhardt and S. W. Nicholsen [Cambridge, MA: MIT Press, 1990].)
most would want to bring in the standpoint of the agent as well, and make morally significant the distinctions between harms she intends and harms she merely foresees, between harms she brings about and harms she merely allows to happen, and/or other distinctions concerning how an agent is related to the effects of her conduct.3

This agent's standpoint seems to have no analogue when the assessment is not of conduct but of ground rules or practices or institutional schemes. The latter are mere means for regulating our interactions so that we all have a better life; thus, there seems to be no reason for giving moral significance to any distinctions between different classes of effects that ground rules might have—for example, to the distinction between the benefits and burdens specifically called for by an institutional scheme, and others that merely foreseeably come about under it. The moral assessment of ground rules should then be entirely recipient-based—a point nicely expressed in the contractarian image of an ideal institutional scheme chosen prudentially by its prospective participants.4 Such prospective participants (or their representatives) will be interested exclusively in results: in the quality of life they can expect under this or that institutional scheme. They will not care how these results come about: through desert or luck, through decree or voting, through birth or the market—except insofar as these mechanisms themselves affect the quality of their lives.

Adopting a prospective-participant perspective, we will take the justice (J) of institutional schemes to be a function of the quality of the individual lives (Q_i) each tends to produce: J = f(Q_i). Much of the discussion in political philosophy in the English-speaking world these days is about how—perhaps through some hypothetical-contract story—to fill in the two key terms (Q_i and f) in this broadly consequentialist equation: Should the relevant notion of the good for individuals be specified in subjective terms (human happiness, desire fulfillment, preference satisfaction, pain avoidance, or opportunity for welfare) or in objective terms (human rights, resources, social primary goods, capabilities, or the fulfillment of human needs)? And should social institutions maximize the sum total, the arithmetic or geometric mean, the bottom level (Rawls's minimax), or equality in the distribution, of quality of life? I will not worry here about which of the many possible combinations within and across

3 We generally do not believe, for example, that neglecting to save someone is morally on a par with killing her. Nor do we think that one ought to kill whenever doing so will save a larger number from being killed. We admit that it might be better for people in general if we accepted these principles. The disposition to kill one for the sake of saving several may be desirable from the standpoint of others at large. But we do not let the recipients' standpoint alone settle the matter.

4 I use "contractarian" in reference to the idea of a hypothetical social contract (as opposed to, e.g., a historical one). Though my primary example is Rawls, I mean to include more thinly veiled social contracts as exemplified by David Gauthier, *Morals by Agreement* (Cambridge: Cambridge University Press, 1987).
the two key terms is morally most plausible and most suitable as a public standard of justice, because I want to present some general concerns about any such contractarian-consequentialist (or CC) way of assessing social institutions.

By coining the term "CC," I do not mean to suggest that consequentialism and contractarianism are at bottom the same. There are nonstandard versions of consequentialism which do not easily fit the contractarian mold—for example, versions that assess consequences in nonindividualistic or nonhumanistic ways. There are also nonstandard versions of contractarianism that do not easily fit the consequentialist mold—for example, Kant's version, in which the contractors, duty-bound, reendorse a preexisting deontological content. My point is that the standard consequentialist conceptions of justice are straightforwardly presentable in contractarian terms, and vice versa. My interest is in this intersection, and the adjective "CC" is meant to single out the conceptions of justice within this broad overlap of consequentialism and contractarianism.

Let me give three brief illustrations of how CC assessments have been deployed. (1) Some have argued for restricting the availability of alcohol (for example, to teenagers). Such restrictions would, they say, significantly reduce the incidence of serious traffic accidents, and it would hence be rational for prospective participants to authorize them. (2) Some feminists have advocated a ban on pornographic materials. This ban would, they say, greatly reduce the incidence of serious crimes against women, and it would therefore be prudent for prospective participants (ignorant of their gender) to insist on such a ban. (3) Proponents of more stringent handgun legislation often argue that their proposal would greatly reduce the number of gunshot victims, so that prospective participants would be well advised to demand such controls.

Of course, some CC conceptions may not recognize as such either the cost or the benefit in some of these cases. Others may recognize both, but take one to be of infinitely smaller weight than, or lexically subordi-

---

5 Pure egalitarianisms, though they accept individualism and humanism, also fall outside the CC mold. It makes sense to stipulate that prudent prospective participants assess their positions in relative terms and care greatly not to fall below the mean. However, the stipulation that this is all they care about—that they do not care at all about their absolute positions (that they are indifferent, for example, between higher and lower levels of equality)—is at odds with the assumption of rational prudence. Though outside the CC mold, pure egalitarianisms may be subject to some of the same problems as CC conceptions of justice—for example, to the first problem discussed below (Section III).


7 Catharine MacKinnon, for example, would presumably not recognize the lesser availability of pornography as a loss. See, for example, her Feminism Unmodified (Cambridge, MA: Harvard University Press, 1987).
nate to, the other. As Rawls's conception attests, the CC approach does not commit us to a narrow cost-benefit analysis, which treats all goods and ills as commensurable.

Even CC conceptions that—like institutional utilitarianism—are committed to the commensurability of all goods and ills can justify rights, and even rights conceived as absolute trumps. It is true that the question of what legal protections, if any, there should be of some activity—say, political speech-making—is settled on the basis of the consequences various answers would in fact produce. It may well turn out, however, that the winning answer is: Political speech-making should be protected by a trump right; that is, any person's effort to give a political speech should be protected by the laws and power of the state irrespective of the net cost or benefit of either the speech or its protection. This would be so because institutional schemes that make the legality of an utterance hinge on its effects have bad effects themselves: they engender abuse, spread fear and insecurity, and hamper free public debate about the common good. CC conceptions of justice may, then, provide reasons for wanting ground rules to include some basic rights that shield specific kinds of conduct from consequentialist calculations. The consequentialist formula is invoked only for the purpose of deciding how the contours of a package of basic rights should best be drawn, that is, how conduct should be categorized, and whether particular categories of conduct should be protected (and, if so, how firmly). 8

Having introduced institutional moral analysis in Section I, I have tried to show in Section II that the two most prominent traditions of institutional moral analysis are not as distinct as is widely believed. Understood as guides to the assessment of social institutions, contractarianism and consequentialism are for the most part not competitors but alternate presentations of a single idea: both tend to assess alternative institutional schemes exclusively by how each would affect its individual human participants. 9 Insofar as this is so, we have only a single worked-out approach to institutional moral analysis.

In the remainder of this essay, I will try to show that this CC approach to the assessment of social institutions is perhaps not as plausible as I and many others have been prone to believe. I see three main problems, which I will, as far as possible, discuss within the framework of one well-known CC conception: that of John Rawls. I hope this will enhance the

8 Whether it can reasonably be confined to this role is another question, which has haunted utilitarianism: Why should a judge follow the best rule even in those few exceptional cases where she knows that violating this rule will lead to a better outcome? See, for example, David Lyons, "Utility and Rights," in Theories of Rights, ed. Jeremy Waldron (Oxford: Oxford University Press, 1984).

