

## **Fact-Sensitivity and the ‘Defining Down’ Objection**

Andrew Lister  
Queen’s University  
andrew.lister@queensu.ca

Final Prepublication version: May 25, 2016

Final Published version: *Res Publica*, Vol. 23, 2017, 117-35.  
DOI: <https://doi.org/10.1007/s11158-016-9332-3>

### **Introduction**

The question of the relationship between facts and principles has been at the heart of recent debates over ideal and non-ideal theory. As noted by Stemplowska and Swift (2012, p.374), Rawls’s theory of justice has been accused of having a status quo bias both because it is too ideal (Mills 2005; Mills 2009), and because it is not ideal enough (Cohen 2003; Cohen 2008). This paper aims to clarify what it means for a normative theory to be ‘fact-sensitive’ (Farrelly 2007, pp. 844-47; Cohen 2008, pp. 281-82) and to explain what if anything is wrong with such sensitivity, by examining the ways in which ‘justice as fairness’ depends upon facts. The paper focuses on Rawls’s principles, and in particular the difference principle, not the methodology of the original position. I construe fact sensitivity as limited generality (Pogge 2008), and argue that there is nothing in general wrong with such limitations. The main exception concerns lack of generality due to facts about human psychology that are assessable from the point of view of justice, and which make otherwise superior principles infeasible. If it should turn out that the correct principles of justice cannot be stably realized because of some feature of human nature, it would be important for us to acknowledge this aspect of our psychology as something we regret. Because the character of relationships depends in part on attitudes, not just what ends up happening, this acknowledgement could be important even if at present greater justice is infeasible. Rawls argued that the fact of motivation by reciprocity made his principles more stable than utilitarianism. However, reciprocity also limits what justice

requires of us where assurance of compliance on the part of others is lacking. If reciprocity is simply a fact about human psychology that limits the range of feasible principles, then Rawlsians may be guilty of defining justice down. However, I argue that reciprocity can be seen as a fact about justice, to wit, that its point is to create a relationship of mutual respect, making strictly unilateral compliance pointless.

### **Constructivism and Principles of Justice**

In his 1992 Tanner Lectures, G.A. Cohen criticized Rawls's incentive argument for inequality on the grounds that it was inconsistent with the difference principle (Cohen 1992, pp. 268-9, 310). The incentives inequality provides are not strictly necessary to aid the least well off, for if the talented accepted the reasoning underlying the difference principle, they would not use their luck in the natural lottery to demand extraordinary wages for ordinary work. In 2008's *Rescuing Justice and Equality*, Cohen questioned whether the difference principle is a fundamental principle of justice at all, by arguing that constructivism confuses principles of regulation with principles of conviction. Constructivism is the view that 'the content of justice' is given by the principles that would be chosen in a privileged choosing situation, e.g. the original position (Cohen 2008, pp. 3, 274-5). Principles of regulation are rules of conduct justified by the consequences of their general adoption, while principles of conviction are normative standards for assessing such consequences. Asking what principles the parties would agree to allegedly conflates the two, allowing contingent facts and values other than justice to determine what we take justice to be (Cohen 2008, pp. 265-6, 275, 290-1). Rescuing the concept of justice from Rawlsian constructivism helps rescue an egalitarian conception of justice from Rawlsian liberalism (Cohen 2008, pp. 2).

I am not concerned to defend the original position, nor to claim coherence for Rawls's

account of the fact-sensitivity of fundamental principles. Many people are unconvinced by the argument from the original position, but remain wedded to Rawls's principles, or something in that neighbourhood. Such Rawlsians might not care if the original position turns out to have more flaws than previously recognized, or if Rawls's pronouncements about the dependence of principles on facts are unsound, but they should care if the principles they espouse have been watered down by the unrecognized admixture of facts and non-justice values. However, I do want to comment on Cohen's account of constructivism for two reasons: to clarify what a Rawlsian principle of justice is meant to be, and to introduce an idea that will be important in the rest of the paper, which is that acknowledgement of the justice-status of counterfactual possibilities can affect the character of social relationships.

