

7 G. A. Cohen’s critique of the original position

David Estlund

7.1 Introduction

G. A. Cohen seeks to “rescue” justice from, primarily, John Rawls. His challenge to Rawlsian philosophy is widely thought to be one of the most powerful we have. Cohen’s writings on this topic over many years culminate in his densely argued volume, *Rescuing Justice and Equality*.¹ While no thorough evaluation is possible here, I will tentatively defend Rawls, but with reservations.

Rawls’s reliance on the original position is at the heart of Cohen’s critique, with Cohen charging that this method makes the principles of justice depend on non-moral facts in ways that are fundamentally misguided. Cohen pursues both metaethical and normative prongs of this attack, and I shall be looking at both. On Cohen’s view, Rawls’s employment of the original position rests on the erroneous assumption that the fundamental principles of justice are fact dependent and, moreover, it leads the content of the principles chosen there to be distorted by categories of non-moral fact that themselves have nothing to do with justice. He argues that both of these shortcomings reveal that the original position method generates a conception of justice that is mistaken because it is less egalitarian than it should be. I will refer to these as the critique of “fact-dependent foundations,” and the critique of “justice as regulation.” I will conclude, briefly, with a discussion of a third form of sensitivity to facts suggested but not developed by Cohen, namely the role that the original position method must give to morally bad facts.

I presume no prior familiarity with Cohen’s arguments, though some familiarity with Rawls is taken for granted. The stage can briefly be set with a reminder of what I will call Rawls’s original position method.² I will

¹ All page and chapter references to Cohen’s writing will be to this work. I am grateful to Amy Baehr, David Brink, Timothy Hinton, Tim Syme, and Paul Weithman for comments on a previous draft.

² There are some later developments, but the canonical presentation of the original position is in *TJ*.

sometimes follow Cohen in referring to it as Rawls's "constructivism." Rawls's seminal development of the argument from the original position stems from the following methodological proposal. Oftentimes, disagreements about what is to be done can be narrowed or settled if the answer is to be provided by a procedure whose authority is less controversial. We might not agree about which restaurant would be the best choice, and yet we might agree that this ought to be decided by voting after discussing it. We don't disagree about that procedure as intractably as we do about which choice would be correct (judged by some procedure-independent standard). Rawls proposes something similar for the case of disputes about the content of social justice: these disputes might be more tractable if we could devise a procedure (it is bound to be only hypothetical, but that would suffice) about which there is less (reasonable, at least) dispute, in the following respect: it might be seen as effectively designed so as to select principles that serve people's interests impartially, when they are conceived as morally free and equal. If such a procedure favors certain views of the content of justice over others, this ought to weigh in favor of those views as a form of evidence or support. It might give pause even to those who hold to their original views with more conviction than they can invest in the proposed procedure, perhaps leading them to appreciate the decency and seriousness with which their opponents' view can be supported. Surely, that would be something.

Famously, Rawls develops an elaborate procedure to serve this role.³ Individuals are represented in this hypothetical original position by parties who are able to know all that science and good sense can tell us about how the competing principles would be likely to bear on certain fundamental "primary social goods." These parties will each prefer the principles that would do best for them, though they are, in their special original position office, not to concern themselves with the fate of the clients of the other choosers. However, the choosers are behind a "veil of ignorance": they are not to know which of the people in the real principle-governed society is their client. They naturally must consider the possibility that it might be anyone.

Cohen's objections to this whole approach stem from the fact that the results are partly, but crucially, driven by the non-moral facts that the parties are aware of. As I have said, Cohen mounts three lines of attack, to which I now turn.

³ For a thorough introduction to the original position in Rawls see Freeman, "The Original Position," *Stanford Encyclopedia of Philosophy*.

7.2 Fact-free foundations

Cohen mounts a general and abstract argument about the nature of moral principles that begins from the claim that facts have whatever moral significance they have in virtue of principles. Thus, fact-independent principles are morally more basic than, because they explain, fact-dependent ones. Certainly, some moral principles depend on non-moral facts.⁴ For example,

Principle 1: It is wrong, except in special conditions, to physically *strike* another person,

and this is partly because of certain facts. That is,

Fact basis 1: Principle 1 is true partly because striking people tends to hurt or injure them.

The fact that hitting will tend to have this effect is not a moral truth, but a non-moral fact derivable (if necessary) from physics, physiology, and psychology. The moral wrongness rests on this non-moral fact in the following respect: hitting people would be less wrong, or perhaps even permissible if, and then because, it did not hurt people. Cohen accepts this, of course. But he argues that all such fact-dependent moral principles rest, in their turn, on principles that are independent of the non-moral facts in question. In the case of the wrongness of hitting, the following principle seems to be implicated, and yet it is not dependent on non-moral facts about how hitting affects people:

Principle 2: It is wrong, except in special conditions, to *hurt* people.

