



## Human Nature and the Limits (If Any) of Political Philosophy

*FINAL DRAFT for Philosophy & Public Affairs*

8/23/2011

David Estlund

Brown University

It is often supposed that a person is not required to do anything they cannot do. “Ought” implies “can,” as this is often put. Let’s accept that for the sake of argument. Still, it is not obvious that if I can’t muster the will to do something, then I can’t *do* it. An inability to will an action might not entail an inability to do the action, in which case the action might yet be required. I will discuss this question shortly. In any case, here is a related but different idea: even if it is not impossible to *do* whatever it is impossible to *muster the will* to do, perhaps it is still implausible to think one could be required to do what one could not muster the will to do. Perhaps we should accept that “ought” implies “can will.” If so, and if there are characteristic things that humans cannot muster the will to do, then human nature (in the sense of the limits of human motivational capacities) would stand as a prior set of facts that constrain what political philosophy can soundly prescribe or morally require.

My thesis, simply stated, is that this is not so. If there are facts of human nature of this general kind, consisting in limits to what humans will be able to muster the will to do, they are not, simply as facts, constraints on what can soundly be prescribed or morally required. The reason is that agents’ abilities and inability to muster their will are subject to moral evaluation in their own right. Some such inability are morally objectionable, others are not. The fact that some men cannot stand to be

at a business meeting with a woman is a fact that must be taken as a constraint in certain practical contexts. It does not, however, force the retraction of a moral requirement to participate in meetings without gender discrimination. Nor would the fact, if it were one (which I doubt), that male sentiments of this kind are part of human nature refute a theory according to which the male-imposed exclusion of women from the halls of economic power is unjust. I want to argue that human nature—more specifically, whatever motivational incapacities are possessed by humans as such—is a constraint on some tasks in political philosophy but not on others. It is not a general constraint on political philosophy.

### *I. The Human Nature Constraint*

Here is the common position that I will argue against:

*The human nature constraint:* A normative political theory is defective and thus false if it imposes standards or requirements that ignore human nature—that is, requirements that will not, owing to human nature and the motivational incapacities it entails, ever be satisfied.

Equally familiar and traditional is a certain way of responding to such charges. An example—one that I will use throughout my argument—is the charge the Marxist and other socialist and egalitarian political theories have long been burdened with: that humans, in their nature, are not and never will be like that. The response to this charge has often been to accept the human nature constraint on political theorizing but deny that human nature is as the critics charge. For example, it is sometimes said that humans as we have observed them up to this point in history might be more partial and selfish than socialist theory would require, but that such observations would not show that our nature

is such that this will continue.<sup>1</sup> I will not be engaging in this kind of dispute about what is or is not in human nature. Rather, I will argue against the constraint itself.

It is important to acknowledge that the term “human nature” can be used to refer to other things about humans, such as their biological needs or their tendency to age. I am using the term in the narrower way I have indicated. However, disputes about whether political philosophy properly respects human nature are not normally about issues like our need for oxygen or our tendency to age. What political philosopher puts these aside? Traditionally, the disputes are very often about allegedly characteristic human structures of motivation, and especially about partiality and selfishness. I am engaging this traditional and central line of dispute about the relevance of human nature to political philosophy.

To keep things simple, I will grant for the sake of argument that it is in the nature of humans to be more selfish and partial than socialist or egalitarian (or whatever other) theory would need them to be. I deny, however, that this refutes any such normative political theories. Far from taking a stand on the content of human nature, it would be open to me to deny that there is such a thing as human nature at all. But I will grant it for the sake of argument. Some have suggested that philosophical criticism is often in one or the other of two categories: “Oh yeah?” or “So what?”<sup>2</sup> When a political theory is alleged to violate the bounds of human nature, many have responded with “Oh yeah?” My response, by contrast, is “So what?”

The appeal to human nature suggests something more than merely making claims about how people *will* act or are likely to act. The suggestion seems to be that there are, in addition, forces at work. But let’s consider the weaker claim first, that people (simply) *won’t* act as the theory requires.<sup>3</sup> In addition to the example of socialist or egalitarian theory, it might be held against a theory requiring significant civic virtue that, for example, people will never do that. It’s hard to see how this is an objection. It would not surprise any of us if people will never be very good or if societies will never be very just. But that very thought requires reference to a standard not met. To show

(supposing that one can show it) that a standard will not be met might count against people's behavior rather than against the standard. Of course, it may be that the thought embodied in such a complaint is that if people won't do it, or if we can know in advance that they won't, then maybe this is because they *can't*. Then, if "ought" implies "can," that would refute the theory. If people can't do it, then they are not required, and the theory that says they are required is thereby refuted.

"Ought" implies "can" is a formula that stands for a number of separable ideas. I want to grant, for the sake of argument, that some version of it could refute a theory of justice. So let's say that if a theory says that a society would be unjust unless it met certain conditions, then that theory of justice would be false if the society lacked the ability to meet those conditions.<sup>4</sup> This is debatable, but even if we grant it, there is no basis for supposing that what people won't do is something they cannot do. People and societies fail to do things they should and could do (even do easily) all the time. If it could be shown that a society lacks the ability to meet a theory's requirements, then I will grant for the sake of argument that the theory would be defeated. But knowing that society will never do certain things is a long way from having any basis for thinking that it lacks the ability.

Here is an important point to keep in mind, if only in the background. For present purposes, when we ask ourselves whether a person can do something, we should not be assuming that causal determinism prevents it. If every event, including every action, is necessitated by causal antecedents, showing that nothing is ever required other than what we actually do, then the issue about justice and human nature isn't of any further interest. This would not be the view that requirements of justice must stay within the bounds of human nature, but the view that they must stay within the bounds of what people and societies actually do. I will be assuming that either determinism is not true or (what seems more likely) that people can often do other than what they actually do even if determinism is true. For example, it might be that a person can do something only so long as they would tend to succeed if they were to try and not give up. That counterfactual situation might obtain even if it is not possible that they try. For example, it is causally guaranteed that Joe, being so uncaring, will never

try to help his neighbor. Still, if he were to try, he would tend to succeed, so it is not as if he can't help her. This sort of view is "compatibilist," in that it claims, compatibly with causal determinism, that people can often do other than what they actually do. (And if they can, then they might be required to.)

A second important background point concerns oppressive threats and incentives. What people can bring themselves to do is very often relative to an environment of threats and incentives. In many cases, what look like motivational incapacities—whether they are morally unfortunate or not—can be overcome by structuring those threats and incentives. In a subset of cases that require inducement, only oppressive threats and incentives would be effective. (I do not assume that coercion, which law normally involves, is always oppressive and wrong.) Even where the initial motivations are themselves shortfalls from what is necessary for justice (such as weakness of will or willing vice), obviously not just any inducement, however oppressive, would be morally permitted. Many inducements would often only substitute one injustice for another. Nor, however, does the fact that the oppressive inducement should be forgone somehow erase the injustice of the behavior that will result. Just because we shouldn't amputate the hands of thieves to eradicate theft (supposing it would) doesn't mean that the theft itself is permissible or that a society rife with it is just. Similarly, just because we shouldn't incarcerate and "reeducate" people who perpetuate bigotry (supposing that would work) doesn't mean that resulting high levels of bigotry are compatible with a just society.

So, returning to the main thread, suppose that "ought" implies "can." Plainly, "ought" does not imply "will" (unless people can only do what they will do, a possibility we are putting aside). So, putting the same points the other way around, "can't do" is *requirement-blocking*, but "won't do" is not.