Cohen presents Rawls with a choice: either principles define justice, or they provide rules for regulating conduct in view of outcomes that are desirable all things considered. Yet there is an intermediate possibility; principles might rank different values that fall under the general heading of social justice. Rawls's aim was to provide a systematic but non-utilitarian alternative to the intuitive balancing of precepts that characterizes ordinary moral thinking (Rawls 1999, pp. 30-36, 267-273). For example, we have conflicting intuitions about what a fair wage amounts to, involving multiple 'precepts' or 'maxims' of justice, e.g. effort, contribution, and need. Mill claimed that only the principle of utility could resolve such conflicts. Rawls argued that the original position would help us identify a superior, non-utilitarian alternative. The parties to the original position are not asked to design the institutional blueprint of the ideal society, all facts and values considered (ignoring only their own place in this structure). Instead, they are meant to select *first principles* for assessing institutions while ignorant of the specific facts about their society, reasoning on the basis of social primary goods alone. Only under these special conditions does Rawls claim that the principles chosen will be principles of justice. The

principles are for us rankings of justice-related considerations, despite being devices for bringing about valuable effects from the special perspective of the parties to the original position.

Recognizing this difference in vocabulary raises the question of whether some of the differences between Rawls and Cohen are merely terminological. To illustrate the evaluative impurity of the difference principle, Cohen imagines a case involving two persons with 5 units of manna each before 3 more units fall from the sky (Cohen 2008, pp.316-17). The only ways to divide this windfall are for A to get 2 and B 1, or for neither to get any, so that the feasible distributions are [5, 5] and [7, 6]. If justice is the avoidance of accidental inequalities, the just distribution is [5, 5]; [7, 6] is Pareto superior, but unjust. The difference principle may be a good rule for us to follow all things considered, Cohen admits, but it compromises justice for the sake of making everyone better off. This rescue of egalitarianism from Rawlsian liberalism is less heroic than it might appear, because on Cohen's account justice is no longer the first virtue of social institutions (Cohen 2008, pp. 302-06); Cohen has rescued egalitarianism only by limiting its importance (Valentini 2012, 658). Is it merely a verbal difference whether we say (with Rawls) that justice comes first, and that justice requires maximizing the level of the lowest social position, or (with Cohen) that justice requires equality, but that Pareto improvement trumps justice (Arneson 2008, p. 398)?

According to Patrick Tomlin, Cohen's vocabulary makes clear that in accepting institutions justified by the difference principle we are trading off one value against another. The distribution [7, 6] is all things considered preferable to [5, 5], but if we call the resulting inequality just, then we may conclude that there is nothing at all bad with [7, 6], losing sight of the fact that in choosing [7, 6] we are sacrificing equality for the sake of making everyone better off (Tomlin 2010, p. 243). Recognizing that [7, 6] involves an unfairness despite being justified

all things considered affects the relations between the persons in the two groups, Tomlin claims, as compared to the situation in which we think ‘since [7, 6] satisfies the difference principle, it is in no respect unfair to anyone.’ I accept that this is a genuine worry, but deny that it applies to the Rawlsian framework.

First, whether a given outcome satisfies the difference principle depends on the set of feasible options. Given that only [5, 5] and [7, 6] are feasible, [7, 6] satisfies the difference principle, but if [6.25, 6.25] were feasible, [7, 6] would *not* satisfy the difference principle. There is no barrier to recognizing that what is just in the limited current set of possibilities would not be just if an expanded set of possibilities were available. In fact, if people understand the rationale for preferring [7, 6] to [5, 5], as they are meant to do in a well-ordered society, they will know that [6.25, 6.25] would be preferable were it possible. As a result, they can say to each other ‘[7, 6] is just in the circumstances, even though it would not be just were [6.25, 6.25] feasible.’

Second, since Rawls’s principles are meant to rank competing considerations that are relevant to social justice, how we divide up principles is not crucial so long as we get the ordering right. We can say that there are two principles so long as we remember that the second principle has two parts, or we can say that there are three principles (equal basic liberties, fair equality of opportunity, difference principle). There is also no barrier to recognizing that the difference principle contains different components, each of which might have intrinsic importance. What the difference principle does is to establish a ranking between avoiding arbitrary inequalities and benefitting everyone. The fact that this happens within one principle is of no more significance than the fact that fair equality of opportunity is located within the second principle, as opposed to being a distinct principle.