This is a deeper principle than (1) in the following sense: (2) grounds (1), and not vice versa. We will not try to say exactly what this grounding relation is, but we can note a few features of it that indicate the sense in which (2) grounds, or underlies (1). Notice that (1)'s truth depends on (2)'s truth. If this were denied, it would be inexplicable why the fact that hitting hurts is stated as a fact upon which (1) depends. How could this fact support (1) unless it was normally wrong to hurt people, which is principle (2)? Citing the fact that hitting hurts as a basis for the wrongness of hitting evidently presupposes, as an explanation, the wrongness of hurting. So (1) must be granted to rest on (2).

⁴ The arguments I attribute to Cohen in this section appear in his own words in Chapter 6, section 5.

But (2) does not similarly rest on (1), of course. The wrongness of hurting does not rest or depend on any facts about whether hitting happens to hurt people. If hitting did not hurt people, then we would have less interest in the morality of hitting, but we are not now asking what moral principles are of interest given the facts as they are. We are asking what things are right and wrong. So there is this asymmetrical relation of depth. We will mark these ostensibly deeper principles with higher numbers, as measures of increasing depth. So (2) indicates more depth relative to (1).

So, the wrongness of hitting rests on the wrongness of hurting, but the wrongness of hurting does not rest on the wrongness of hitting. Principle (2), the principle about the wrongness of hurting, obtains independently of the facts that ground (1). But (2) might be thought in turn to rest on some other non-moral facts. In that case we would not yet have to accept that there are any fact-independent moral principles. Cohen's argument, however, contains two importantly different claims, one of which we have just seen an argument for:

Relative claim (of fact-independence): For any principle that does depend on certain facts, there is a deeper grounding principle that does not depend on *those* facts.

This leaves open the possibility of an infinite stack of deeper principles, each of which depends on some facts, albeit owing to a deeper principle yet. But Cohen denies that this is so in his second claim, which is the

Ultimate claim (of fact-independence): Every principle that depends on facts is grounded, directly or indirectly, in a principle that depends on no facts at all.⁵

Unless the ultimate claim is true, principle (2), which grounds the wrongness of hitting in the wrongness of hurting, might itself depend on some non-moral fact. It is not, as we saw, grounded in the fact that hitting hurts, so it would have to be some other fact. What might it be? There might be something deeper to say *morally* about what is wrong with hurting people, but we are asking whether it is wrong to hurt people owing to some *non-moral* facts. I cannot think of any plausible candidates, but even if there were such grounding facts, Cohen challenges us to recapitulate these questions at the next deeper level and to see if we can find grounding facts yet again. Cohen

⁵ I paraphrase what Cohen says at p. 232 and elsewhere.

argues, from such examples, that this procedure will always eventually (and usually quite quickly, he thinks) terminate in a principle that does not rest on any non-moral facts at all.

Since it will figure importantly below, we want a handy name for Cohen's movement (in the manner of the "relative claim") from any fact-dependent principle to a deeper, relatively fact-independent principle that explains the moral relevance of those facts. I will call this "unearthing" the deeper principle.⁶ In asking what grounds the fact that hitting is wrong in virtue of hurting, we can unearth the deeper principle that hurting is wrong. Cohen hopes to use this strategy against the original position method.

7.3 A formal objection: Cohen's unearthing strategy

As we have seen, Rawls argues that principles of social justice are "constructed"⁷ partly through engagement with non-moral facts. The question that guides the Rawlsian approach is how things would work out in practice, in certain ways, if the basic social structure met certain principles rather than others. "In practice" is not meant to tie the principles to all the expected contingencies of our world going forward. It idealizes in particular ways. For example, parties to the original position are to choose principles of justice on the assumption, clearly contrary to fact, that there would be, among other "favorable conditions," publicly recognized full compliance with the rules and norms of the basic structure. Still, the parties who are selecting the principles in the hypothesized choice situation will bring to bear their general knowledge of such non-moral facts as characteristic human motives and concerns, cognitive abilities, characteristic patterns of moral development, predictable strains in keeping certain commitments, not to mention the whole universe of facts about how nature itself operates.⁸

The derivation of the fundamental principles of social justice in the original position renders those principles dependent on, and explained partly by, the non-moral facts about people and nature that lead the parties to select them. Rawls is explicit about this. He writes, "There is no necessity

⁶ The pun is intended, though no part of my present point. Cohen agrees with Plato that, "justice transcends the facts of the world. . . [and] that justice is the self-same thing across, and independently of, history. But that extreme anti-relativism is no part of the doctrine here defended that justice is, ultimately, facts-free" (pp. 291–2).

⁷ *PL*, p. 103.

⁸ Freeman pulls together the main textual sources for roughly this list in "The Original Position."

to invoke theological or metaphysical doctrines to support [these] principles . . . Conceptions of justice must be justified by the conditions of our life as we know it or not at all.”⁹

At this point, of course, we should bring to bear Cohen’s strategy of unearthing the fact-independent principles that are alleged to ground any fact-dependent ones. If Rawls is right that the two principles of justice as fairness are grounded wholly or partly in the non-moral facts about the “conditions of our life as we know it,” there must be, Cohen argues, a deeper principle that explains the particular moral relevance of those facts. This would suffice to show, Cohen thinks, that the principles generated by the original position cannot be the fundamental principles of social justice. The normative force of Rawls’s two principles would be explained by some more fundamental principle, one that does not depend on those facts. To get to the fundamental principle or principles, we must unearth them by scraping away the facts whose relevance is grounded by something deeper.