9 I think the analogous claim could be made about contractarianism and consequentialism understood as guides to the assessment of conduct and character, but this is not my topic here.
clarity of my challenges by making them more concrete and by placing them in a familiar setting. My claim is, however, that the difficulties I discuss arise for almost any presently proposed moral conception that focuses on appraising social institutions in a contractarian or broadly consequentialist way.

III. THE PROBLEM OF NATURAL INEQUALITIES

Rawls's conception of justice has been criticized for being insensitive to natural differences among persons—to the greater health-care needs of the sickly, the greater textile needs of the tall, and the greater food needs of those with higher metabolic rates, for example. This criticism can be presented as an internal critique of Rawls's position: what quality of life each participant can attain under any institutional scheme depends not only on her share of the benefits and burdens whose distribution is regulated by this scheme (her rights and duties, income and wealth, powers and opportunities), but also on her natural (genetic) features. Prospective participants therefore have reason to want the institutional scheme that takes account of these features in an optimal way. If they use Rawls's maximin rule, for example, they have reason to insure themselves against natural handicaps by preferring schemes under which the naturally handicapped are favored in the distribution of social goods.

At first blush, this critique looks like a friendly amendment which Rawls should welcome with open arms. After all, aren't decent people agreed that (at least in a wealthy society) social institutions should favor those born with physiological handicaps? But then there are other dimensions of natural difference with regard to which the consensus would seem to go the other way: A homely physique, for example, or a disposition toward melancholy, clearly reduce quality of life; few believe, however, that justice mandates social institutions that favor persons with these natural features. Yet such institutions are exactly what prospective participants would prefer, if the internal critique of Rawls is valid.

Here is a simple illustration. Imagine a society with three social classes. Higher education is free, but scarce: 50 percent of each age cohort are admitted by lot. Those so educated will—depending on their endowments, motivation, and luck—end up in one of the top two classes. Those

---


11 Someone using the maximin rule of decision making strives to maximize her minimum pay-off. In the case at hand, prospective participants using the maximin rule would rank alternative institutional schemes by the worst social position each would produce: if one institutional scheme produces a superior bottom position than another, then the former is preferred to the latter because it secures a higher minimum.
not so educated will—depending on the same factors—end up in one of the bottom two. There are then four segments: the fully educated in the top class (FT), the fully educated in the middle class (FM), the less educated in the middle class (LM), and the less educated in the bottom class (LB). Each segment contains a quarter of the population.

Sixty percent of the population have attractive natural features (A), while the remaining 40 percent are unattractive (U). While the unattractive bear some genetic handicap, which reduces the quality of their private lives (love and friendship, self-respect), they are not sick or otherwise dysfunctional. Nor does their handicap result in any economic disadvantage: since the unattractive do not differ from the attractive in regard to talents, motivation, and luck, and are not treated any worse in the public sphere, they are proportionately represented in all four segments. There are then altogether eight relevant groups: FTA (15 percent), FTU (10 percent), FMA (15 percent), FMU (10 percent), LMA (15 percent), LMU (10 percent), LBA (15 percent), and LBU (10 percent).

The following reform is proposed. The unattractive should be entitled to automatic admission to college, and only the remaining openings should be filled by lottery. Given our stipulation, this reform would not affect economic performance (the unattractive are just as talented and motivated as the population at large), so incomes in all segments would remain the same. The only change would be in how the attractive and the unattractive distribute over the four segments: FTA (5 percent), FTU (20 percent), FMA (5 percent), FMU (20 percent), LMA (25 percent), and LBA (25 percent). Groups LMU and LBU would not exist.

Rational prospective participants using the maximin rule would, on the face of it, choose principles that support this reform. For the worst-off group under the reformed scheme is either FMU or LBA; and both of these groups are clearly better off than LBU under the unreformed scheme.

This result would be welcomed by those who believe that the institutions of a just society must compensate persons not merely for severe handicaps that affect health, but also for all other involuntary factors neg-

---

12 We can specify this notion in various ways, for example, as physical attractiveness, as being disposed toward a cheerful temper, or as some combination of the two.

13 Variants of this example work for CC conceptions that use an aggregation function other than maximin. If a CC conception specifies $f$ as averaging (conceives prospective participants as maximizing probability-weighted expectations), for instance, the example could stipulate (as seems plausible empirically) that the attractive and the unattractive differ in the extent to which education and income differentials affect their quality of life: a given advantage in education or income means more to the attractive, because it contributes more to their chances of achieving rewarding personal lives. As in the example in the text, institutions should then be designed so as to favor the unattractive in the distribution of such advantages. For the sake of simplicity, the examples are somewhat crude. The same point could be made through illustrations involving more fine-grained or even scalar dimensions, more dimensions than our three (education, class, genetic attractiveness), or correlations between class and talents or class and motivation.
atively affecting quality of life, such as homeliness, cheerlessness, and low intelligence. But most of us do not share this belief. If we are committed to the CC approach nevertheless, we must explain why prospective participants would not find it rational to adopt principles that take account of all natural handicaps and deficiencies.

Rawls does not give any such explanation. He presumes that prospective participants have three higher-order interests. Though the extent to which a person can fulfill her third higher-order interest—can “protect and advance some determinate (but unspecified) conceptions of the good over a complete life”—is clearly affected by her natural features, including attractiveness, Rawls concludes that the parties would adopt a criterion of justice (the “two principles”) that is formulated in terms of (social) primary goods among which attractiveness is not included. He leaves unexplained why the parties do not choose natural attractiveness as a primary good to be incorporated into their criterion. Why do they not


16 *Ibid.*, p. 74. The other two higher-order interests are the interests in developing and exercising the two moral powers (*ibid.*), namely, a capacity for a sense of justice and a capacity for a conception of the good. A sense of justice is the capacity to understand, to apply, and to act from the public conception of justice which characterizes the fair terms of social cooperation. Given the nature of the political conception as specifying a public basis of justification, a sense of justice also expresses a willingness, if not the desire, to act in relation to others on terms that they also can publicly endorse. . . . The capacity for a conception of the good is the capacity to form, to revise, and rationally to pursue a conception of one’s rational advantage or good. (*Ibid.*, p. 19)

17 In *Political Liberalism*, Rawls formulates his two principles of justice as follows. First principle: “Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value” (*ibid.*, p. 5). Second principle: “Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity [the opportunity principle]; and second, they are to be to the greatest benefit of the least advantaged members of society [the difference principle]” (*ibid.*, p. 6). The first principle is to have lexical priority over the second and, within the second principle, the opportunity principle is to have lexical priority over the difference principle (Rawls, *A Theory of Justice*, p. 302f.). The word “lexical” is short for “lexicographical” (*ibid.*, p. 42f.). The paradigm case of a lexical priority is this: when arranging words alphabetically, as in a lexicon, one gives priority to the first letter over the second, to the second over the third, and so on. For example, one lists all words beginning with an “a” ahead of all words beginning with a “b” regardless of the second and subsequent letters—as when “azure” is listed ahead of “baal.” Rawls conceives the priority of one principle over another along the same lines: in the comparison of basic structures, an advantage in terms of a lexically prior principle overrides any disadvantage in terms of lexically subordinate ones.
insist that poor and unattractive persons should count as worse off than equally poor but attractive ones, seeing that the former clearly face inferior prospects of fulfilling their third higher-order interest than the latter do?