Third, it may be that the less egalitarian interpretations of the difference principle get these priorities wrong, but this is only an objection to those versions of the principle. Cohen

pointed out that Rawls's various formulations of the difference principle were ambiguous as to whether inequalities were *required* or merely *permitted* if they *raised* or merely *did not lower* the lower social position (assuming for simplicity's sake a two-position case; Cohen 1992, 266, note 6). The least egalitarian version requires all inequalities that do not lower the lower position, and so does not ascribe any intrinsic importance to equality (in the dimension of social primary goods). The most egalitarian version does, because it permits inequalities only if they raise the lower position. According to the latter view, inequality is in itself a bad thing, but something that can be justified by making the worst off better off than they would otherwise be. The 'required' + 'not lower' version of the principle may be insufficiently egalitarian, but that is simply an objection to that version of the principle.

For these reasons, I do not accept that the evaluative impurity of the difference principle leads to ethical inarticulacy. Yet I want to keep in mind Tomlin's main insight, which is that it can matter to the quality of our social relations that we recognize not just which of our currently feasible set of options is most just, but what would be just in alternate, counterfactual conditions. Social relationships are affected not just by what ends up happening but by the attitudes we have towards one another. These attitudes are expressed by the reasons we accept, and publicly acknowledge as the bases for our political decisions. Given our current set of feasible choices, it may be that a variety of different reasons all support the same option. Even though our choice of policies doesn't currently depend which reasons we accept, it may be important for us to distinguish and acknowledge these reasons. Many different principles would support the preference for [7, 6] over [5, 5], e.g. maximize the average, the maximum, or the minimum. Knowing which of these principles we accept can affect our social relations even if currently [7, 6] and [5, 5] are the only feasible options, and for the moment our choice of principles has no

effect other than its constitutive relational effect.

The idea that it can be important to recognize what would be just in circumstances that do not presently obtain was not foreign to Rawls. Immediately after saying that contract theory agrees with utilitarianism that the choice of first principles depends on facts (Rawls 1999, p. 137), Rawls points out that the choice of principles is *less* fact-sensitive on the contract view than it is under utilitarianism, and that the mutual recognition of this partial fact-insensitivity has ethical importance. It would be important to embed ideals of justice in our first public principles even if we were certain that slavery would not maximize happiness (Rawls 1999, p.139). Our selecting the two principles of justice because they would be chosen in the original position expresses our nature as free and equal moral persons, whatever the facts of human psychology and society may mean for the level of happiness these principles will generate (Rawls 1999, p.226). Use of the original position was supposed to make it common knowledge that we would continue to support the equal basic liberties even if the utilitarian's empirical assumptions turned out to be false. Thus it can make a difference to our social relationships both that we recognize that what is just in our fortunate present conditions would still be just in *worse* conditions (Rawls; equal liberties would still be just even if they did not maximize happiness), and that we recognize that what is just in our present, *unfortunate* conditions would *not* be just in better conditions (Tomlin; arbitrary inequalities are only just because in the circumstances they are necessary to raise the worst off).<sup>1</sup>

---

1 Valentini argues (contra Farrelly) that the distinction between fact-sensitive and fact-insensitive principles is not a matter of degree but a dichotomy, because fact-insensitive approaches ask what is justice as such, independent of empirical facts, whereas fact-sensitive approaches ask what principles should govern the exercise of political power here and now, assuming various facts about society and human nature that make the exercise of political power necessary (Valentini 2009, p.335). Within Valentini's category of fact-sensitive approaches, however, we can distinguish degrees of fact-sensitivity. I am arguing that recognition of partial or relative fact-insensitivity can matter, ethically speaking. I will

### **Fact Dependence as Lack of Generality**

Rawls's principles depend on a number of factual assumptions. To limit the scope of the discussion, I will focus on the difference principle.