Suppose Cohen’s argument were conceded up to this point. It might seem to undermine the original position as an appropriate way to generate or arrive at the content of principles of social justice. And suppose a Rawlsian were to acquiesce in the demand for further unearthing. She then reports on her findings:

The unearthed Rawlsian principle: The fundamental principle is this – institutions ought to meet principles, whatever they are, that would be chosen in the original position, with its sensitivity to the facts whatever they might be.

This principle is now independent of any of the facts that the original position brings to bear, just as the wrongness of hurting is independent of the fact that hitting hurts. It is the kind of grounding principle, possibly independent of any facts at all, that Cohen has urged us to seek.

Obviously, a key question is whether or not the unearthed Rawlsian principle would necessitate any revision of the content of Rawls’s two principles of justice. On the face of it, it seems that their content would remain exactly the same. And indeed, Cohen acknowledges this.¹⁰

Moreover, the unearthed principle also grounds the content of justice by grounding the exact same moral relevance of the facts of our life that Rawls had

⁹ *TJR*, p. 398.

¹⁰ Pp. 262–3. My point here can allow that there would be even deeper principles required to “justify the use of the original position machine” underlying even the unearthed Rawlsian principle.

originally maintained. It grounds the whole original position methodology unchanged. It merely adds that in a certain sense, the famous two principles, which depend on facts, are not fundamental because they are explained by a deeper principle – the unearthed Rawlsian principle – which does not depend on the facts. But this is uninformative normatively, and it leaves in place the grounding relation between the facts and the principles Rawls had alleged.

Cohen still has a complaint here. Let's call it a *formal complaint*, because it is a complaint about how Rawls formulates the theory, not about its normative substance. (In that sense, we could also call it "metaethical.") Let's look more closely at the formal complaint before turning to a *substantive complaint* as well, but one quite independent of the unearthing strategy.

The unearthing move to the fact-independent principle is so simple that it does not appear to make any substantial difference in Rawls's theory at all.¹¹ But there is, first of all, an important metaethical lesson here, and Cohen is right to think it is philosophically significant. It is structurally similar to the point (as I see it) of the famous Euthyphro problem:¹² If "[i]t is wrong to murder" depends on the fact that God forbids murder, there is a deeper principle that does not depend on that fact about God, namely: "[i]t is wrong to disobey God's will, whatever it might require." On the divine command view this principle is more fundamental than "it is wrong to murder," because it morally explains the force of that prohibition. Similar points can be made for any view according to which moral principles stem from the outcome of a certain specified agency or procedure of any kind. Such views ascribe a certain moral authority to the agency or procedure. It looks to be a deeper (maybe fundamental, maybe not) normative truth that they have that authority to begin with. And that deeper normative force or authority does not stem from or depend on what those agencies or procedures do or say at all.¹³ The structure of Cohen's point, then, has an ancient resonance in a broadly Platonic cast of thought.¹⁴

¹¹ Several authors have explored the simplicity or triviality of the ungrounding move, for example Pogge, "Cohen to the Rescue."

¹² Plato's *Euthyphro*.

¹³ This is related to Valentini and Ronzoni's argument in "On the Meta-ethical Status," p. 403. They argue that the unearthed Rawlsian principle might be grounded in a fact, but a methodological one about how to construct justifications. Cohen would surely ask what it is that accounts for the particular moral significance of that method of construction, if not a deeper principle?

¹⁴ Cohen aligns himself with Plato in certain respects at p. 291. For Cohen's views about Plato's views more generally, including his "reactionary" political position relative to the sophists, see his *Lectures*, Chapter 1.

Cohen believes that, in this way, he has identified a deep metaethical disagreement between himself and Rawls.¹⁵ He believes that his own view, in which the unearthed principle is the fundamental one, is in conflict with Rawls's view in which the original position-generated principles are the fundamental principles of justice. I take Cohen to be arguing that Rawls is committed to denying that moral principles are ultimately grounded in fundamental fact-independent principles. This appears to be a metaethical disagreement in the sense that it is independent of disputes about what the content of moral requirements might be. So Cohen takes Rawls to be committed to a faulty (because fact-bound) metaethics.