## *II. What Humans Can't Will*

I hear appeals to human nature as something more than predictions, as suggesting that, in addition, there are forces at work. Often, the more specific suggestion, and the one I will concentrate on, is that the reason people won't behave in some way is that they are not motivationally made up that way. That is, because of their motivational natures they are unable to get themselves to do so. Consider Plato's theory of justice, part of which requires parents to permanently surrender their infants to community care and upbringing.<sup>5</sup> This might be requiring something contrary to human nature in the following respect: many parents, however persuaded they might be by the conception of justice, might find themselves unable to bring themselves to do it. Or consider a theory that requires people to devote their time and skills impartially to the common good rather than to personal benefit. Even people who accept this conception of justice would, perhaps, still find themselves unable to bring themselves to comply. This motivational incapacity might be in the very nature of human beings.

Call these examples cases of an inability to will the action. We have spoken of "won't do" cases, which are not requirement-blocking, and "can't do" cases, which we will grant are requirement-blocking. What about "can't will" cases? Are they requirement-blocking? Even if they are not requirement-blocking in individual cases, what about cases in which humans, by their very nature, can't will? Is this requirement-blocking?

In "can't will" cases, there might not be any obstacle to the agent's complying except the structure of the agent's own motivation. On one hand, this might be a reason for supposing that the motivational incapacity does not change the fact that the agent can do the action. She would succeed at doing it if only she thoroughly willed to do it or tried and didn't give up, and that might be a perfectly good interpretation of "can do," even if she "can't will." On the other hand, if she is incapable of mustering the will all the way through to action, then she might not succeed at doing

the action she wants to do. So, that is some basis for supposing that if she can't will it, then she can't do it. Which is it? Does "can't will" imply "can't do" or not?

It will be helpful to have a working account of what it means to say that a person can do, or is able to do something—I will use those interchangeably. I propose, as previous suggested, the following:

A person is able to (can) do something if and only if, were she to try and not give up, she would tend to succeed.

The mere fact that it is possible that I do something doesn't establish that I am able. It is possible that I draw a jack of hearts from a shuffled deck in a single draw, but I don't have the ability to do so. Actuality proves possibility, but not ability. Even if I try without giving up, I will not tend to succeed. The same goes for staying awake for four days straight, and much else that is not beyond possibility. A person might sometimes be able to do something even if they are unable to bring themselves to do it—even if they have that motivational incapacity. They will tend to give up, but if they had not given up, they would have tended to succeed.

Notice that we took for granted that if "ought" implies "can," and "can't will" implies "can't do," then "ought" would imply "can will." That is, if a person is unable to bring herself to do something even if there are no other obstacles, then she is not required to do that thing. "Can't will" would be requirement-blocking. And if humans have, by their natures, certain motivational incapacities such as those associated with the parental bond or with ineluctable self-concern or partiality, then normative political theory that required behavior incompatible with those facts about human nature would be refuted. All this hangs on whether we agree that "can't will" implies "can't do" (assuming "ought" implies "can").

Since I propose to argue that “can’t will,” even when it is part of human nature, is *not* requirement-blocking, I must hold either that “can’t will” does not imply “can’t do” or, if it does, deny that “ought” implies “can.” Indeed, since I’m granting that “ought” implies “can,” I must hold that the inability to bring oneself to do something (to will to do it) might coexist with an ability to do that thing. I must hold this because otherwise motivational incapacity would be requirement-blocking after all. But denying that is my central thesis. I have not argued for that thesis yet, but it is important at this stage to ask the reader to leave open the question whether “can’t will” implies “can’t do.” I grant that we sometimes say things such as “I can’t” in cases where, apart from our own motivational structures, there is no obstacle at all to our doing the action. We might say this when we are in the grip of fear, or revulsion, or a contrary temptation, for example. But it would be hasty to let this way of talking convince us that these are genuine inability to do those actions, since, if “ought” implies “can,” we will hastily be committing ourselves to the moral position that there is no requirement to do them. To postpone that moral question, we should, I think, remain open to supposing that these agents can still do these actions even if they can’t will them or “bring themselves” to do them. We’ll come back to this point after I have argued that even motivational incapacities that are part of human nature (if any) are not requirement-blocking—that is, that “ought” does not imply “can will.”

Our question is this: does it refute a requirement if it is shown that the agent, while (at least in all other respects) able to do the action, is unable to will it—unable to *bring herself* to do it? I take this to be one central meaning for the charge that certain theories of, say, justice fail to respect human nature with its characteristic selfishness or partiality.

### III. *Human Nature vs. A Socialist Theory of Justice*

Let’s look at a specific theory of justice that requires a strong kind of impartiality or selflessness. Joseph Carens has invented a fascinating and (he admits) utopian social arrangement.<sup>6</sup> His guiding question is how distributive equality might be compatible with both efficiency and individual

freedom. He doesn't assess these desiderata, alone or together, as principles of justice, although it will be useful for our purposes to think of them in that way. I simplify things for my own purposes here, but there are essentially three parts to Carens's reconciliation of these three desiderata:

(1) The underlying *distributive principle* is equality of income, along with principles about individual freedom.

(2) *Institutionally*, there is taxation or another means of redistributing income that is earned in the familiar market manner, so that income is equal (this can be measured per household or per individual, it doesn't matter here).

(3) Finally, there is a *behavioral* requirement: citizens are required to devote themselves to maximizing their pretax income, even knowing that it will be redistributed equally.

Carens's interest in this kind of market was about the advantages of markets. My interest is in an interesting feature of such a conception of social justice. Call the theory that society ought to implement and comply with the Carens Market the "Pretax Max" theory of justice. Carens characterizes the theory as "utopian" to signal that he does not believe it to be relevant to actual policy in any existing society.<sup>7</sup> He does argue, however, that there is reason to believe that with proper socialization people might well develop the motivational and behavioral characters needed for the reconciliation of equality, efficiency, and freedom.

Suppose, however, that people, in virtue of human nature, would not be able to get themselves to comply. Suppose people can't, even with the most propitious socialization, bring themselves to devote their energies to the common good in this way—they can't fully will it. Would that be requirement-blocking, thus refuting the Pretax Max theory of justice? I will argue that it would not.

It is important to acknowledge that any institutional proposal that ignores the facts about how people will actually tend to behave is worthless. It might seem as though the Carens Market is an institutional proposal and, as such, worthless for this reason. However, the Carens Market is not proposed by Carens for actual implementation under anything like actual or foreseeable circumstances. It is not an institutional proposal. So, even if actual behavior would make such institutions a disaster, that is no objection to the theory.

But if the Carens Market is not an institutional proposal, what is it? It doesn't look like a fundamental principle of justice, a status held, perhaps, by the egalitarian principle of distribution. It is institutional, but not a proposal. To understand its status better, it will be helpful first to introduce Professor Procrastinate.