One might base such an explanation on pragmatic grounds: The proposed reform would make being-naturally-unattractive socially more salient and thereby render the handicap more hurtful (imagine tax forms inquiring about ugliness as they now do about blindness). The reform would be socially divisive, because there are no clear and generally acceptable standards by reference to which citizens could be graded for natural handicap in a publicly accessible way, and also because citizens are liable to disagree about how significantly such handicaps reduce one's quality of life and about what would constitute adequate compensation. Finally, the reform might lead to perverse incentives: to persons (being suspected of) pretending to be naturally handicapped by altering their appearance and/or conduct. These pragmatic considerations cannot be lightly dismissed. Still, I wonder whether enough pragmatic arguments can be found to convince prospective participants that they should not take account of any natural inequalities beyond the uncontroversial ones involving serious physical handicaps. Moreover, such pragmatic arguments seem to involve a distortion of the grounds on which it is widely believed that our social institutions need not favor the homely and the cheerless in order to compensate them for their handicap: people see no moral reason why citizens should have to compensate one another for such misfortunes.

For these reasons, I have previously tried to explain in a principled way why only social goods and their distributions should be deemed relevant to the comparative assessment of alternative institutional schemes. This strategy now strikes me as ad hoc, however: in order to derive the contract we want, we would have to tell the hypothetical contracting parties what is blatantly false—namely, that natural (genetic) differences among persons do not, under any circumstances, affect their quality of life. Without this disinformation, rational parties would demand that, insofar as quality of life is affected by differentially distributed natural features, a basic structure should be considered more or less just depending on how well the distribution of social goods it generates fits the distribution of natural features—with fit measured by reference to the distribution of overall quality of life that arises from the superimposition of the two (the social upon the natural) distributions.

---

18 See Thomas W. Pogge, *Realizing Rawls* (Ithaca: Cornell University Press, 1989), pp. 44-47, 114-16, 183-89. I also show there that CC conceptions meeting this constraint (I called them "semiconsequentialist") can accommodate our conviction that the distribution of health care should favor those with greater health problems by defining the relevant social good as "access to health care when needed."

19 This demand would generally lead to institutions favoring the naturally handicapped, even if the parties do not employ the maximin rule. But there could be exceptions. It might
IV. The Problem of Group Inequities

While the CC approach would seem to be too sensitive to certain natural inequalities, it may not be sensitive enough to group inequities. CC conceptions assess institutions in terms of their effects upon individual persons: if two alternative institutional schemes would generate identical distributional profiles, then they are equally just. Intuitively we are convinced, however, that the justice of an institutional scheme also depends on the extent to which it allows certain involuntary and morally arbitrary factors (such as gender, race, the status of one’s parents) to influence who ends up where on the distributional scales—the greater this influence, the more unjust (ceteris paribus) the scheme. This moral conviction cannot be reaffirmed within a CC conception. A prospective participant cares solely about the share she can expect under each scheme (that is, about the distributional profile it produces) and not about how these shares (distributional profiles) are produced. Rawls, of course, has a principle of fair equality of opportunity that forbids some group inequities (those based on initial social class—the status of one’s parents); but I see no way of supporting anything like his opportunity principle through the original position.

Let me give an illustration which involves the choice of a basic structure for a relatively simple society. We take for granted that Rawls's first principle and formal equality of opportunity must be satisfied. Applying the difference principle within the remaining option space, we get a society with two classes: the more privileged (P) and the less privileged (L). As children, persons belong to the class of their parents; as adults, their class is determined by their professional position. Thus, there are four possibilities: PP (born among the privileged and holding a privileged position as an adult), LP, PL, and LL. Talents, motivation, and luck are roughly equally distributed among children of the two classes. The top-flight education necessary for a privileged position is expensive, however, and only very few less-privileged parents make the sacrifices necessary by argued, for instance, that lack of access to higher education generally reduces the quality of life of gifted persons more dramatically than, and (other things equal) below, that of the less gifted. Contracting parties presumed to know this would then have reason to prefer institutions under which the naturally better endowed are favored in the distribution of the social good education.

20 The principle requires that “those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system, that is, irrespective of the income class into which they were born” (Rawls, A Theory of Justice, p. 73, cf. pp. 275, 301). Does fair equality of opportunity rule out group inequities involving race and gender as well? Rawls does not incorporate this constraint—perhaps because he believes that formal equality of opportunity (together with the first principle of justice) suffices to wash out race-induced and gender-induced inequalities over time.

21 This illustration uses class-induced inequalities in order to stay as close as possible to Rawls's formulation of fair equality of opportunity.
to give their children such an education. This basic structure (S₁) engenders the following income distribution:

<table>
<thead>
<tr>
<th></th>
<th>Average Annual Income over Lifetime^{22}</th>
<th>Percentage of Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP₁</td>
<td>180</td>
<td>9%</td>
</tr>
<tr>
<td>LP₁</td>
<td>170</td>
<td>1%</td>
</tr>
<tr>
<td>PL₁</td>
<td>110</td>
<td>1%</td>
</tr>
<tr>
<td>LL₁</td>
<td>100</td>
<td>89%</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

S₁ violates fair equality of opportunity. Those born into less privileged families—though they exhibit the same distribution of natural talents, motivation, and luck—have much less of a chance (one in ninety) to attain a privileged position than those born into privileged families (nine in ten). Nevertheless, S₁ satisfies economic efficiency and the difference principle: enough children are born into privileged families and well educated at their parents' expense to ensure that leadership positions (including those that require not only a superior education but also exceptional talents and motivation) are filled well.

Fair equality of opportunity could be achieved by reforming the education system so that the same limited number of superior educations are provided at public expense and distributed through a lottery scheme. This reform will have some economic cost, as people would have less of an incentive to work hard to attain privileged positions: one advantage of occupying such a position—the freedom to guarantee a superior education to one's children—would be absent.^{23} Applying the difference principle within this much narrower option-space (readjusting tax rates, etc.), we get an alternative basic structure (S₂) which engenders the following income distribution:

^{22} I have designed the example so that the inequality in average annual incomes is very modest indeed (compared to present realities), in order to reassure the reader that the fair value of the political liberties is really being maintained and that the difference principle is really satisfied.