There is a way of interpreting Rawls that would avoid any metaethical tussle with Cohen, one hinted at in the quotation just above, in which Rawls says, "There is no necessity to invoke theological or metaphysical doctrines to support [these] principles." This is importantly not the same as asserting the metaphysical view that *there is no deeper grounding principle* for the relevance ascribed to the facts in the original position method. Rawls says, instead, more modestly, that the metaphysics (and here, metaethics) is beside the point if we are interested in the content and justification of the principles of social justice. Whether because God says so, or because there is a quasi-Platonic fact-free principle to this effect, Rawls may still assert that the principles of justice are justified by their appeal to the parties in the original position in light of the facts of human life. Nothing *about what would be just* is added by pointing to the unearthed fact-independent principle (whether God-given, Platonic, or otherwise) that says, simply: "Those are the principles of justice, justified by the original position argument which appeals to those very facts." The quoted passage from Rawls, "Conceptions of justice must be justified by the conditions of our life as we know it or not at all," is no support for Cohen's evident suggestion that Rawls is metaethically committed against this kind of "rational intuitionism." Indeed, that very quotation could be read as an endorsement by Rawls of a fact-independent principle, one that can be unearthed, in Cohen's own fashion, in looking for further grounding for the method of the original position. That is, to say that principles of justice are to be justified by the facts of human life or not at all, is not to deny that there may be some deeper philosophical support for this very view.¹⁶ Contrary to

¹⁵ Chapter 6, especially section 18.

¹⁶ As Cohen points out, Rawls "disparages" rational intuitionism (p. 258), but my suggestion is that his target is only rational intuitionism about the content of justice. "[O]nly the substantive principles specifying content of political right and justice are constructed. The procedure

what Cohen argues, the original position method, in which principles are justified by facts, is not committed to denying that there is some fact-free basis for any such moral relevance of the facts in a deeper normative principle.

7.4 A substantive objection: regulation and non-justice values

Cohen denies that the facts of human life rightfully play any role in determining the content of justice, and we will consider his arguments shortly, but this is neither implied by, nor does it imply, the unearthing point about fact-free fundamental principles. We should have names for two separate issues about the relation of facts to principles. Cohen's unearthing strategy uncovers principles that do not *rest on facts*. This leaves entirely open whether those principles *operate on facts*. To say that principles operate on facts is to say that their normative implications vary in accordance with relevant variations in the facts. Consider, again, the unearthed Rawlsian principle. It says that the content of justice is given by principles chosen in the original position in light of the non-moral facts of human life and nature, whatever they might be. As we have seen, this principle does not rest on any facts, such as those of human nature. Rather, the principle grounds the moral significance of those facts. But this shows that the principle operates on certain non-moral facts. It gives them a certain moral significance. The principle's normative implications vary depending on facts about "the conditions of our life."

Every normative principle operates on facts, by saying which facts have what kind of moral significance. If this is right, then the way to understand Cohen's substantive complaint about the original position method, as represented by the unearthed Rawlsian principle, must be that it operates on the wrong facts and/or that it operates on facts in the wrong way. Cohen's substantive objection is that the original position method can be shown to mix questions of justice with other considerations, as if one were choosing rules of social regulation rather than (the real project at hand) trying to ascertain the true or genuine principles of justice. We might keep the issue clearer if we distinguish the formal *objection from facts*, which we discussed in the previous section, from this substantive *objection from regulation*. In a way, it is a coincidence that this second issue also happens to be about facts

itself is simply laid out" (*PL*, p. 104). On the question of whether there is a deeper principle that explains how facts have the moral importance they are given in the original position method, I believe Rawls can and does stay neutral in affirming a "constructivist conception" that is "political and not metaphysical" (*PL*, p. 97).

allegedly playing too great a role. I believe we get a clearer view of Cohen's substantive worry if we refrain from continually casting it, as Cohen tends to do, as (to put it roughly) anti-fact. The crux of this point is that the original position method wrongly assimilates principles of justice to rules of social regulation.

The explicit distinction between rules of social regulation and principles of social justice is, as far as I know, original to Cohen and it is powerful. A precise account of rules of regulation is not needed in order to see the distinction Cohen is after.¹⁷ It rests on the powerful observation that when we make choices, there are often reasons in place that count for or against certain alternatives. Rules of social regulation are things we choose or adopt, and we do so for certain reasons. Among the reasons to consider are reasons of justice. Once adopted, rules of regulation bear on how certain things are to be done, sometimes in the form of laws, sometimes in less formal norms. Being normative in that way, they are easily confused with principles of justice. But we do not choose principles of justice. Rather, they are among the considerations we consult in our choice of rules of regulation.

This important distinction allows that our convictions about social justice might include principles that we do not, on balance, have reason to adopt as rules of social regulation under the conditions we happen to face. Such regulatory reasoning might recommend adopting (in practice) rules with a different content from the principles of justice we have compelling reason to accept as a matter of conviction, and we will consider examples in the next section. Cohen argues that this dissonance can arise because among the great variety of considerations that will bear on the question of which rules to adopt, it will be relevant how one set of rules of regulation would work out in various respects – engaging values other than justice, such as efficiency or stability – as compared to an alternative set of rules.

Here is one interpretation of Cohen's line of thought:

- 1 The original position, with its regulatory reasoning, requires that the choice, by the parties, of principles be made in light of whatever might affect people's interests – effects stemming from the adoption of one or another set of rules.
- 2 Not just anything about how adopting certain rules affects individual interests is a consideration of justice.

¹⁷ For more on the distinction in Cohen, see pp. 276–7.