#### *IV. Interlude: Professor Procrastinate*

Following Frank Jackson and Robert Pargetter, consider the case of Professor Procrastinate.<sup>8</sup> He is duty-bound to accept and complete an assignment to write a book review for a journal. (Suppose he owes the journal, or he promised, or whatever.) But he is the sort of person who, even if he accepts the assignment, will put it off and never do it. Should he accept the assignment? Many agree with Jackson and Pargetter that Procrastinate ought not to accept. This forces them, awkwardly, to accept both that he ought not to accept the assignment, and also that he ought to accept and perform the assignment. If you hold that this is not only awkward but self-contradictory, then you would be making a logical objection.<sup>9</sup> Accordingly, Jackson and Pargetter defend the consistency of these seemingly discordant positions with the following claim about the logic of "ought" statements: "ought" does not distribute over conjuncts. That is, even if it ought to be the case that A and B, it does not follow, as it might at first seem to follow, that it ought to be the case that A. That would be to treat "ought" distributively, and this is what they reject. By rejecting it, they save their claim that the professor ought not to accept the assignment, even though he ought to accept it and perform it. If

“ought” were distributive, then it would of course be problematic to say both that he ought to accept and perform and that he ought not to accept. For my purposes, all that matters is that distributivity should be rejected, not whether it is correct to say that Procrastinate ought not to accept. I will have to wait to explain why that is so, but for now I want to argue briefly that distributivity is properly rejected, leaving a fuller discussion to another occasion.

This issue is closely related to the familiar “problem of second-best.” When there are several desiderata that are desirable as a package, if one of them is not satisfied, the value of the rest of them is thrown back into question.<sup>10</sup> More generally, values are often holistic in this way. It is good to have steak sauce and steak. Is it good to have steak sauce? Not necessarily. It depends on whether you will also be having steak. Reflection on this very general structure of values is, I believe, a strong answer to the understandable impression that if Professor Procrastinate ought to accept and perform, then it would be contradictory to say that he ought not to accept.

The Professor Procrastinate example was devised not only to dispute the principle of distributivity, but also to support the case for “actualism,” the view that a person ought to perform whichever action is such that things will turn out best given what will actually happen (i.e., he won’t perform) rather than what could happen (i.e., he could perform even though he won’t), as “possibilists” argue. This is not the place to explicate, much less attempt to settle, the dispute between philosophers who agree with Jackson and Pargetter that Procrastinate ought not to accept the assignment (“actualists”) and others who argue that he ought indeed to accept (“possibilists”).<sup>11</sup> I take no position on that, but only on the narrower question of whether “ought” is logically distributive over the components of a conjunction. This does not commit me to actualism for the following reason. It is not necessary to agree with the actualists that Procrastinate ought not to accept in order to hold the following two things that I want to hold: First, contrary to the principle of distributivity, it doesn’t follow from the requirement to do A and B that he is required to do A. He might or might not be required, but nothing can be inferred about this from the conjunctive

requirement. Second, if someone holds that society ought not to implement some arrangement (call this implementation A), this would not refute a theory of justice according to which society ought to implement the arrangement and comply with it (call the compliance B), since “ought” is not distributive. So, I am not siding with or against the actualists, who argue that Procrastinate ought not to accept. I am holding only that *if* Procrastinate ought not to accept, this leaves open the possibility that, nevertheless, he ought to accept and perform. So, in the case of the Pretax Max theory, I am *not* arguing that it ought not to be implemented (supposing, for argumentative purposes, that the theory is true). My position is only that even if, as many will believe, it ought not to be implemented because it will not be complied with, the claim that it ought to be implemented and complied with is not contradicted. So, the fact, if it is one, that we shouldn’t institute the Carens Market because people won’t comply doesn’t refute the theory of justice.

We can now return to our earlier question: if the Carens Market is not an institutional proposal, what is it? Let’s distinguish between three things:

*Fundamental political principles:* These are abstract principles without institutional content such as a distributive pattern (equality, priority, sufficiency, and the like), historical principles (free transfer, legislative proceduralism, and the like), or combinations of these.

At the more institutional end of the spectrum:

*Institutional proposals:* These propose the implementation of rules and arrangements such as election and legislation procedure, economic rules and regulations, laws of property, marriage, employment, and much else.

Distinct from either fundamental principles or institutional proposals is what I shall call:

*Institutional principles:* An institutional principle describes institutional arrangements as part of a broader prescription or proposal, even if the described arrangement itself is not proposed or prescribed. In the form that matters for our purposes, an institution might be described as one that ought to be instituted and complied with, even if, because it will not be complied with, it ought not to be instituted. Thus, it is not an institutional proposal.

The Carens market is a case of institutional principle. It is institutional but it is not a proposal to build the institutions—because they won't be complied with.

Summing up this point: The Pretax Max theory requires us to implement and comply with a Carens Market. It is observed that we will not comply with it, and so we ought not to implement it. But we still ought to build and comply with it, and so the Pretax Max theory of justice and all its requirements are intact. It never did require us to implement a Carens Market. It required the more involved thing: implement and comply. In general, my claim is that institutional principles, which tell us to build and comply, are unlike institutional proposals because they are *not refuted by any facts about whether people will build or comply with them*.

But what about human nature? This discussion of the Carens Market has so far not engaged with the questions about human nature that are my central topic. We have, for the moment, only considered the fact that people *won't do* as the Pretax Max theory requires. As we saw earlier, that fact would not block the requirement. Our question is whether a certain stronger claim is requirement-blocking, namely, the claim that people, by virtue of human nature, are unable to bring themselves (at least in sufficient numbers) to behave in the impartial and energetic way that theory requires: maximizing pretax income when it will be redistributed equally. So, the position to be contended with is not only that we shouldn't implement the Carens Market (a claim that is

compatible with the truth of the theory of justice), but also that the theory of justice is false anyway. The argument I have in mind goes like this:

It's not only that people won't comply with such a market. That would allow that the theory of justice might be true even though, knowing what we do about people, we ought not to implement it. It is worse than that for the theory. It is *human nature* to be more self-interested than that. That is, it is not just a prediction but a claim about something in us that explains our behavior, namely, a *characteristic motivational inability to will to behave that way*. This refutes the theory of justice because theories of justice must respect such facts about human nature.

Recall that we have considered "can't do" and we allow that it is requirement-blocking. We looked at "won't do" and concluded that it is not requirement-blocking. We are now looking at "can't will." The question is whether this is requirement-blocking.

#### *V. Selfishness Is Not Requirement-Blocking, and Typicality Adds Nothing*

There is important variety among cases in which people can't bring themselves to act in certain ways. Some cases are "clinical" such as addictions, phobias, compulsions, and the like. I will put these cases aside and allow that, for all I say, it might be that some such cases are indeed requirement-blocking. I don't need to take a stand on this because I will assume that any motivational incapacities that are part of human nature are not in that (admittedly vague and intuitively defined) category. It is only these characteristically human motivational incapacities that I will argue are not requirement-blocking. In particular, I concentrate on (what I grant for the sake of argument to be) characteristic and essential human selfishness.

Before looking at characteristic and essential human selfishness, let's look at the selfishness of a single person. Consider a plausible moral requirement to refrain from dumping your household garbage by the side of the road. Suppose Bill pleads that he is not required to refrain from dumping because he is motivationally unable to bring himself to do it. There is no special phobia, compulsion, or illness involved. He is simply deeply selfish and so cannot thoroughly will to comply. Dumping his trash by the road is easier than wrapping it properly and putting it by the curb or taking it to the dump. He wishes he had more willpower, and yet he doesn't have it. Refraining is something he could, in all other respects, easily do, except that he can't thoroughly will to do it. It would be silly for Bill to propose this as requirement-blocking. This motivational incapacity is patently powerless to block the requirement in the individual case. (Notice that if we had let "can't will" count as "can't do," then the principle that "ought" implies "can" would have yielded the answer that the requirement is blocked here. This is counterintuitive, and there is no sufficiently weighty reason to proceed in that way in light of this implication.) If Bill's selfishness doesn't block that requirement, the reason is that it simply doesn't block any requirements. If, in addition, he is too selfish to comply with a Carens Market, then this is again no challenge to that institutional principle or the associated theory of justice at all. I allow, for the sake of argument, that some requirements are refuted by their being too burdensome or demanding, or some such thing, and I will return to the question whether this is itself a concession to human motivation incapacities (hint: no).