^{23} It may be said that I have not told a plausible story showing that the maintenance of fair equality of opportunity can have net costs; if so, another such story can be substituted. One might further object that the difficulty I indicate cannot arise at all, because maintaining fair equality of opportunity has important benefits that will always outweigh any costs. This strikes me as unlikely. Even if it were a fact, however, it would not support the inherent moral importance that Rawls (and most of us) attach to the avoidance of class inequalities. Fair equality of opportunity would not be demanded in its own right by justice (by prudent prospective participants), but would come about as a merely contingent by-product of the realization of economic efficiency or the difference principle. This scenario would confirm, not defeat, my claim that CC conceptions cannot explain why fair equality of opportunity should have any moral standing in and of itself. I am indebted to Ling Tong for this objection and subsequent discussion.
The reform eliminates the intuitive injustice Rawls meant to rule out with his requirement of fair equality of opportunity. It ensures that those similarly endowed and motivated have similar chances of success irrespective of their initial social class. But why in the world should rational prospective participants care about this effect of the reform? Who benefits from the elimination of class-induced inequalities?

Rawls is specially constrained in answering this question: given that his parties use the maximin rule, he must be able to say that the worst off benefit. Now the worst off under both basic structures are the members of LL. Comparing across schemes, we must conclude that the members of LL₂ are surely worse off than the members of LL₁: they are just as much excluded from a superior education and their income is lower by 10 percent.

One might say that under S₂ at least there are fewer persons in LL. This will not help, however, because group PL₂ is also worse off than LL₁; thus, the union of LL₂ and PL₂ is both larger and worse off on average than LL₁.

It might also be said that prospective participants would care not about lifetime income (ex post), but about expected lifetime income (ex ante). But this suggestion does not work either: it entails that prospective participants would not care how unequally given benefits and burdens are distributed so long as the distribution is based on a fair neonatal lottery (so that everyone has the same expected pay-off); this is absurd. Moreover, it does not even support the desired conclusion. For the expected lifetime income of those born less privileged under S₁ is \((170 \times 1\% + 100 \times 89\%) + 90\% = 100.78\). Under S₂ it would be worse, namely \((153 \times 9\% + 90 \times 81\%) + 90\% = 96.30\).

Another tack is to say that group LL₂ is better off than LL₁ because they live in a society without class inequities, but this just begs the ques-

---

24 To ensure comparability between this and the previous table, let us assume that any expenditures for education are counted not as income of those adults who pay them (parents under S₁, and taxpayers under S₂), but as income of the youngsters who are receiving the educational benefit.

tion: a society should maintain fair equality of opportunity because that way its members will enjoy the benefit of living in a society that maintains fair equality of opportunity. The stipulation that the parties to the social contract care very much that life prospects not be influenced by initial social class simply builds the desired conclusion into the premises. I will say more about such "gimmicky"—as Robert Nozick likes to say—argumentative strategies below (Section VC, penultimate paragraph).

Perhaps this stipulation can be defended in a more empirical way, however. Perhaps it is true that the members of LL$_2$ derive considerable comfort and self-respect from the knowledge that others, born into the same class as themselves, do and will receive a superior education and lead a life of privilege. It is clear why the members of LP$_2$—and today's beneficiaries of affirmative action—should like to think so. Nevertheless, it is an empirical question: Are the lives of ordinary workers from working-class backgrounds really improved by the knowledge that others from a working-class background succeed in becoming doctors, lawyers, and executives? Do black cleaning ladies, for example, really gain (or might they even lose) self-respect from knowing that other blacks, and women, ascend through the glass ceiling? And is the benefit conferred by this knowledge sufficient to support a preference for LL$_2$ over LL$_1$?

CC conceptions other than Rawls's focus not only on the least advantaged but on other social groups as well. Still they all face the same difficulty and challenge. To be able to reaffirm the deep and widespread commitment to fair equality of opportunity, a CC conception must show how the required elimination of group inequities benefits individuals and how the benefits for those who gain outweigh both (a) the burdens on those who lose and (b) the general burdens on the economy. One can see most clearly how my example is troublesome for all CC conceptions, if one divides the populations under each of our two basic structures into percentiles and then compares down the list, percentile by percentile, across the two schemes. While there is no difference in education and relative incomes, absolute incomes are smaller under S$_2$ at each and every percentile.

It may be thought that my discussion of the moral comparison between S$_1$ and S$_2$ should carry no weight because I have freely stipulated the

---

27 The answer may well depend on whether they have offspring or not. Those who do will probably care about their children's and grandchildren's chances of occupying privileged positions. Interestingly, this offspring argument could work only for race and class, but not for gender: in the first two cases, fair equality of opportunity benefits the offspring of the worst off—those from the lowest class or of the currently most disfavored race. Fair equality of opportunity for women, by contrast, benefits one's offspring only if one's offspring are predominantly female. Assuming zero correlation between social disadvantage and predominantly female offspring, and assuming also that disadvantaged mothers do not care more for their daughters than for their sons, we cannot then argue that fair equality of opportunity for women is in the interest of the worst off because it benefits their offspring.
empirical facts that lead to the numbers presented in the two tables to suit my convenience. But then Rawls's thought experiment is supposed to give plausible guidance for institutional comparisons under all reasonably possible empirical assumptions about the net costs of maintaining fair equality of opportunity. This universal claim can be defeated by producing one set of assumptions under which it does not give plausible guidance. As to whether my empirical assumptions are reasonably possible for at least some societal context, see footnote 23.

A last remark: Even if, as I have tried to show, CC conceptions fail to reaffirm our convictions about how social institutions should accommodate natural inequalities, and also fail to reaffirm our commitment to fair equality of opportunity, their adherents may still claim that the fault lies not with the CC approach, but with our convictions. I will come back to this point in Section VI.

V. THE PROBLEM OF NONCOMPLIANCE

I have so far briefly discussed two problems that arise within the CC approach: the problem of natural inequalities and the problem of group inequities. The third and final problem—which is actually a set of related difficulties involving noncompliance—will be discussed at greater length. If Rawls himself has not run up against these difficulties, it is perhaps because he has largely and increasingly confined his attention to a rather ideal version of ideal theory. In this first approximation to a full theory of justice, the choice, specification, and application of a criterion for the comparative assessment of institutional schemes is based upon the unrealistic assumption that any chosen ground rules will be fully complied with by every participant at all times. The difficulties I will now discuss arise in what Rawls calls "partial-compliance theory," where the comparative assessment of sets of ground rules begins to take account of the extent to which each set of rules would actually engender compliance. The rules prospective participants would prefer given universal compliance may not be the ones they would prefer once they consider the degree of compliance each set of rules would actually engender. Rawls has largely neglected this important topic,28 and the advocates of competing CC conceptions have followed him in this.