The difference principle depends on the factual assumption that the 'contribution curve' (Rawls 1999, p. 66) is not too flat or shallow, i.e. that the slope on its upward segment is not too small. The contribution curve is the function that relates increases in the level of higher social positions (due to decisions about the basic structure of society) to the level of the lowest social position, measured in terms of representative persons' shares of social primary goods. A very small positive slope means that large increases in the level of the better off bring small increases for the worse off. Considered in isolation, the difference principle justifies such inequalities because they raise the lowest social position, even though the facts are such that it takes big increases in inequality to generate small gains for the worst off. The difference principle is not universally valid, however, but holds only under two empirical conditions. First, it must not be the case that the inequalities that raise the lower position are so large that they undermine the first principle's guarantee of the fair value of political liberty. The priority of liberty sets a limit to the inequalities that would otherwise be justified as benefitting the least well off (Rawls 1999, p.70; see also Estlund 1998, p. 110). The proposition 'inequalities that raise the worst off are just' is true only where these inequalities are within the range permitted by the first principle's guarantee of democracy.

Second, it must not be the case that the inequalities that raise the lower position are so large that they undermine fair equality of opportunity. If it should turn out that raising the lowest

---

also later argue that we should distinguish facts that limit the realizability of justice from facts that make the concept of justice applicable.

social position a little requires raising higher positions a lot, this extra inequality might increase the intergenerational transmission of advantage. Higher levels of economic inequality do seem to lead to higher levels of intergenerational transmission of advantage (Corak 2013). The difference principle depends on this effect not being too large; the proposition ‘inequalities that raise the lowest social position are just’ is true only if these inequalities are not so large as to compromise equality of opportunity. The standard of equality of opportunity is of course a normative standard, just like the fair value of political liberty is a normative standard. Yet it is an empirical, economic or sociological fact how steep or flat is the function that relates increases in inequality between positions to the level of the lowest position.

The fact-dependence that results from the difference principle’s subordination to prior principles provides a possible justification for Rawls’s claim that the difference principle ‘assumes the operation of other principles’ and ‘presupposes as well a certain theory of social institutions,’ in particular ‘the idea that in a competitive economy (with or without private ownership) with an open class system excessive inequalities will not be the rule’ (Rawls 1999, p. 137). Cohen takes issue with this statement on the grounds that there must be a prior principle that defines what counts as ‘too much,’ and that this more general principle would not be dependent on the economic facts whose relevance it explains (Cohen 2008, p.259). From the perspective of citizens committed to Rawlsian principles, the more general principle is the first principle’s guarantee of political liberty, and the principle of fair equality of opportunity’s limitation on intergenerational transmission of advantage. However, in the passage in question Rawls implies that the fact that excessive inequalities will not be the rule is one of the ‘general facts of human society’ of which *the parties* are aware (Rawls 1999, p. 137). Since they are assumed to have already selected the prior principles, it is not clear why their knowing this fact should matter to their choice. With the principles of equal liberties and fair opportunity already

on hand, the parties should know that any subsequent principles they choose will only be binding where conditions are such that prior principles are secure. Conversely, if the parties *know* that the slope of the contribution curve cannot be so shallow as to make the difference principle violate fair opportunity / political liberty, why would they bother insisting on the priority of fair opportunity over the difference principle?

Whatever one makes of Rawls's methodological statements, the difference principle does depend on facts, and that this fact-dependence limits its generality. This result should not be surprising, since Cohen's own analysis of facts and principles leads to the same conclusion. Cohen maintains that a principle P can be grounded in a fact F only in virtue of a prior principle that explains why F is a reason for accepting P. For example, if the principle 'we should keep our promises,' is grounded in the fact 'only when promises are kept can promisees successfully pursue their projects,' that must be because of a prior principle such as 'we should help people to pursue their projects' (Cohen 2008, p. 234). The implication of Cohen's analysis is that the principle of keeping promises is not entirely general, but only valid only where promise-keeping is in fact essential to project-pursuing. Similarly, the difference principle is not generally valid, but only where inequalities that raise the worst off are not so big as to undermine the two prior principles.