I said there were two parts to the idea that a Carens Market requires things that are incompatible with human nature. The first part is a claim of motivational inability due to selfishness. I have argued that in an individual case this is not requirement-blocking. The second element is the claim, which we are granting for the sake of argument, that this motivational inability is *typical* (or even essentially typical) of humans. So, even if the inability is not requirement-blocking in an individual case, what about if it is typical of humans, or even of humans as such? I see no reason to say anything other than that:

*Typicality adds nothing:* If S's being unable to will  $\emptyset$  is not requirement-blocking, then this is still the case if all humans are (even essentially) like S in this respect.

Suppose people line up to get your moral opinion on their behavior. Bill is told that his selfishness is indeed a motivational incapacity, but that it does not exempt him from the requirement to be less selfish. But behind Bill comes Nina with the same query. Again, we dispatch her, on the same grounds as Bill. Behind Nina is Kim and so on. Since each poses the same case, our judgment is the same. The line might contain all humans, but that fact adds nothing to any individual's case. I take this to show that even if the reason people will not comply with Carens Market is because there is a motivational inability to do so that is part of human nature, this is not requirement-blocking and so does not refute the theory of justice. Its requirement to build and comply with a Carens Market would not be refuted. I take this to show the following rather significant thing:

Even if a large dose of selfishness is part of human nature, this does not refute theories of justice that require people to be less selfish than that.

What would it mean for selfishness to be not only typical but essentially typical of humans?<sup>12</sup>

One possibility is that an agent does not count as human if their motivations are otherwise. There are historical examples of people who were extremely unselfish, such as Mother Theresa. The objection we are considering is this: a theory that requires people to be as unselfish as Mother Theresa is false because anyone who is that unselfish does not count as human. But this is absurd.

## *VI. Are Prerogatives a Concession to Human Nature?*

It would be supremely difficult for many parents to surrender their infant children to be raised by the state. Perhaps people could be socialized so that fewer parents felt this way and so that those who did feel it did so less intensely. Or perhaps not. I don't want to try to resolve this, but to accept for the sake of argument that human nature would resist such socialization and that many parents would always be unable to bring themselves to give up their children. Few accept that justice would require that children be raised by the state, and it may seem as though this is a concession to human nature: humans cannot bring themselves to surrender their children and so, for that reason, justice does not require it. This would be a concession to human nature of precisely the kind I am denying.

Indeed, a similar point can be made against the Pretax Max theory itself. To maximize pretax income, a person would have to forgo any projects that did not (directly or indirectly) contribute to that goal. Maybe reading bedtime stories to one's kids increases one's productivity, but if not, it would have to be skipped. Spending a long evening over a fancy meal, or buying a nice camera with money that could be used to take more professional development courses would be forbidden. Isn't it absurd to require maximizing pretax income if that means never reading to your children at night, or never going out to dinner with your partner, or never buying a nice camera? Donald Moon writes, "The composer who lives to write great music, but whose work is not appreciated, would be failing to perform his or her social duty. The person who chooses a life of scholarship over investment banking is morally blameworthy under this principle. We may well wish to demand some minimal level of social service from the composer or the scholar, but Carens's moral incentives seem to require far more than that."<sup>13</sup> Carens himself concedes the point and proposes a revised ideal in which people are obligated to contribute to the common good, but not necessarily as much as they can.<sup>14</sup>

Similar concerns have been pressed against highly demanding moral theories, and many believe that it is plausible to accept, in effect, a moral prerogative to pursue certain of one's own ends to some extent, even when this is not the optimal way to contribute to the agent-neutral good.<sup>15</sup> There

are a number of different possible positions in this area. Notice that Moon's examples do not militate for a prerogative to do (to some extent) just as one wishes but for a prerogative (or maybe even a duty) to do certain things that have significant value even if they do not maximize one's contribution to the social product. The rationale for that kind of prerogative might be derived from a theory of value, not from any observation that humans, by their nature, are too self-concerned to be able to bring themselves to meet the original common-good-maximizing requirement. There are various concerns an agent might have that might seem to have a claim on the agent even when they conflict with some more impartial or social good, including one's own aims, one's own valuable aims, the welfare of loved ones, local duties in the context of a family or neighborhood, and others.

Without venturing too deeply into these matters here, I will accept for the sake of argument that Carens's original duty—to maximize pretax income—is refuted by reflection on the ways in which it is, as he puts it himself, “too demanding.”<sup>16</sup> Let's say, then, that “ought” implies “not unreasonably demanding.” The question for present purposes is whether this would be a concession to human nature. More specifically, is this to be granted on the grounds that humans (by their natures) will be unable to bring themselves to conform to such a demanding requirement? Carens conjectures that people might well be able to bring themselves to conform if society has moved along a path in which they have been properly socialized to understand and care about their duties to others in this way.<sup>17</sup> So, when he grants that the original duty is too demanding, it is no concession to human nature. He says, “The kind of work that a person finds most meaningful and important may not always be the work that contributes most to the productive output of society. A principle that gives absolute priority to society's demands is too demanding.”<sup>18</sup> In other words, whether or not all people, or even any people, will especially value work that is not socially optimal, any who do are permitted to pursue that work to some extent even at some cost to the common good. What moral basis one might give for this moral judgment is a complex and difficult question. What matters here is that there is no indication that the reasoning must proceed from observations about what humans can or cannot get

themselves to do. The suggestion is not that the more partial motivations are widespread or natural but that they are morally sound or reasonable, and this is a substantive moral judgment about their content.

Indeed, if the argument for a less demanding duty were to proceed from the observed motivational incapacities of humans, it would be a deeply troubling argument, having the following form:

This tendency to partiality is what people are motivationally like, as a matter of human nature. *Therefore*, requirements to be otherwise are specious and false.

If that's a good form of argument, then the following is a good argument:

People tend to a certain degree of cruelty, and this is part of what they are motivationally like as a matter of human nature. (Suppose this is so.) Therefore, requirements to be otherwise are specious and false.

That is an absurd argument. The form of argument is bad, and so no prerogative in favor of partiality can be inferred from the fact, if it is one, that humans are naturally partial or selfish. I conclude that if the partiality of the parent is permissible when he refuses to deliver the infant to Plato's nursery, this is not because it is a characteristic feature of human motivation. It must be for some other reason.

Perhaps it is because the theory of justice is flawed in other ways that undermine this aspect of it. Or maybe there is a deep moral case that can be made for the proposition that *whether or not people are typically partial in this way*, it would be permissible for them to be so. That, manifestly, is not a principle that depends on whether people are, in fact, like that.

I am denying that principles of political justice depend on facts about human nature, construed specifically as facts about what people can or can't bring themselves to do. No such facts are requirement-blocking. Some readers will be reminded of G. A. Cohen's provocative argument that all normative principles ultimately have a fact-free basis.<sup>19</sup> I am not committed to his more general thesis. I take no stand on whether the search for deeper grounds will ultimately rest on facts or not. I have argued only that facts about characteristic human motivations neither block putative requirements of justice, nor are they the basis for whatever prerogatives of partiality there might be. Cohen allows that institutional proposals, which he calls "rules for the regulation of society," are highly sensitive to facts of all kinds, and I agree. So, it might seem that Cohen has limited the class of fact-free principles to noninstitutional principles such as fundamental principles of justice. What I have argued is that institutional *principles*, such as those that make up the Carens Market, are not sensitive to the facts that institutional *proposals* would be sensitive to, such as facts about human motivational incapacity. The important line is not between the institutional and the noninstitutional, but between principles and proposals.