Rawls holds that prospective participants (characterized by the three higher-order interests) would be primarily concerned to enjoy certain basic freedoms, such as liberty of conscience, freedom of movement, and

28 He envisions a second stage, after ideal theory, in which the preferred conception of justice and the ground rules it selects are tested for stability. This test is conceived as an either-or affair and is never integrated with the results of the first stage. Rawls never aims for an overall appraisal of alternative institutional schemes, an appraisal that would take account of how each would actually work in practice (degree of stability, incidence of non-compliance, etc.). See, for instance, his Political Liberalism, pp. 64, 133f., 140f.
integrity of the person. Guaranteeing the enjoyment of these freedoms is then the preeminent task of social institutions, as laid down in his first principle of justice.

Once we bear in mind that full compliance with legal norms is not a universal feature of human societies, it is clear that prospective participants would not settle for a written commitment to these freedoms in assorted legal documents. They care not about legal rights as such, but about their objects: about whether they will actually enjoy the freedoms in question. Legal enumeration is a means to this end—perhaps a necessary means under modern conditions, but certainly not a sufficient one, as is demonstrated by various "showcase" constitutions around the world. Rawls himself has made this point on occasion: "Whether the basic structure guarantees equal liberty of conscience, or freedom of thought, is settled by the content of the rights and liberties defined by the institutions of the basic structure and how they are actually interpreted and enforced." What matters in regard to the first principle, then, is not whether a citizen is legally entitled to certain freedoms on paper, but whether she securely enjoys these freedoms themselves.

While it is quite clear—textually and systematically—that this is his view, Rawls has paid far more attention to the scope of these freedoms than to the security of their enjoyment. He thereby bypasses a crucial problem for the lexical priority he postulates for his first principle. The problem is this. There is no realistic hope for a human society in which basic freedoms are absolutely secure, yet Rawls wants it to be realistically possible for his first principle to be fully satisfied. Therefore, this principle must be read as requiring that the basic freedoms it covers merely be reasonably secure. The principle will then have a clear meaning, will fulfill its intended function within a standard for assessing social institutions, only if, for each basic freedom, some security threshold is specified—a threshold one can refer to in order to judge whether the enjoyment of some particular basic freedom by some particular participant under some particular institutional scheme is reasonably secure or not. Without such thresholds, the theory is overly vague. But with such thresholds it involves a highly implausible discontinuity. Insecure enjoyment below the threshold is an injustice of the first order, and societies must enhance security through institutional reforms regardless of the economic cost. Insecure enjoyment above the threshold is a matter of moral indifference,

29 I have in mind constitutions whose words are highly protective of civil rights and democratic procedures but bear little or no relation to the realities in the relevant country. The successive constitutions of the USSR are a clear-cut example, as are many constitutions in present and former Third World dictatorships and one-party regimes.


31 Such a security threshold would presumably be defined in terms of frequency-based probabilities. Here one may need to take account of various personal characteristics. Thus, it is quite possible that a certain basic freedom is reasonably secure for some citizens (defined by wealth, geography, race, gender, or age, for example) but not for others.
and societies need not achieve greater security even when doing so is cheap and easy.\textsuperscript{32} However the requisite security thresholds may be defined, it is unclear why we, or prospective participants, should attach such tremendous importance to them.

Rawls does not specify or envisage security thresholds, and so his conception faces the problem of excessive vagueness. Every institutional scheme leaves citizens' basic freedoms insecure to some extent, and we have no way of judging whether it nevertheless satisfies the first principle or not. This is not merely a philosophical point, but also an eminently practical and political one. What must a society be like to secure the freedom and integrity of the person,\textsuperscript{33} for example? What if opponents of government policies are occasionally roughed up by police when they demonstrate for their convictions; how often can this happen without contradicting the justice of a society's basic structure? What if 20 percent of all women are raped, or 1 percent of all black men murdered, within ten years of reaching age sixteen? Without answers to such questions, the consensus around Rawls's first principle of justice is in large part deceptive: it is unclear what it would mean to apply or realize this principle.

The partial-compliance problem I have just described is specific to Rawls's CC conception—it arises from his postulate of the first principle as both satiable and lexically prior. The related difficulties I shall now discuss arise for any CC conception. Citizens' quality of life—however conceived—may be affected by the government and its officials on the one hand and by other (individual and collective) agents on the other. Citizens are roughed up by the police and by gangsters, killed by executioners and by drunk drivers, imprisoned in jails and in basements, tortured by secret services and by private death squads.

In aiming for institutions that minimize such abuses, a CC conception will treat both kinds of abuses as on a par; that is, it will be insensitive to how given effects of social institutions come about. Taking the perspective of a prospective participant, one will consider social institutions under which one's physical integrity is endangered by crime to be, other things equal, just as bad as ones under which one's physical integrity is equally endangered through government officials. And this leads to the final problem I shall discuss: a CC conception of justice may require social institutions that permit and even encourage government officials to invade citizens' basic freedoms or fundamental interests, if these institutions minimize such invasions overall. It is quite likely that such trade-offs

\textsuperscript{32} The discontinuity could be somewhat mitigated: justice demands, one might say, that the basic liberties be made more secure than the first principle mandates, and this demand is on a par with, and to be balanced against, the demands of either the opportunity principle or the difference principle.

\textsuperscript{33} This is said to include that one be free "from psychological oppression and physical assault." See John Rawls, "Reply to Alexander and Musgrave," Quarterly Journal of Economics, vol. 88, no. 4 (November 1974), p. 640.
would be worthwhile in our own social world. I will illustrate this claim in regard to three topics within the criminal law.

A. The definition of criminal offenses

For the first topic, we can begin with an example Rawls gives in a brief foray into partial-compliance theory:

Suppose that, aroused by sharp religious antagonisms, members of rival sects are collecting weapons and forming armed bands in preparation for civil strife. Confronted with this situation, the government may enact a statute forbidding the possession of firearms. . . . And the law may hold that sufficient evidence for conviction is that the weapons are found in the defendant’s house or property, unless he can establish that they were put there by another. Except for this proviso, the absence of intent and knowledge of possession, and conformity to reasonable standards of care, are declared irrelevant. . . . Now although this statute trespasses upon the precept ought implies can it might be accepted by the representative citizen as a lesser loss of liberty, at least if the penalties imposed are not too severe. . . . Citizens may affirm the law as the lesser of two evils, resigning themselves to the fact that while they may be held guilty for things they have not done, the risks to their liberty on any other course would be worse.34

Rawls refers to cases of this sort as “extreme eventualities,”35 but they are not so far from our reality today. With firearms being used in the United States to wound and kill tens of thousands of persons each year, it may certainly seem that it would be, on balance, a gain for the representative citizen, if possession of firearms were prohibited with the kind of strict-liability statute Rawls contemplates.