The difference principle also depends on the assumptions of close-knitness and chain connection, but at least with respect to chain connection, this dependence is not explained by any explicit prior or more general principle. Close-knitness obtains when any change in one position affects all the others, i.e. it is not the case that we face a choice between between [5, 10] and [5, 12] (for example). Chain connection obtains when any inequality between a higher position and a lower position that raises the lower position also raises all intermediate positions, i.e. it is not

the case that we face a choice between [1, 5, 10] and [2, 3, 20].<sup>2</sup> Rawls initially insists that the difference principle is still valid where close-knitness and chain connection do not hold (Rawls 1999, p. 71), which would imply that the principle is in these respects a general principle. When chain connection fails, however, increases for the least well off can come at the expense of decreases for the next worst off. When the benefit to the least well off is small, and the cost to the intermediate position large, this strict priority becomes implausible. Is [1, 5, 10] really worse than [1.0001, 1.0002, 30]? Rawls later responds to the two-position analogue of this objection. The difference principle assumes that the ‘contribution curve’ does not have a very small negative slope, past the point of inflection, i.e. it is not the case that tiny gains for the worse off require massive losses for the better off (moving back toward the summit of the contribution curve). Rawls says that ‘the difference principle is not intended to apply’ to such situations (Rawls 1999, p.136). But this is just to say that the difference principle is not valid under these circumstances. Similarly, it is probably not valid where inequalities that raise the worst off a little lower the next worst off group a lot while exacerbating inequality overall. The difference principle is a special case of more general principle, one that Rawls unfortunately left unspecified.

One might object that we can’t know whether the more local, specific principle is correct unless we know the more general principle of which it is a special case.<sup>3</sup> Typically, however, we will have reason to believe a more specific theory before we know what general theory it is a

---

2 If close-knitness and chain connection hold, then when the difference principle is satisfied, any inequality between any two positions raises the lower position; (Rawls 1999: 69-70) Without close-knitness the worst off do not *benefit* from all inequalities, because where the contribution curve is flat over some range, additional inequality does not raise the lowest position. Without chain connection, the worse off party does not benefit from *every* inequality, because some people who are worse off than some others will lose from an inequality that is necessary to benefit those who are even worse off than themselves.

3 Thanks to an anonymous reviewer for this objection.

special case of. Einstein came up with the special theory of relativity prior to the general theory. Presumably one of the reasons for believing the general theory is that it explains the special theory. If Einstein had needed to know the general theory first, he would not have been able to come up with the special theory in the first place. Similarly, one could have reason to think that the difference principle is the right principle where chain connection holds, without yet knowing what principle is correct when chain connection fails. One of the rationales for ideal theory is to start with simpler, unrealistic cases before trying to move to more complex, real world scenarios, e.g. to figure out what would be just were we living in one isolated society, before tackling the much harder problem of justice in a world of interacting societies. Clearly, we do not want to misconstrue locally-valid principles as general principles. So long as we are aware of the ways in which our principles depend on facts, however, we will not be guilty of overgeneralization.

David Wiens has recently argued that the difference principle's lack of generality makes it unsuitable to guide political practice. Relaxing the assumption that we are dealing with a single closed society composed of people who have a sense of justice, Wiens imagines that capital but not labour is mobile, and that elites have no compunction about threatening to transfer wealth abroad unless tax rates are lowered and regulations gutted. In these circumstances, the slope of the contribution curve might be very shallow, in the sense that small gains for the worst off require massive gains for the better off. Wiens questions whether such inequalities would in fact be just.

An institutional scheme that satisfies the difference principle... publicly surrenders to the threats of a class of mercenary elites in exchange for a modicum of gain for the least advantaged. In so doing, such an institutional scheme acquiesces in the creation of an economic hierarchy that seems in tension with

Rawls's commitment to the moral equality of citizens and a stable system of cooperation for mutual advantage (Wiens 2015, 443).