#### *VII. Circumstances of Justice and Strains of Commitment*

On one way of thinking about justice, it is, essentially, not fully ideal. Justice is a virtue that a society might possess in light of certain circumstances, but which loses interest or application if those circumstances were overcome. In the classic conception of "circumstances of justice," deriving from Hume and Rawls, only a society that confronts problems such as scarcity and limited altruism confronts any questions about justice. Should these difficulties be overcome, the issue of justice is overcome as well. It is natural to assume that any human society will indeed be characterized by these conditions and that justice will therefore always be an issue. However, we may not reason the other way around: observing the conditions we expect never to be overcome, then building these for that reason into our conception of the circumstances of justice. It is not because certain conditions

will persist that they shape the very idea of justice, which would oddly be to assume that injustice itself will not similarly persist. It would also be to allow even the most evidently justice-tainting facts, if they are likely to persist, to drag justice down to their level. So, there is no basis in the idea of circumstances of justice for observing (if one does) that humans, by their nature, have certain motivational incapacities, and for deriving from this the conclusion that they do not count against justice.

There remains the question whether a theory such as Pretax Max depends on motivations that go beyond the constraints of the circumstances of justice, a question which is not at all settled by the supposition that they will never, in fact, be overcome. "Limited altruism," we can suppose, is a circumstance of justice. But the motivations required to meet the Pretax Max theory hardly overcome limited altruism. People are assumed to work as productively under the egalitarian tax scheme as they would under a less egalitarian one, but each can still be assumed to have an interest in having more rather than less of the scarce social goods. The change in motivations is, as compared with overcoming limited altruism, modest, even if we might doubt that it will ever come about. I conclude that a theory does not move beyond the circumstances of justice merely by depending on motivational structures that are, owing to human nature, never going to come about.

There is a feature of Rawls's theory of justice that is more plausibly at odds with the view taken here. By conceiving of the principles of justice as the outcome of an imaginary contract, Rawls finds an elegant way to account for the conviction that the claims made on individuals by the standards of justice should not be too demanding. In particular, if the parties to the original position were to know of motivational incapacities that are part of human nature (and as such not overcome in any institutional setting), they could not honestly commit, as parties to a contract, to complying with demands that conflicted with these known incapacities. If the demands of a conception of justice such as Pretax Max were beyond human motivational capacities, then the parties would not commit

themselves to it, which means, on Rawls's theory, that it does not give us the correct principles of justice.<sup>20</sup>

Surely, society should not implement institutions that people will not be able to bring themselves to comply with (assuming their value depends on that compliance). The question is whether that is a constraint on the content of justice. The rules and institutions that should be constructed given what is known about everyone's likely compliance are hardly guaranteed to be rules and institutions that qualify a society as just.

Suppose, for example, that it is part of human nature that people who recognize their own superior talents will tend to resent and envy people who are, owing to the operation of social institutions, as well-off as they are but without as much talent and ability. This resentment, suppose, would undermine the levels of allegiance and compliance on which the operation of those institutions depends. Rawls, of course, would have doubted that this is so, but his theory of justice has a distinctive response to such hypothetical conditions nevertheless. If it were so, then the parties should (other things being equal) reject principles that do not tend to apportion the distribution of social goods according to levels of talent and ability. Not only should society so apportion things under the circumstances on Rawls's approach, but this apportioning, according to the Rawlsian theory, would constitute the content of perfect social justice. It is this last point that is the crucial and, I believe, damaging one. The implausibility does not lie in the suggestion that in such unfortunate conditions there might be important reasons to accommodate those untoward but ineluctable motivational constraints.<sup>21</sup> But what institutions a society should implement given such practical constraints can hardly be thought to deliver the content of the idea of social justice. As I say, I am not relying on Cohen's sweeping argument that Rawls goes wrong by letting justice depend on any facts whatever. Whether or not that is correct, it is doubtful that the content of social justice is sensitive in this way to untoward motivational features of people. Rawls's doctrine of strains of commitment in his

contractual framework silences concerns about whether some motivational structures—however much they might be part of our natures—might be justice-tainting rather than justice-shaping.<sup>22</sup>

The human nature constraint, as I stated it, is a universal generalization and is refuted if there are any motivational features of humans by nature that are not (or would not be), as such, constraints on the content of justice. I have argued that certain motivational features that are themselves moral defects are counterexamples of that kind. Strictly, this leaves open the possibility that some other characteristically human motivational features do, as such, constrain the content of justice. However, I take the counterexamples to refute that possibility as well. So, there is a weaker human nature constraint that I have also tried to refute, which says that at least *sometimes* a motivational feature's status as characteristic of humans by nature constrains the content of justice. I have argued that, on the contrary, it is *never* a feature's status as characteristic of humans by nature that constrains the concept of justice, since a judgment is also required in every case about whether the feature's moral value or significance suits it to have this kind of weight.<sup>23</sup>

Limited altruism, as I have said, is plausibly among the circumstances of justice. The ideas of justice and injustice only have application for beings of such a kind. So, justice cannot coherently call for the elimination of limited altruism. It takes it for granted. However, this does not mean that justice capitulates to the existence of limited altruism in human nature. Facts about human nature are not determining the content of requirements of justice even in this case. Rather, since justice has no application except in circumstances of limited altruism, if humans were not that way (i.e., for any beings that are not that way), justice would fail to apply to them. The content of justice, however, might—for all that is implied by the points about circumstances of justice—precede any facts about what humans or any beings happen to be like, including whether limited altruism is part of their nature.

We aren't looking for an account of justice for martians or angels but for humans. One often hears this observation used against moral or political theory that would only be satisfied if people were different. The idea seems to be that if we want to understand justice-for-humans, then we must ground it in the nature of humanity. If that's right, then I must be stuck with a justice that is *not justice-for-humans*, since I deny that justice should be grounded in human nature. But, while I may be stuck with an account of justice according to which people, owing to human nature, will not bring about justice, that hardly shows that the theory isn't about what justice would be for humans. Human justice might well be something that human societies will fall short of. What makes it justice-for-humans is simply that it applies to them and is normative for them. If they ought to meet certain principles, then those are principles-for-them.

By eschewing any foundation of justice or morality in human nature, my approach shares something significant with Kant's moral philosophy. Kant is rather vehement that morality is grounded not in human nature, but in rationality. Imperfectly rational beings (unlike nonrational beings) are subject to requirements or imperatives stemming from reason, and among these is the categorical imperative, the fundamental principle of morality. Since humans are imperfectly rational beings, they are subject to moral imperatives. So, morality, whatever else it is, is morality for humans. It might also be morality for martians, if there are any and they are imperfectly rational beings. So, observing that we want an account of morality-for-humans or justice-for-humans offers no particular support for the idea that morality or justice must be constrained by human nature. Kant writes, "we must not let ourselves think of wanting to derive the reality of [the principle of all duty] from the *special property of human nature*. For, duty is to be practical unconditional necessity of action therefore (to which alone an imperative can apply at all) and *only because of this* be also a law for all human wills."<sup>24</sup>

Putting my thesis in Kantian terms, those to whom we owe justice do not lose their claim on us just because it might turn out that we are not, perhaps even by our nature, disposed to deliver it. Our

own will is no more authoritative over the content of the requirements of justice than theirs is. They are just as “legislative” as we are.