What is at stake in his hypothetical example is the venerable precept that mens rea (a “guilty mind”) ought to be a necessary condition of criminality and hence a prerequisite for any severe punishment. This precept does not require that the defendant must have intended or foreseen the harm she is to be held legally accountable for, or a significant probability of such harm (as in recklessness). Even merely negligent conduct is normally deemed sufficient to satisfy the mens rea condition: agents can act without negligence. The problem with the strict-liability law Rawls entertains is that even those who take great care to abide by it, and thus

34 Rawls, A Theory of Justice, p. 242. The quote shows nicely how his commitment to the prospective-participant perspective attenuates the uplifting idea (chosen for the cover of the paperback) that “[e]ach person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override” (ibid., p. 3).
are not even ever so slightly negligent, may be convicted under it—for example, a woman who searched her house and grounds frequently for weapons and, though she knows someone else must have put the weapon on her property, cannot prove that this is so.

Once we countenance the broadly consequentialist reasoning suggested by a prospective-participant perspective, we are likely to find ourselves attracted to strict-liability laws in cases where criminal conduct with mens rea can be significantly reduced through the extra deterrence that comes from the general knowledge that pretending to lack mens rea will be to no avail.\footnote{One may also be attracted to strict-liability laws that criminalize harms through clumsiness, mental confusion, and the like, where the main benefit is not deterrence but prevention of like harms caused by the same offender.} Cases in which such pretensions are hard to refute are many: speeders often claim that their cruise control or speedometer malfunctioned; drunk drivers claim that they had no idea the punch contained (so much) alcohol; drug runners claim not to know how the drugs got into their luggage; rapists claim that they honestly thought their victim had consented; and killers claim that their victim's death was accidental (for example, "during rough sex"). In some of these cases, of course, the extra deterrence gained may not be worth the cost in terms of having to find persons guilty who genuinely lacked mens rea and who constitute no abnormal danger to others. In other cases, however, it will be worth it—it will be rational for prospective participants to run a certain risk of being convicted of some crime without mens rea in exchange for reducing more significantly their risk of being harmed through crimes of this type.

B. The apprehension and judgment of suspects

We are quite firmly convinced that it is much more important, morally, to avoid convicting the innocent than to avoid acquitting the guilty. We impose very high standards of evidence ("beyond a reasonable doubt") in criminal cases, especially where severe punishments are at stake. We categorically reject imposing lower standards ("clear and convincing evidence," say, or "preponderance of evidence") that might significantly reduce the number of false acquittals at the cost of only a few more false convictions.\footnote{There is a similar categorical commitment to various other procedural safeguards—for example, to the prohibition of double jeopardy (Fifth Amendment to the U.S. Constitution) and to the prohibitions on informing the jury about the accused's prior convictions.} Once we adopt the institutional view, however, such a categorical stance is hard to justify: it may lead to a society in which many crimes are committed by persons who have been falsely acquitted and many crimes are encouraged by the well-grounded belief that, with a resourceful lawyer, one has an excellent chance of escaping conviction even if caught. Prudent prospective participants would not impose this requirement for social worlds in which the additional crimes engendered
by the higher standard of evidence do more to reduce the citizens' quality of life than the additional false convictions engendered by the lower standard.  

Analogous reflections apply to apprehension work by the police. The incidence of crime may well be higher when the police must not conduct so-called unreasonable searches and seizures and when any evidence so acquired is inadmissible in court: as a consequence of this constraint, some criminals are never caught and others cannot be convicted. Thus, even if freedom from unreasonable searches and seizures counts as a basic freedom, it is still arguable that the objects of all basic freedoms would on the whole be better protected if the relevant restraints upon the police were somewhat looser.

C. The punishment of convicted criminals

Our current punitive practices are still, by and large, informed by the value of proportionality: the severity of punishment  should match the culpability of the offender, with culpability a function, mainly, of the magnitude of the harm (or risk of harm) inflicted and of the offender's state of mind. Prospective participants, by contrast, would want punitive practices to be informed by the idea that harms overall should be kept to a minimum; hence, they would want the severity of punishments to be justified by their (negative) impact on the rate of incidence of the offense. According to CC conceptions, the punishment (type) fits the crime (type) when there exists no alternative punishment offering either (1) a decrease

---

38 In nonideal theory, we must settle what kinds of abridgments of basic liberties are called for when it is impossible to safeguard all basic liberties for all. What sorts of abridgments of the basic rights of the accused would the contracting parties countenance for the sake of reducing to a certain extent the excessive incidence at a certain level of crimes of a certain kind? It may be said that we are not (yet) in a nonideal situation, that current crime rates are not excessive, do not undermine the claim that the basic right to physical integrity is fulfilled for all; and it may then be concluded that the first principle prohibits the lower standards of evidence, assuming it mandates evidence beyond a reasonable doubt in ideal theory. The issue cannot be so easily avoided, however. For we must also justify how, in ideal theory, the basic liberties are carved out of a much wider domain of candidate freedoms, how one particular design of a universalizable scheme of basic liberties is chosen. This involves showing that the parties would indeed consider the existing insecurity of the right to physical integrity consistent with the first principle being fully satisfied, thereby protecting the basic rights of the accused from abridgment. At some level, the parties must trade off between citizens' freedom from being innocently convicted and citizens' freedom from being innocently victimized by crime. For some reflections about this trade-off, see Jeffrey Reiman and Ernest van den Haag, "On the Common Saying that It Is Better that Ten Guilty Persons Escape than that One Innocent Suffer: Pro and Con," Social Philosophy and Policy, vol. 7, no. 2 (Spring 1990).

39 A punishment is more or less severe (as I use this term) the more or less it hurts the person punished. In many jurisdictions, rich persons receive larger fines than poor ones for the same offense. This does not mean that the severity of punishment varies with wealth. To the contrary, those differentials accommodate the concern that severity should not vary with wealth. Only a larger fine can punish a rich offender as severely as a smaller fine punishes a poor one.
in expected punishment costs that outweighs any decrease in expected crime-reduction benefits, or (2) an increase in expected crime-reduction benefits that outweighs any increase in expected punishment costs. Consider two types of crime, which, in accordance with the prevailing proportionality standard, draw roughly equivalent punishments. Suppose their rates of incidence differ significantly in how elastic they are vis-à-vis the severity of punishment: the rate of incidence of the first crime would rise/drop by several percentage points for every 1 percent decrease/increase in punishment (jail time, say) from the prevailing level, while the rate of incidence of the second would be affected only minimally if at all. Rational prospective participants would want prevailing punishments to be revised so that the first crime is punished more severely than the second.