This example is hard to interpret because it is overdetermined. If massive gains for the better off are required to generate small gains for the worse off it is likely that the additional inequality violates prior principles: fair value of political liberties, and fair equality of opportunity. Wiens stipulates that this is not the case, because (e.g.) all students are educated in public schools of equal quality. However, if it is really true that there is equal opportunity and that the society is genuinely democratic (not controlled by a wealthy elite), then it is not clear that there is an economic 'hierarchy' that is inconsistent with the equal standing of citizens. Setting that issue aside, I concede that the difference principle is not a perfectly general principle. Wiens insists that if the difference principle is to serve as 'basic' evaluative function it must rank all possible worlds, not just a subset (Wiens 2015, 443-4). But lack of complete generality does not make a principle something other than a principle. A principle can still serve an evaluative function even if it is not basic because not perfectly general. Greater generality is of course desirable. But greater generality means greater complexity, and it is unlikely that there is any path to greater generality except via special cases that we then seek to generalize.<sup>4</sup>

There is a problem with unexplained fact-sensitivity, which is that lack of generality implies a degree of ethical inarticulacy. Recall our simple example of manna from heaven in

---

<sup>4</sup> Wiens also argues that the difference principle cannot rank worlds in which capital mobility is high and worlds in which it is low (Wiens 2015, 444). He imagines a given society under both factual scenarios. In both, the position of the worst off is maximized, and at the same level, but in the high-mobility scenario the contribution curve is much flatter, with the result that it takes a lot more inequality to get the worst off up to their maximum level, i.e. the better off are *much* better off, in high-mobility scenario. Wiens says that the difference principle does not rank one world above another. That is true if all the difference principle says is "maximize the position of the worst off." If we think of the principle as ranking gains for the worst off against equality, however, we should prefer the low-mobility world (the one with the steeper contribution curve), on the ground that it requires less inequality to bring the worst off up to the same level.

which the status quo option is [5, 5] and the alternative is [7, 6]. If all we knew was that [7, 6] > [5, 5], and so could not say whether [6.25, 6.25] or other currently infeasible options would be preferable, this would imply that we did not know whether our underlying principle was to maximize the minimum, the average, or the maximum. The same is true for the unexplained factual limitations of the difference principle. Not knowing what the right principle is for the condition in which chain connection fails also involves a lack of ethical articulacy; we do not know how to rank different inequalities, of different sizes, between worst off and next worst off.

I have argued that it can be important to our social relations that we publicly acknowledge what would be just under currently infeasible conditions, and that such acknowledgement involves understanding the principles that justify our political choices. It would be cognitively too demanding, however, to require that citizens acknowledge what would be just under all possible counterfactual conditions, or to expect people to articulate completely general principles. It may also be unnecessary. Although it matters to the relations between the better and worse off that they understand that [7, 6] is preferable to [5, 5] because it raises the minimum, not because it maximizes the average, and that they acknowledge that [6.25, 6.25] would be better still, it may not matter that the better and worse off have explicit views about more distant counterfactual possibilities. From a theoretical point of view greater understanding is always desirable, but the fact that a principle is not fully general does not necessarily make it a phony principle, something less than or other than a principle of justice.

There is however a special case in which a worry about defining justice down does arise. The real problem with fact-sensitivity arises when the limited generality of a principle is explained by feasibility constraints that are due to facts about human nature that one might assess as just or unjust.

### **Feasibility, Stability, and Psychology**

The issue of feasibility arises in connection with the ‘strains of commitment’ and the problem of stability. The parties to the original position must not agree to principles unless they are confident that they would be able to abide by them no matter what social position they end up occupying (Rawls 1999, p. 153). Determining whether this is so involves appealing to the general facts of human psychology. I am setting aside the original position, however, in order to ask how the principles themselves depend on facts. In this respect, the strains of commitment matter because they speak to the relative stability of competing conceptions of justice.

Stability is a property of equilibria. A system is in equilibrium if its current state will persist indefinitely if not disturbed by external forces (Rawls 1999, p. 400). An equilibrium is more or less stable depending on the extent to which it has a tendency to return to equilibrium when disturbed – when, as Rawls says, departures from equilibrium ‘call into play forces within the system’ that bring it back to this state. Many different kinds of systems can be stable or unstable, with respect to different kinds of conditions. The systems that were the focus of Rawls’s concern were the basic structures of political societies well-ordered by different conceptions of justice. A conception of justice is stable if when realized by a social system (and known by its inhabitants to be realized) departures from justice tend to correct themselves.