### *IX. Is Human Nature Excusing?*

One kind of defense of behavior that is caused by a motivational incapacity or difficulty is, as we have seen, to hold that the motivational structure (say, one that is not perfectly impartial) is substantively sound or reasonable. This seems to be what Carens has in mind by saying, “A principle that gives absolute priority to society's demands is too demanding.”<sup>25</sup> Perhaps such a case could be made in defense of the reluctance to surrender a child to the state, though how typical the reluctance is would not be directly to the point since some unreasonable motivations might be common or even natural. A different kind of defense can be illustrated with the example of the Skywalk at the Grand Canyon. This horseshoe-shaped glass-bottomed structure extends out from the edge of the canyon at a point that is four thousand feet above the river. Needless to say, some people will never want to go near it, and some who do won't be able to bring themselves to walk out onto it. Suppose someone could only save some lives by crossing such a high glass bridge, but can't bring themselves to do it. Is that motivational incapacity requirement-blocking? Should we say that they are not morally required to save those lives because they are unable to bring themselves to do it? If so, might human nature block requirements of justice in a similar way? The challenge for my argument is this: if the glass bridge case is requirement-blocking, then why aren't motivational incapacities generally requirement-blocking?

What is relatively clear is that the fear of heights mitigates the agent's liability to blame to some extent, which means that there might be similar mitigation in contexts of injustice due to human nature. This, a plea for excuse, would be the second kind of defense behavior owed to motivational incapacity (the first being to argue that the motivations are substantively sound or reasonable). Even

without trying to say which motivations are excusing and which are not, two points about this line of defense are important here.

First, even when certain motivations are excusing, this is still no basis for thinking that the mitigation or excusatory power is owed to their being widespread or part of human nature. Recall the line of people behind Bill pleading unsuccessfully (so I claimed) that their acts were justified (not merely excused). If Bill moves over to another queue, to plead only for excuse, his case is not strong or weak depending on how many people line up behind him ready to make the same plea. If a certain fear of heights excuses the failure to step onto the glass bridge as required, this is so whether or not anyone else (or all other humans) suffer from the same fear. Human nature, in the sense under discussion in the article, is not germane. Just as I argued in the case of justification, letting excuses be set by human nature would rely on an abhorrent principle that would excuse cruelty or bigotry if only they are characteristic or natural.<sup>26</sup>

Second, the category of excused acts includes acts that are not justified.<sup>27</sup> They remain wrong even as the agent's liability to blame is mitigated. For example, you might not be to blame if it is a fear of heights that kept you from saving someone, but that doesn't justify it. In our terminology, it isn't requirement-blocking. The morally requiring reasons remain in place, and a properly oriented agent will continue to appreciate them. It is allowed, by the appeal to excuses, that the requirements of justice are still not bent or relaxed by excusing or mitigating factors such as motivational incapacities. In that case, my thesis stands unchallenged: requirements of social justice are not blocked by facts (if there are any) about human motivational incapacities. That is not rebutted even if the failures are, to whatever extent, excused.<sup>28</sup>

Many will ask: but then, who cares what justice requires if failures are (supposing, for the sake of argument) excused and thus not blameworthy?<sup>29</sup> The answer, I believe, is that the agent must care. If the failure to save someone out of a fear of heights is excusing, then presumably the agent will have been motivated (incompletely, of course) by the morally requiring reasons. In this case, for example,

we might say that the requiring reason is the fact that you could save an innocent life with very little cost to yourself.<sup>30</sup> If the agent is not motivated by this consideration, then it is not true that what has prevented him from acting is his fear of heights. When we say that the agent's failure was caused by his fear of heights, we seem to be assuming that the cause is not any failure to appreciate the moral reasons to act. Were there such a failure, the fear of heights, even if present and sufficient to cause inaction, would not be excusing. A morally proper practical orientation involves a proper motivational engagement with moral reasons for action (including any requiring reasons). The fact that a failure to act as required might yet be excused, in some cases, by such things as a fear of heights does nothing to delete the role of the morally requiring reasons in a morally proper practical orientation. Facts about what is morally required, or required by social justice, do not fade to irrelevance just because an agent, or even human agents generally, are, to whatever extent, excused for failures to do as required. The question of what justice requires remains important. And it would be a moral deficiency in any agent subject to those requirements (whatever excuses might or might not come into play) if they did not care what justice required.

We can allow that a fear of walking on the glass bridge can be excusing, because it doesn't necessarily indicate a morally improper practical orientation even though it sometimes leads to morally wrong acts. There are variants of such a fear, however, that are morally more troubling. Suppose that you cannot bring yourself to cross to save an innocent person, but you can do so to save any cat. Or you can bring yourself to cross but only to save someone of your own race. Either of these might indicate morally deficient motivations in a way that a more general fear of heights does not. The fact that a failure to cross was produced by a fear of heights does not ensure that the agent is excused.

In the case of social justice, we must similarly recognize that even if some failures produced by motivational incapacities are excused, many are not. How much partiality is compatible with a morally proper practical orientation, for example, is something I am leaving open, except in my

insisting that there is such a thing as too much. Consider cases in which social justice is not met, even though it is within a society's abilities to do so, because people can't bring themselves to do what is needed. In principle, this failure might be entirely excused since it is possible that there is no morally inappropriate orientation to the moral reasons. Even in the case that the failure is excused, however, the standard of justice is unbowed. And that standard remains important too, as something that must matter to moral agents even if they might, in the end, be excused for failures. But, of course, the failure to meet the standards of social justice might not even be excused, since it might be owed to a morally improper orientation to the relevant standards and reasons. This might indeed give rise to motivational incapacities, but they would not, in that case, be excusing. If people can't bring themselves to do as social justice would require, then, while it might be that they are properly motivated by reasons of social justice but excused by something akin to a general fear of heights, it might instead be the case that they are not. It might be that their motivational incapacity is owed to a morally improper indifference to the legitimate interests of others. In this case, not only do the standards of justice stand, condemning the society as unjust, but in addition, the society is fully to blame (i.e., not excused) for the injustice, since no excusatory motivational incapacities explain the failure.

Some typical human motivational inabilities, such as the parental bond, might excuse certain kinds of failure to do as justice (nevertheless) requires, even if they don't justify (a question I am not addressing). Even if so, this would not refute the theory of justice that claims that the surrendering is morally required. Perhaps, of course, it can be refuted on other grounds.

What about people's motivational incapacity—supposing there was one—to work as productively under the egalitarian tax system as they would under a system that allowed the more able and talented to have more post-tax money? Is this morally like the person who could not bring himself to walk on the glass bridge to save anyone, including his own beloved son? In the bridge case, the motivational structure probably strikes us as an excusing condition. It is, however, rather

different from the doctor who can't bring himself to work an extra four hours a week in order to help an extra dozen patients, even though he would have worked that long and more if only he were paid slightly more—and more than others. In any case, that is the question.