3 Therefore, the original position's regulatory approach lets the choice of principles be determined by non-justice considerations, and so does not reliably identify justice.

(To be clear, (3) does not say that the original position approach *guarantees* that the choice of principles will actually turn on non-justice considerations; it says that such considerations are included among the ones that will, taken together, determine the choice.) On one meaning of "a consideration of justice" (in premise (2)) this would be a puzzling complaint. In the original position method it would be fatally circular to have the choosers bring ideas of justice to their deliberations. The aim of the original position method is to understand justice in terms of other ideas. A less puzzling interpretation of the complaint, and the one I believe Cohen intends, does not suggest that the parties should employ the idea of justice. Rather, the objection is that some of the considerations relevant to the parties' regulatory reasoning arise from values that are (so we philosophers determine) no part of justice.

Of course, the Rawlsian should flatly deny this. So we want to know what argument Cohen gives in support of this charge, and what can be said against it. We know that Cohen endorses an interpretation of distributive equality that differs from and conflicts with Rawls's with respect to a number of values and considerations that might move the parties, but that is beside the point Cohen is focusing on. He denies that the stylized regulatory reasoning is suited to track or constitute justice in any case, whether or not it happened to select principles he (Cohen) endorses.

The polemical situation is similar to the one raised by Nozick in his early charge that the original position method is designed from the beginning in a way that is blind to the possibility that justice is simply whatever would result from free market exchanges under protection of a minimal state. As Cohen writes, "the original position also excludes a concern for how much one person gets compared with somebody else: what I get by comparison with others finds no representation within that position, and believers in the claim to justice of relational equality should therefore be as wary of the original position, as a criterion of justice, as Nozick is."¹⁸

There are two subtly different lines of objection to consider here. One is that the original position argument is, roughly, question begging by, in effect, premising its argument, whose conclusion rules out Nozickean entitlement theory and Cohen-style egalitarianism, on the assumption that they are false.

¹⁸ Pp. 159–60.

It is not clear, though, which side (putting Nozick and Cohen on one side, with Rawls on the other) is best seen as begging the question. All sides are supposing for the sake of the present argument that Rawlsian premises do indeed entail the Rawlsian conclusion, thus ruling out the others. And, as Nozick notices,¹⁹ any good deductive argument will be logically valid, in which case no one who rejects the conclusion can consistently accept the premises. To show that the argument is question begging, then, must take more than showing that the premises logically entail the conclusions, but neither Nozick nor Cohen rises to this challenge. And, in fact, they both suggest that their doubt about the premises is based on and justified by their own theories of justice. It is patently question begging, as an argument against the Rawlsian original position premises, to object on the grounds that those premises do not entail Cohen's or Nozick's preferred principles of justice. What is needed by these critics is an argument against the original position which does not presume the truth of any particular account of justice.

A second line of objection by Cohen can be seen as taking this more promising form, arguing that the original position involves the parties in (hypothetically situated) regulatory reasoning, and that such reasoning is bound to introduce considerations that have nothing to do with justice as intuitively understood (and not assumed to be Cohen's egalitarianism or anything like it). So the question becomes this: which of the considerations that must influence the parties to an original position can be persuasively shown by Cohen to be considerations that do not, intuitively, bear on social justice?

After all, here is the Rawlsian position, put into Cohen's terms, which he must argue against: principles of justice are properly identified with possible rules of social regulation that parties to an original position would choose behind a veil of ignorance and under certain idealizing assumptions such as full compliance.

Notice that even what Cohen casts as regulatory reasoning is itself a morally defined enterprise. The question of what rules of regulation we should have in our society is, surely, a moral question. For example, in answering this regulatory question presumably no one's interests should count more heavily than anyone else's, and this is for moral reasons. And, arguably, the sum of interest-satisfaction is not a morally significant quantity, since it does not represent the good of any agent at all, again a moral point. So

¹⁹ *ASU*, p. 203.

the original position might seem to emerge as a good method for answering this moral question: which rules of regulation should we have for our society? Indeed, Cohen writes,

The present charge is not a criticism of the particular device, that is, the original position, that Rawls employs to answer the question, namely, what rules should we choose, that the denizens of the original position answer . . . Instead, I protest against the identification of the answer to *that* question with the answer to the question "What is justice?"²⁰

But importantly, Rawls is not proposing the original position or the two principles as answers to the latter question, for (at least) the reason that the parties are choosing under the false assumption of full compliance. Cohen tends to exaggerate the regulatory character of the original position method in this way, and it is a distraction from his main point. For example, he often writes of "identifying justice with optimal rules of regulation."²¹ Rawls is certainly not asking what rules of regulation we should have in the real world. He is enquiring into what justice would be for this world. The parties, while they are sometimes said to know all "general" facts – those that would not allow any of them to know which individual they will be – are actually fed at least one important factual falsehood, namely that there will be full compliance with the chosen principles. For that reason, the resulting principles, if translated into social rules, would not be sure to serve their intended purpose in a world in which compliance was only partial. Thus the original position is neither intended nor suited to selecting good principles of regulation for any plausible society. It is not perspicuous to say, then, that the account identifies optimal rules of regulation for a society with the principles of justice for that society. The original position identifies justice with appropriate rules for a certain hypothetical scenario. Indeed, Rawls could perfectly well agree with Cohen that principles of justice are not the sort of things we get to choose.