I will here focus on high-elasticity crimes, with respect to which more severe punishments work especially well—generally by being especially effective in deterring potential offenders and/or in taking would-be repeat offenders out of circulation. To avoid the problem of heterogeneity of costs and benefits, I consider the death penalty for offenses that result in deaths. Here we might ask whether prospective participants, guided solely by rational prudence, would mandate the death penalty for murder if this minimizes premature deaths overall. Let me entertain, however, another death-penalty proposal which, though in a sense perfectly plausible, is also rather wild. In the U.S., drunk driving causes at least ten thousand premature deaths per year. A practice pursuant to which

---

40 The word “expected” is to serve as a reminder that costs and benefits have to be assessed from the perspective specified by the relevant CC conception. If the hypothetical contracting parties use the maximin rule, for example, they define the expected net benefit of a punishment scheme in terms of its impact on the quality of life of the worst off (in comparison with the quality of life that the worst off under alternative schemes would enjoy).  
41 I assume that, for CC conceptions, the imposition of punishments will involve expected costs in terms of quality of life. (Otherwise prospective participants would want equal, namely maximal, punishments for all crimes, and the prospective-participant perspective would be in even greater trouble than I allege.) This assumption is plausible, because prospective participants run some risk of being (correctly or falsely) convicted and subjected to any punishment they authorize, and may have to pay a share of the cost of inflicting punishments. It does not follow, however, that more severe punishment for a particular crime will entail greater cost. When elasticity is high, the opposite may be true: though each person convicted of the crime receives more jail time (say), the crime is less frequent and fewer persons are punished for it. Since the increase in the first factor may be overwhelmed by the decrease in the second, total jail time imposed for the crime may actually decline.  
42 I assume, as seems inevitable, that any reasonable CC conception counts involuntary premature death as a major evil.  
43 The figures provided by the American Automobile Association are as follows: The annual death toll on U.S. highways hovers around forty-three thousand. About half these deaths are caused by accidents in which at least one intoxicated person is relevantly involved (i.e., involved not merely as passenger, but as driver or pedestrian). Of course, in some of these cases there is no drunk driver, and in others the intoxication of one or more drivers did not contribute to the accident. There is some dispute over what the death toll on U.S. highways would be if there were no drunk driving, but the lowest estimate seems to be that ten thousand premature deaths (plus one hundred thousand injuries and one billion dollars in property damage) would be avoided every year. For these figures, see H. Laurence Ross, Confronting Drunk Driving (New Haven: Yale University Press, 1992), pp. 35-38. Over
some two hundred of the worst offenders are executed annually would presumably have a tremendous deterrent effect and thus would greatly reduce the frequency of the offense. No one knows how great this reduction would be. Still, I am quite confident that it would vastly exceed 2 percent and that the total number of premature deaths would decline.\(^44\) Here we have, then, a draconian proposal that looks quite mad and yet would seem entirely plausible under institutional viewing.

We can render the "reform" even madder: to maximize the deterrent effect, the law should be that anyone repeatedly caught heavily intoxicated at the wheel of an automobile is subject to the death penalty, but that only a fraction of those so caught, selected by lottery, will actually be executed. The percentage to be executed is adjusted so as to minimize the overall loss of lives: executed offenders plus vehicular-homicide victims plus killed police officers.\(^45\) This idea of an execution lottery, by the way, goes back to the ancient Romans, who punished cowardliness on the part of whole armed units with a one-in-ten chance of execution. We still have the word—"to decimate"—but the practice has long since gone out of fashion.

Can adherents of the CC approach stick to their respective CC conceptions of justice and yet consistently reject the mad reform? They might propose that the optimal punishment for drunk driving lies somewhere between our current practices and the death penalty regime I have entertained: repeat offenders should, perhaps, receive multi-year jail sentences. This proposal runs into two difficulties, however. First, it does not detach the CC approach from the view that the death-penalty regime, though perhaps not optimal, is at least morally preferable to the status quo. This view is a considerable embarrassment. Second, it is doubtful whether the proposal is empirically sustainable. For suppose multi-year jail sentences would reduce drunk-driving fatalities to an annual rate of three thousand. It would seem likely, once again, that the execution of sixty of the worst offenders every year could further reduce this annual rate by well over 2 percent.\(^46\)

\(^44\) We must consider that the projected "reform" would increase the killing of police officers: notorious drunk drivers, when flagged down, may believe that they have little left to lose. I owe this point to Michael Levin.

\(^45\) Note that the argument for using the death penalty here extends, for purposes of assessing social institutions, the same concern and respect to all citizens—criminals and innocents alike. It counts the execution of a guilty person as no less bad than the vehicular death of an innocent. A higher execution rate would be called for, if the lives of innocents were weighted more heavily.

\(^46\) For the record, let me say that—though I am unconvinced by the draconian practices that (I have argued) are entailed by the CC approach—I do believe that punishments for grave and repeated drunk-driving offenses ought to be significantly more severe than they currently are and than they ought to be according to Douglas Husak, "Is Drunk Driving a Serious Offense?" *Philosophy and Public Affairs*, vol. 23, no. 1 (Winter 1994), pp. 52-73. But then this essay is not about appropriate punishments for the various drunk-driving offenses.
Adherents of the CC approach might also claim that deaths by execution are worse than traffic deaths. But how can this claim be supported? Morally, the opposite would look more plausible: most of those executed for drunk driving would have inflicted serious harms or risks of harm on others, while those killed by drunk drivers are mostly innocents. Prospective participants, too, might have the opposite valuation: one can gain a good deal of security against the risk of being executed by leaving one's car at home when going out for entertainment; but one cannot, without enormous losses in quality of life, attain the same degree of security against the risk of being killed by a drunk driver. Here it might be said that a premature death with social stigmatization is worse than an otherwise similar death without. But how much worse? Is it so much worse as to make it rational to prefer our current one in four hundred chance of premature death through drunk driving to a one in one thousand chance, say, of death by execution for drunk driving?

A third strategy for blocking the mad conclusion stipulates that prospective participants are strongly averse to living in a society with extensive use of the death penalty. This stipulation is entirely ad hoc, however, unless one can connect the stipulated aversion to interests that members of modern societies can plausibly be assumed to have. Without such a connection, the stipulation tends to trivialize and undermine the contractarian thought experiment: in this manner we can, after all, derive anything we like simply by stipulating that the hypothetical contractors like it as well. The core idea of modern contractarianism is to subject our moral judgments to an independent test by comparing them to the prudential judgments that rational persons would reach in a fair initial situation. And why should it be prudent to prefer a greater risk of premature death through drunk driving to a significantly smaller risk of premature death through execution for drunk driving?