It might be tempting even to deny that it is a motivational incapacity at all rather than just a simpler unwillingness to do as justice requires.<sup>31</sup> For my purposes, I am willing to let my opponent have that description of the case as a motivational incapacity. Even if that is granted, it is far from clear that it has excusing power (much less requirement-blocking power), since it's far from clear that it doesn't reflect poorly on the moral quality of the agent's motivations in justice-destroying ways. This is the question that is illegitimately silenced by supposing that justice takes motivational incapacities as given and shapes itself around them. Some motivational incapacities, if sufficiently widespread, and even if they were to characterize all humans by nature, undermine the very possibility of anything that deserves to be called social justice. To use Cohen's distinction again, principles of justice would differ from the most advisable principles for the regulation of society in that case by consisting in principles for the regulation of society that would be advisable if (at least) conditions weren't characterized by justice-damaging individual motivational incapacities.

Even if some motivational incapacities are excusing, this does not make them justifying, and so there are moral standards that survive intact. None of this would be affected by how widespread or natural the incapacities might be. Other motivational incapacities are not even excusing, much less justifying, since they reflect a morally deficient structure of concerns. Again, they would not become excusing, much less justifying, simply by being widespread or even natural. It remains open to argue that political theories that demand more than we expect people, in appropriate numbers, ever to comply with have specious requirements because, say, they require an unreasonable degree of impartiality or sacrifice. So, the view defended here is not essentially a kind of moral rigorism. It all depends on what grounds might be supplied for less strenuous moral requirements. My argument is meant to show not only that there would be no support in mere statistical facts about how people do

or might act, but that even appeals to natural motivational incapacity would not be decisive. The question is whether or not those incapacities are, for all their naturalness, forms of morally deficient will or concern. Vicious, or complacent, or selfish concerns are not somehow morally sanitized if they should happen to be characteristic of humans.

#### *X. Ability and Excuse in Individuals and Collectives*

We have been considering questions of individual excuse for failures to do what social justice requires. It is important to be clear about the relations between excuse and justification on the individual level on one hand and on the social level on the other. Social justice is a moral standard for societies and, strictly speaking, not for individuals. An individual cannot be characterized by any distribution of social goods or any institutional arrangement, and these are among the sorts of things that are assessed by standards of social justice. Requirements of social justice morally require things of societies as such. Obviously, whether a society meets certain standards will depend on how the society's members act. It is required, in the nonmoral sense, that they act in certain ways if society is to be just. There is an ambiguity in saying that individuals are required by social justice to act or refrain in certain ways. On one meaning, this is to say that a society's justice depends on how individuals act. This is not a moral judgment. On another meaning, it is to say that individuals are morally required to act in certain ways by standards of social justice. That is not settled by facts of the first kind. Individuals may be under certain moral requirements to promote or comply with standards of social justice (even though standards of social justice don't apply directly to them, but to societies), but that is not a question I will be considering in any detail. I have, however, been assuming, quite blandly I hope, that for any plausible standard of social justice, some morally bad individual behaviors, if sufficiently widespread, would entail that society fails to meet standards of social justice. This leaves open what moral requirements on individuals, if any, stem in some way from standards of social justice.

I have assumed that societies as such are not required to do anything they cannot do. The idea of ability, or what an agent can or could do, is familiar in the case of individual agents, but less so for collective agents such as societies themselves. I assume:

*The ability bridge principle:* Meeting some standard is within a society's abilities only if any individual behaviors that would be necessary if society is to meet that standard are within those individuals' abilities.

So, a moral requirement of social justice would be blocked if individuals could not behave in the necessary ways, because in that case society is unable to meet the standard. The principle that "ought" implies "can," in this sense, I take to apply at the collective level.

Of course, if individuals won't behave in the necessary ways, society won't meet the standard. I have argued, however, that even if people won't do something, that doesn't imply that they can't. Moreover, even if people can't will the act or bring themselves to do it, that doesn't imply that people can't do it. So, while society can't do something if people can't do what would be necessary, the mere fact that people won't do it or can't will it (i.e., bring themselves to do it) does not imply that society can't do what is putatively required, because it doesn't show that people can't do what would be necessary. Showing that people can't bring themselves to do something is not yet to show that they are unable to do that thing, and its relevance for refuting standards of justice is blunted in that way.

The ability bridge principle supplies the needed bridge between individual and collective ability, with its implications for requirements at the collective level. I turn next to excuse at the individual and collective levels. My central concern is standards of justice that apply to societies rather than moral standards that apply to individuals. Still, the question of morally excused individual behavior

is important for the following reason. While I leave open whether a society's injustice is itself ever excused, I am assuming at least:

*The excuse bridge principle:* If society's injustice is produced by individual behaviors that are morally wrong and not excused, then the social injustice is also wrong and not excused.

If the individual failures are not excused, then the social failure is not excused (though I take no stand on the converse proposition). Now, if society should fail to meet a standard of justice because of excusable individual failures, then perhaps society's failure is excused. I also take no stand on this question, since it is no challenge to my main thesis, that requirements of social justice are not refuted by motivational incapacities even if these happen to be owing to human nature. First, even if the incapacities are excusing at the individual level, this would not be owed to their being widespread or even natural (as I have argued). Second, even if individual excusable failures render the social failure excusable, the moral requirement of social justice stands unchallenged. And third, it is important to emphasize that many familiar individual motivational incapacities are not even excusing, much less justifying. Much of the behavior whose alleged ubiquity in human life is held against high standards of social justice is simply selfish, or bigoted, or otherwise powerless to excuse failures to act as social justice would require whether or not they are owed to human nature. By the excuse bridge principle stated above, the social failure would not be excused either.

## *XI. Conclusion*

Justice-for-humans could, in principle, set standards that are beyond human nature. It might require things that, while not beyond the abilities of societies or their members, are nevertheless beyond what individuals can, in virtue of human nature, bring themselves to do. Even if these motivational incapacities should be excusing, which (emphatically) is not guaranteed, the results would still fail to

meet moral standards that matter to properly oriented agents. If we are more optimistic about human nature than this, that would still be no defense of the theoretical approach to justice that would bend the standards of justice to whatever unfortunate motivational incapacities humans might turn out to have.

---

I am grateful to many people and audiences for helpful discussions of this work while it was in progress, including University of Oslo, Goethe University Frankfurt, the Political Philosophy Workshop at Brown University, CIDE and UNAM in Mexico City, Legal Theory Workshop at University of Pennsylvania, Canadian Political Science Association meetings in Montreal, the seminar of John Ferejohn and Janos Kis (whose comments were especially instructive) at the law school at New York University, the Harvard Government Department colloquium, and a lecture at Franklin and Marshall College. I am also grateful to Nomy Arpaly for numerous helpful discussions of these ideas, and to Charles Larmore, Sharon Krause, and two anonymous referees for valuable discussions of an earlier draft.

<sup>1</sup> See, e.g., Joseph Carens, *Equality, Moral Incentives, and the Market* (Chicago: University of Chicago Press, 1981), p. 104.

<sup>2</sup> Nicholas Sturgeon reports this observation as lore among Cornell graduate students. See “What Difference Does It Make Whether Moral Realism Is True?” supplement, *The Southern Journal of Philosophy* 24 (1986).

<sup>3</sup> Here I go through some points from “Utopophobia: Concession and Aspiration in Democratic Theory,” Chapter 14 of my *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2008).

<sup>4</sup> My concern, then, is with requirements on societies as such. Much will depend on points about abilities and motivational capacities of individuals, however, and in Section 11 I will be more explicit about the relations between these questions.

<sup>5</sup> Plato, *Republic*, bk. 5, lines 457c-d.

<sup>6</sup> Carens, *Equality, Moral Incentives, and the Market*.

<sup>7</sup> *Ibid.*, p. 3.

<sup>8</sup> Frank Jackson and Robert Pargetter, “Oughts, Options, and Actualism,” *Philosophical Review* 95 (1986): 233-55.