Cohen's central point is not damaged by this clarification, however. Even with the idealization about full compliance, the parties are engaged in what we might call "regulatory reasoning," and this is where the issue lies. The hypothetical society they seek to regulate is an unrealistically compliant one, but nevertheless, by selecting principles for that society on the basis of what would best promote their interests the parties are (so Cohen argues) necessarily concerning themselves not only with considerations of justice, but

²⁰ Pp. 277–8. ²¹ p. 277.

also with non-justice considerations such as efficiency and stability. That is the more important charge, rather than the charge that Rawls thinks that principles of justice are subject to real social choices, or that the choosers in the original position are selecting appropriate rules of regulation for real societies.

We should accept Cohen's argument here up to a point. It is hard to deny that such values as efficiency and stability will influence the parties' choices. What remains disputable, then, is whether Cohen has strong arguments showing that these values are not considerations that bear on justice. Cohen appears to believe that it is simply obvious. He writes, "I have . . . asked you to agree . . . with the . . . overwhelmingly intuitive claim, that the sorts of facts about practicality and feasibility that control the content of sound rules of regulation do not affect the content of justice itself."²² He argues that, "Constructivism about justice lacks the conceptual resources to describe justifiable trade-offs between justice and other desiderata, because those desiderata (improperly) constrain what constructivism deems to be just."²³ but what we are looking for is an argument that justice is the sort of thing that should be balanced against practicality and efficiency rather than *being the appropriate way of balancing them*, as constructivism, with its regulatory reasoning, would hold. Cohen's argument for this lies mainly in his discussion of a number of problems of social policy or regulation, including the structure of tax rate schedules, the problem of differential care (roughly, moral hazard), and issues around publicity and stability.²⁴ For reasons of space, I consider only the first two here, arguing that they do not succeed.

7.5 Tax brackets and exactness

Cohen argues that step-wise tax brackets are bound to be less than perfectly just, and yet they must be irresistible to a constructivist theory of justice for reasons of administrability. This puts daylight between the original position method and considerations of justice. How could the person whose property or income is greater than another person's by the single dollar that kicks him into the higher tax (or rate) owe, *as a matter of justice*, much more tax – not just a little more – than that other person, and yet owe exactly the same as

²² Pp. 278–9. ²³ P. 312.

²⁴ Richard Arneson, by contrast to my reading, finds Cohen's "terminological" choice here to be undefended. See "Justice Is Not Equality."

significantly richer rate-mates? Of course, it might be impractical to spend vast sums of public money to implement a highly refined tax schedule, and in that case it ought not to be done. That does not, however, make the step-wise tax rates as just as the more refined ones would be.

Cohen argues that the parties to the original position would not insist on an extremely fine-grained tax schedule if it would be vastly more expensive than a moderately fine-grained schedule. This is because they will be sensitive to how those extra resources might be used to benefit them in other ways. He takes this to show that a constructivist method must be prepared to trade greater justice off against gains with respect to other values such as efficiency.

Does this objection really show that Rawls's original position method forces the parties to engage in regulatory reasoning? It is fair to ask to what extent the original position addresses such things as tax rates at all. The answer is that it mostly leaves tax rules aside as a subsidiary question about how the basic social structure could be brought to meet the difference principle (subject to the other Rawlsian principles). So Rawls's view seems to be that tax rates are not directly matters upon which justice takes a position. A given tax rate is just in the purely procedural sense if it is the product of a just basic structure where legislators duly aim to maintain the structure's justice by conforming it to the basic principles.

Cohen might reply that Rawls's original position argument forces him to deny that there is anything unjust about, say, taxing the poor at a higher rate than the rich, if that should turn out to be the most sensible policy all things considered. But, of course, it doesn't quite say that. It says that there would be nothing unjust about doing that *so long as* that scheme is part of a basic social structure that meets the difference principle and the other principles. Rawls could argue that no such tax system is remotely likely to meet that proviso, thus explaining the absurdity in the suggestion that it might be just. But some will agree with Cohen that this does not accommodate the deep intuition that some tax rates are unfair, even if there are other good reasons for adopting them, irrespective of their downstream effects on distribution.