While it is difficult, if not impossible, for CC conceptions to avoid endorsing the mad reform, it can easily be rejected on interactional viewing. Here we can distinguish, in good old deontological fashion, between deaths that we, acting through the state, bring about and deaths that merely come about as a consequence of what we, through legislation, do. "Excess" homicide deaths are, on this account, still morally regrettable, but we simply must not kill even a few to save a greater number. (This claim is often embellished by the Kantian thought that, insofar as we impose superproportional punishments in order to sustain deterrence, we are illicitly using convicted criminals merely as means.) The paradoxical implications of the CC approach are then connected to the fact that it focuses not on actors (the people or the government), but on the societal basic structure which produces various good and bad events in particular frequencies and is to be held morally responsible for them: for execu-

47 A good many victims of drunk driving are, of course, victims of their own drunk driving.
tions as well as for traffic deaths, for punishments as well as for crimes, for taxes as well as for bankruptcies.

D. Summary

With each of our three criminal-law topics—defining offenses, apprehending suspects, and establishing punishments—we feel, I believe, strong moral resistance to the idea that our practices should be settled by a prospective-participant perspective. This becomes quite clear when we look at how the "reforms" suggested by institutional viewing might combine once we consider all three topics together: imagine a more extensive and swifter use of the death penalty in homicide cases coupled with somewhat lower standards of evidence; or think of backing a strict-liability criminal statute with the death penalty. Of course, such "reforms" would increase the execution of innocents; but, their proponents will tell us, any penal system involves the punishment of some innocents, and, if it provides for the death penalty, the execution of some innocents. Moreover, why is it worse for innocents to be punished than for innocents to suffer an equivalent harm in some other way? Formulated from a prospective-participant perspective: Why not run a small risk of being innocently executed in exchange for reducing, much more significantly, the risk of dying prematurely in other ways?

While the criminal law in the countries of the developed West is clearly concerned to enhance citizens' quality of life, it is also informed by a plurality of further, conflicting and competing values and principles which, from a prospective-participant perspective, deserve no independent weight. Examining our three topics in the criminal law, it does not appear that these further principles are in decline. While strict liability is important, and increasingly so, in tort law, there is still great reluctance to use it within the criminal law. In the second topic, the picture is similar: the standards of proof protecting criminal defendants ("beyond a reasonable doubt") are virtually unchallenged, and the protections of suspects have even been significantly expanded over the past thirty-five years, despite the evident negative impact of this expansion upon crime rates. In the third topic, the prospective-participant perspective seems to have made some inroads. Nevertheless, our penal institutions on the whole

48 In the long passage I quoted from Rawls at the beginning of Section VA, he writes that the representative citizen can accept the strict-liability statute "at least if the penalties imposed are not too severe." But whence this limitation? The harms at stake ("civil strife") are very serious, and, if only the death penalty can avert them, then it is possible that the death penalty is called for to protect optimally the representative citizen's right to life.

49 For the United States, witness, for example, punishments of so-called "accessories after the fact" and punishments under the Racketeer Influenced and Corrupt Organizations (RICO) statute. In both cases the punishment for given criminal conduct can become much more severe on account of circumstances beyond the offender's knowledge and control (an element of strict liability) and can become grossly disproportionate to the offender's culpability.
diverge dramatically from what a plausible CC conception of social justice would presumably require, and we seem to be quite strongly committed, morally, to maintaining this divergence.

It may be said, of course, that there is no such divergence, that our criminal law does correspond, reasonably closely, to the hypothetical preferences of a prospective participant. More specifically, it may be said that the penal reforms we have looked at would have negative collateral effects outside the penal system and thus would lose their plausibility once we compare, as we certainly should, entire institutional schemes in light of their likely overall effects. I have not considered this objection for two reasons. I do not believe it has a chance of being true, and I do not believe that complex empirical considerations reconciling our current practices with a prospective-participant perspective can possibly lie behind our strong and unhesitating moral condemnation of the penal “reforms” I have entertained. Most people think that imposing the death penalty even for the most notorious drunk drivers is absurd, and they think this without having remotely studied the relevant statistics, thought through deterrence effects, or the like. This firm conviction may well be a deontological hang-over of interactional viewing, which has long been in decline in political philosophy but appears to be alive and well in the subfields of political philosophy connected with the criminal law. This hang-over leads us to think of ourselves as much more responsible, morally, for harms we impose through the state than for harms that merely come about as a consequence of our social institutions. We say: “People driving while intoxicated surely deserve punishment; but they do not deserve to die for it. So we simply cannot kill them, no matter what good this might do.” When pressed, we readily acknowledge that those who die on our roads because we fail to impose a death penalty for drunk driving do not deserve to die either. We do not feel morally responsible for these excess road deaths, however, or at least we feel much less responsible than we would for the proposed executions.

VI. Conclusion

Political philosophy today stands before the following tetralemma:
1. We can bite the bullet by following the CC approach, which seems so attractive under institutional viewing, wherever it may lead. But this option runs into the problems discussed in Sections III through V: it could well force us to accept social institutions that compensate persons for unattractiveness, institutions that violate fair equality of opportunity, and various draconian revisions of our criminal-law system.
2. We can abandon institutional viewing, go back to the old ways of interactional moral analysis, and then make morally significant the distinctions between what we, as a society, intend and what we merely foresee—and between what we, through the state, do and what merely
comes about as a consequence of our legislation. But this option abandons the great moral and explanatory advantages (briefly sketched in Sections I and II) of institutional viewing in the modern world.

3. We can try to find some plausible rationale for delineating the spheres of institutional and interactional viewing so as to block the difficulties. Thus, one might say, as Thomas Nagel has suggested to me, that in the criminal-law domain the burden of intending and producing a particular outcome falls on particular persons—police officers, judges, executioners—who must enforce the optimal ground rules against particular defendants. This is not the case in the economy, where people can be poor or unemployed without anyone having intended and produced this outcome. That is why it is appropriate to view legal and economic institutions differently: the former interactionally and the latter institutionally. But this option, whatever its merits in regard to the set of difficulties discussed in Section V, would not seem to work for the problems discussed in Sections III and IV.

4. We can search for a principled way of reconciling our considered convictions about natural inequalities, group inequities, and the criminal-law domain with institutional viewing by abandoning the CC approach to the moral assessment of social institutions. To make this work, one could consider a backward-looking account à la Nozick, which assesses an institutional scheme through an appraisal of its history. One could also aim for a forward-looking account that breaks out of the $J = f(Q_t)$ mold, as Thomas Scanlon seems to do—though so far his conception is still rather sketchy.\(^5\)

On the face of it, a resolution exemplifying the third or fourth option (or some combination of these) would seem most appealing. The task is, however, to formulate a plausible resolution of this sort in detail. This is one project, I think, that political philosophy should now undertake.

*Philosophy, Columbia University*

---