<sup>9</sup> It is only self-contradictory if deontic dilemmas are logically impossible, which is controversial. A weaker objection would be that distributivity would lead to a deontic dilemma where, intuitively, there is none. The denial of distributivity, as discussed in the next sentences, would respond to either form of the objection.

---

<sup>10</sup> The classic statement, formulated in economic terms, is R. G. Lipsey and Kelvin Lancaster, “The General Theory of Second Best,” *The Review of Economic Studies* 24 (1956-1957): 11–32.

<sup>11</sup> For two recent short discussions, see Christopher Woodard, “What’s Wrong with Possibilism,” *Analysis* 69 (April 2009): 219–26; and Ralph Wedgewood, “Against Actualism,” *PEA Soup*, <<http://peasoup.typepad.com/peasoup/2009/09/against-actualism.html>>.

Accessed February 10, 2010.

<sup>12</sup> Thanks to Paulina Ochoa for discussion of this modally stronger position.

<sup>13</sup> J. Donald Moon, review of *Equality, Moral Incentives, and the Market*, *Ethics* 94 (October 1983): 146-50, at p. 149. See the reply by Carens: “Rights and Duties in an Egalitarian Society,” *Political Theory* 14 (February 1986): 31-49.

<sup>14</sup> See Carens, “Rights and Duties,” p. 35.

<sup>15</sup> Classic pieces includes Samuel Scheffler, “The Rejection of Consequentialism” (Oxford: Clarendon Press, 1982); and Shelly Kagan, “Does Consequentialism Demand Too Much?” *Philosophy & Public Affairs* 13 (Summer 1984). For some use of these issues in political philosophy, see G. A. Cohen, “Where the Action Is: On the Site of Distributive Justice,” *Philosophy & Public Affairs* 26 (Winter 1997): 3-30; and David Estlund, “Liberalism, Equality and Fraternity in Cohen’s Critique of Rawls,” *Journal of Political Philosophy* 6 (March 1998): 99-112. See also Cohen’s replies to criticisms stemming from his granting of a “personal prerogative,” in *Rescuing Justice and Equality* (Cambridge, Mass.: Harvard University Press, 2008), pp. 385-94.

<sup>16</sup> Carens, “Rights and Duties,” p. 34.

<sup>17</sup> See *ibid.*, esp. pp. 103-8.

<sup>18</sup> *Ibid.*, p. 34.

<sup>19</sup> Cohen, *Rescuing Justice and Equality*.

<sup>20</sup> For Rawls’s original discussion of the strains of commitment, see *A Theory of Justice*, 1st ed. (Cambridge, Mass.: Harvard University Press, 1971), Section 29, pp. 153-60. In *Justice as Fairness: A Restatement* (Cambridge, Mass.:

---

Harvard University Press, 2001), the strains of commitment might seem to be calculated by attending only to whether principles guarantee the “fundamental interests” of people seen as free and equal citizens. See, e.g., p. 103. However, it is unclear how additional facts about human nature could appropriately be kept from the parties if they were facts established (say, by science) beyond reasonable dispute. If known, such facts might well affect which commitments the parties would believe they could keep and so make in good faith. This challenge to the theory is serious even if there happen to be no such troubling facts (a question I do not take up). I’m grateful to Erin Beeghly for pressing this difficult interpretive question, and I do not take this short note to sufficiently resolve it.

<sup>21</sup> See *ibid.* Cohen’s distinction between principles of justice and rules for the regulation of society is a perspicuous way of thinking of this distinction. The distinction figures throughout the book but is introduced in the first few pages.

<sup>22</sup> Cohen’s more sweeping thesis is less plausible (though maybe still correct) than this more limited argument. Letting justice be shaped by at least some facts is not initially troubling in the way that it is troubling to let justice be shaped by certain apparently justice-tainting facts such as the sort of talent-supremacist motivations I hypothesize. The approach taken here depends on identifying some subset of facts as justice-tainting, a larger project I will not attempt here. In effect, I believe this approach is suggested by Cohen’s earlier work (though later abandoned) about how a Rawlsian theory seems forced to let selfish (and thus, presumably, justice-tainting) behavior by the “talented” dull the egalitarian edge of principles of justice. See G. A. Cohen, “Incentives, Inequality, and Community,” in *Tanner Lectures on Human Values*, Vol. 13, ed. Grethe B. Peterson (Salt Lake City: University of Utah Press, 1992), reprinted in *Rescuing Justice and Equality*.

<sup>23</sup> It is often suggested that the putative requirements of highly idealized theory of the kind under discussion here must not count as normative because they are not “action-guiding.” I assume that there is no doubt about the normativity of the moral duty that Professor Procrastinate has to accept the assignment and write the review. So, that normativity is apparently to count as sufficiently action-guiding, even though that duty has nothing to offer about what, under the circumstances in which he is not going to write the review, he ought to do. It is action-guiding at one level: he ought to accept and perform. It might fairly be said not to be action-guiding on another level: ought he to accept if he will not perform? But this would never be held to count against the very existence of the duty to accept and perform, and so it is granted to be action-guiding in whatever way may be demanded of normative statements.

---

In that case, there's no problem about the action-guiding-ness of the Pretax Max theory of justice, which is structurally the same: society is required to build and comply with certain institutions. This is action-guiding, just as the professor's duty is. Obviously, showing that some putative requirement is no guidance with respect to *some* choices hardly shows that it fails *tout court* to be action-guiding in the way that is constitutive of normative considerations.

<sup>24</sup> Kant, *Grundlegung zur Metaphysik der Sitten*, Vol. 4 of the Academy edition, p. 425. This translation is from *Practical Philosophy*, ed. Mary J. Gregor, The Cambridge Edition of the Works of Immanuel Kant (Cambridge: Cambridge University Press, 1996), p. 76.

<sup>25</sup> Carens, "Rights and Duties," p. 34.

<sup>26</sup> It is important to remember here that the characteristics' being natural—meaning characteristic of humans as such (perhaps only as a statistical fact, such as right-handedness)—would not entail that it is beyond agents' ability to refrain. If that were the case, it would suggest an entirely different line of justification or excuse, and I've granted for the sake of argument that it might succeed.

<sup>27</sup> There is dispute about whether an act can be both justified and excused, but that doesn't affect my claim that not all excused acts are justified. For a recent argument that all justified acts are excused but not vice versa, see Mark McBride, "Justifications and Excuses: Mutually Exclusive?" *Journal of Ethics and Social Philosophy* (June 2011). <[www.jesp.org](http://www.jesp.org)>.

<sup>28</sup> On the distinction between excuse and justification, see Paul Woodruff, "Justification or Excuse: Saving Soldiers at the Expense of Civilians," supplement, *Canadian Journal of Philosophy* 2 (1982): 159-76; and Marcia Baron, "Justifications and Excuses," *Ohio State Journal of Criminal Law* 2 (2004-2005).

<sup>29</sup> I am indebted in what I say in this paragraph to Nomy Arpaly, "Moral Worth," *Journal of Philosophy* 99: 223-45, which is also Chapter 3 of her *Unprincipled Virtue: An Inquiry Into Moral Agency* (Oxford: Oxford University Press, 2003).

<sup>30</sup> Some will deny there is a requirement of that sort. Add, if you like, that you happen to be specially responsible, for some reason, for the person's safety. The precise form of this requirement doesn't matter for present purposes.

---

<sup>31</sup> I thank Thomas Nagel for this observation.