7.6 Differential care

The case of what Cohen calls "differential care," similar to what is often called "moral hazard," is a second example he uses to support his view that Rawls's constructivist method in the original position incorporates values that have nothing to do with justice. He sketches an example, which I slightly simplify here. Suppose that there are two possible schemes S1 and S2 for publicly

compensating homeowners should storms damage their property. Under S1 everyone gets fully compensated for any damage. However, some people might rely on this program and reduce the amount of care they take to prevent storm damage. When compensation is provided, this would seem to be unfair to homeowners who had, at their own expense, taken greater care and minimized their property damage. To reduce that kind of unfairness, we might prefer scheme S2, which requires anyone who claims benefits to bear the first \$200 of repairs themselves. This provides people with an incentive not to skimp in their preparations in the hopes of being bailed out later, thus reducing the unfairness produced by compensation under S1. Of course, the “deductible” in S2 is crude in that it is not scaled to each homeowner’s incentive. Therefore, there might remain homeowners who will still do little or nothing even though the first \$200 will be their own responsibility, calculating that their preventive costs would be considerably more than their expected compensation (the amount in excess of \$200 multiplied by the probability of its occurring, say). Finally, suppose that while we could, at great expense, determine just how conscientiously each homeowner prepared her house for storms, and thereby tailor compensation so as to avoid such free riding,²⁵ this scheme, S3, would be very expensive. To recap: S1 compensates for damages with no deductible, S2 compensates but with a \$200 deductible to discourage skimping, and S3 expensively compensates each, partly according to how thoroughly she prepared. Cohen argues that, (a) if there are free riders, S3 would be the most just, but (b) the original position approach to justice, with its reliance on regulatory reasoning, would, if S3 is expensive enough, select S2. The Rawlsian method, then, trades off justice against non-justice values such as efficiency.

S3’s allegedly greater justice seems to presuppose that unequal outcomes are only just if they reflect factors for which individuals are responsible.²⁶ It might look as though Cohen’s argument here assumes that “luck egalitarian” standard of justice, indicting the original position method on the grounds that it doesn’t select that principle. That would be question begging in a dispute about whether justice truly has that content. But there is another way of reading Cohen’s complaint against constructivism, namely as asserting only that to count against a policy the fact that it would be very expensive is to give weight to a consideration – total expense – that is not an ingredient of justice at all, but a different value altogether. The original position method’s reliance

²⁵ See Cohen’s more elaborate example at pp. 308ff.

²⁶ P. 313.

on regulatory reasoning evidently selects for less expensive options other things equal, and it thereby disqualifies itself as a reliable method for determining the content of justice. To avoid the question-begging form of argument, we must be willing to grant – whether or not we accept luck egalitarianism – that the overall expense of an arrangement does not count against its justice.

It might be asked how we supposedly know that a scheme's overall expense is not any part of its justice? Grant that parties to the original position would, other things equal, disprefer an arrangement that is more expensive. Can we infer that the original position set-up forces Rawls to be indifferent to comparative benefits? Supposing that we were to grant to Cohen that social justice, whatever more precisely it is, cannot be completely indifferent to comparative benefits, that would be damning. Surely we cannot infer it, though, since the parties in the original position might give some weight to inequality and also some weight to aggregate benefit – overall cost. Cohen tells us that he rejects the view, “that distributive justice doesn't have a comparative aspect at all. And as long as it (at least also) does so, then [constructivism's sensitivity to the non-comparative issue of efficiency] will be (at least) in one way a deviation from justice.”²⁷ That is a dubious argument, which resembles the following fallacious reasoning: a crust is an aspect of what it is to be a pie. So pumpkin pie, by including pumpkin, which contains no crust, deviates from pie-ness “(at least) in one way.” The fact that a certain consideration (such as Pareto efficiency or overall cost) is not a comparative consideration does not establish that an account in which it is one consideration among others is not comparative.

It is worth noting that Cohen eschews, without explanation, the name “moral hazard,” for the issue he calls “differential care,” even though in earlier drafts of the book he did use the more familiar name.²⁸ One possible explanation leads us into an important issue. The term “moral hazard” suggests that when certain protective policies induce reduced care or increased risk-taking this operates by way of some moral deficiency on the part of those who reduce their care or take greater risks. Cohen points out, however, that while some reduced care might be intended to exploit the care of others, this needn't be the case. Some people might reduce their care in response to the protective policy without any intention of exploiting others, and so there needn't be any element of moral defect, but simply an economic calculation. An alternative

²⁷ p. 323.

²⁸ See September 2003 and September 2006 drafts, on file with the author.

explanation is that strictly speaking, the familiar problem of moral hazard is not about “differential care” – some exercising less care than others – at all, but about some or all (maybe even equally) reducing their care in response to the policy, still with the clear implication that their doing so is a “moral” wrong. In any case, we can see that Cohen wishes to focus on the case of differentially reduced care, and without the assumption that it is morally wrong. His point, as we have seen, is that even if no one is morally misbehaving, a policy that leads to such differentially reduced care will be unfair to those who reduce their care less or not at all, even though it might be the appropriate policy in light of the costs of fairer policies.

Summarizing, the differential care example of Cohen’s might show that Rawlsian constructivism will tend to trade off equality of a luck egalitarian kind for the sake of efficiency (less overall cost). But that doesn’t show that this deviates from justice unless it is question-beggingly assumed that justice is luck egalitarian in content. The simpler point that the original position method will be sensitive to overall expense, or to Pareto efficiency (which is not a comparative consideration), shows neither that justice is essentially comparative in nature, nor that, even if it is, constructivism is indifferent to (other) comparative considerations.²⁹

7.7 Bad facts

We have seen two ways in which Cohen criticizes the original position method for bowing to facts. The first was his formal objection that Rawls fails to recognize that facts are only morally relevant owing to principles that determine their relevance. Call this his objection to Rawls’s reliance, in a certain way, on *facts as such*. The second and independent complaint was that the original position bends its results to facts that have nothing to do with social justice, facts concerning such things as efficiency or stability. Call this his complaint about constructivism’s reliance on *facts irrelevant to justice*.

In his treatment of the case of differential care, we glimpse a third kind of complaint about reliance on facts, but one that Cohen does not clearly demarcate as different from the other two. Repeatedly, as I will illustrate, Cohen rhetorically leverages the moral deficiency represented by some of the

²⁹ Space prevents consideration of Cohen’s treatment of stability and publicity. His argument that the original position method deviates from justice succeeds overall if it succeeds in any of these cases.

fact to which constructivism bends its results. Call this a complaint about constructivism's reliance on *bad facts*.³⁰

In discussing the differential care case discussed above, he observes that

The root cause. . . that induces a compromise with justice in the "exploiter" variant of the differential care phenomenon is a certain human moral infirmity: Constructivists are, therefore, in the questionable position that they must defer to facts of human moral infirmity in the determination of what fundamental (nonrectificatory) justice is.³¹

Cohen acknowledges that under full justice-compliance, which Rawls assumes to be in place, there might not be any exploitation, but only innocent differential care. Even if there is no exploitation, there might be reasonable fear of exploitation, which he says is enough for his purposes. And yet he continues to suppose that "moral infirmity" might well be part of the differential care profile: "It would be transparently wrong to say that the facts about moral weaknesses and so on make S2 just (without qualification), as opposed to more worthy of selection."³²

It might have seemed from the very beginning that Cohen objected to the original position's sensitivity to, in particular, bad facts. He complained that if talented citizens decide to withhold socially productive labor unless they are paid more than others, the original position will lead to the conclusion that justice requires paying them more even though such a demand seems to put them outside the publicly shared sense of justice. To that extent, they look bad. But Cohen never quite embraced this route. Even in the earliest setting, he did not complain that justice is being bent to accommodate bad behavior, but complained, more obliquely, only that in that case the talented would not count as in "justificatory community" with the other citizens – quite a different point.³³ And true to form, in the much later treatment of differential care in *Rescuing*, he refers only glancingly to the inadequacy of "moral weaknesses *and so on*" (emphasis added) to qualify the original position's results as just. He does not make clear whether the original position is disqualified partly by the fact that its results are bent to make the best out

³⁰ I briefly discuss a part of this question in "Human Nature," pp. 225ff. ³¹ P. 309.

³² P. 311.

³³ "Incentives, Inequality, and Community," pp. 263–329. There, he says that the original position "generates an argument for inequality that requires a model of society in breach of an elementary condition of community" (p. 268).

of *morally* bad situations – adjusting not only to facts as such, or to justice-irrelevant facts, but to bad facts.³⁴

Even if Cohen never offered that line of criticism in full voice, it poses an important challenge to the original position method. If a certain predictable moral deficiency in some or all people – maybe indefensible selfishness or partiality, or maybe a certain ineluctable level of bigotry – leads the original position to reject certain arrangements as infeasible or unreasonably expensive, it may strain credulity to accept that justice is being reliably tracked. That is, suppose one disagreed with Cohen's view that feasibility or expense are values foreign to justice, holding that they are among the values that ought properly to be balanced in the constitution of justice, as the original position method holds. Nevertheless, when an arrangement is rejected because it would be too expensive, but where the expense stems from moral deficiencies people would have under the hypothesized conditions, there is, arguably, an untoward capitulation to vice that seems foreign to the idea of full social justice. (Cohen seems to me unclearly to be running both objections at once in the passages I have quoted above concerning differential care.) That would be a third of three lines of objection to letting justice be sensitive to facts: to facts as such, to justice-irrelevant facts, and to morally bad facts.

One possible reply is to say that even if the original position procedure is, in principle, sensitive to such bad facts, as it happens, it turns out they make no difference to the choice of principles of social justice. It seems likely that Cohen would reject this as a defense, being essentially the kind of defense of utilitarianism that he explicitly rejects.³⁵ Whether a reply along those lines is successful is a further question. A second possible reply would be to say that an account of justice would not be tainted by its sensitivity to justice-*irrelevant* bad facts, and that justice-*relevant* bad facts are put aside by the Rawlsian assumption of full justice-compliance. Why think that moral defects that are not themselves any violation of the principles of justice are justice tainting in the way the bad facts objection alleges? There is more worth thinking about in this third, barely broached, line of objection to the original position method's sensitivity to facts, but I must leave the matter here.

³⁴ At pp. 178–80, Cohen trenchantly scrutinizes several short texts from across Rawls's career on this question, but still does not suggest that his (Cohen's) critique of Rawls's argument relies on the badness of the facts in question.

³⁵ Pp. 263–7.