Rawls’s argument that justice as fairness would be stable (and more stable than utilitarianism) relied on claims about the attitudes and motives people would develop growing up in a society organized according to this conception. When the principles of justice as fairness were realized by the basic structure of society (and known to be so), people would develop an appropriate and effective sense of justice. This psychological conjecture relied on the empirical premise that people are motivated by reciprocity, a backward-looking (i.e. non-strategic)

tendency to answer in kind. Rawls speculated that this trait was a result of natural selection (Rawls 1999, p. 440). Our sense of justice extends this trait; we want to comply with the rules of institutions when we have benefited from others likewise complying (Rawls 1999, pp. 238, 296, 429-33, 440). Rawls claimed that justice as fairness could use this disposition as a stabilizing force more easily than could utilitarianism. If social inequalities benefit all, people will tend to see rule-compliance on the part of others as a benefit that ought to be reciprocated, whereas if inequalities simply maximize aggregate utility, some people may not see rule-compliance as a benefit, and so may be less inclined to reciprocate (Rawls 1999, p.155). Rawls's principles are 'closer to the tendency of evolution' than is the principle of utility (Rawls 1999, p. 440)

Cohen objects that stability is a relevant consideration for rules of regulation, not for fundamental normative principles. If stability were a requirement for fundamental principles of justice to count as principles of justice, then it would be impossible to say obviously sensible things such as 'justice is fragile,' or 'we want justice to last' (Cohen 2008, p. 328). This objection is decisive if principles are meant to define the idea of justice. However, as we have seen, Rawlsian principles are meant to provide rankings of considerations relevant to the justice of social institutions. If institutions designed in accordance with a particular set of rankings would not last, then one might say that these are not the right rankings for the factual circumstances in which we find ourselves. Nonetheless, the requirement of stability does threaten to generate a problematic kind of factual impurity, one that may involve watering justice down.

Rawls initially presents stability as a property that is desirable 'other things equal' (Rawls 1999, p. 6). Similarly, he later says that considerations of stability are 'not decisive' (Rawls 1999, p. 399), and that his argument that justice as fairness would be more stable than utilitarianism was not intended as a justification (Rawls 1999, p. 441). However, in the same

passage, Rawls also speaks of checking to see whether the principles identified by the original position are feasible, i.e. ‘not so unstable that some other choice might be better’ (Rawls 1999, p. 441). This statement implies that stability is not a tie-breaker for conceptions equally well-supported by the original position, but a threshold that all conceptions of justice must pass, a threshold that a conception ranked higher by the original position might fail. Moreover, Rawls is not consistent in presenting stability as a consideration that arises after choice of principles in the original position. As Jon Garthoff points out, the issue of stability arises behind the veil of ignorance, in Section 29 (on ‘some main grounds’ for the two principles), where Rawls says that ‘psychological stability’ is a ‘strong point’ in favour of a conception of justice (Rawls 1999, p. 154; Garthoff 2016). If the parties take stability into account as a ‘strong point,’ they do not appear to be treating stability as a threshold requirement all candidate conceptions must pass, let alone as a mere tie-breaker. Garthoff argues that stability was always for Rawls a ‘greatly significant’ and often ‘decisive’ consideration in assessing conceptions of justice of justice (Garthoff 2016). Even if stability is merely a threshold condition, however, a worry about defining justice down can arise, so long as stability can trump what Garthoff calls ‘independent convictions about fairness,’ or what Rawls referred to as ‘ideals of justice’ (Rawls 1999, p. 138). As we have seen, Rawls claimed that his principles are in fact stable, and more stable than the principle of utility. What if the result had been different, however – perhaps because the facts of human psychology turned out to be different?

To illustrate the worry about defining justice down, suppose that people are by nature more selfish and less strongly motivated by reciprocity than Rawls claims. If evolution has made us more selfish than Rawls assumed, it might be that a society successfully established according to his two principles would not endure (i.e. would not remain just according to that conception). Perhaps the better off simply will not support laws and institutions animated by the difference

principle. It might be that the best we hope for is a society characterized by equal basic liberties, formal equality of opportunity, a social minimum, and beyond that institutions that maximize the average level of social primary goods. In this scenario, Rawls's method apparently calls for us to select less ambitious principles that are more stable, principles more in tune with our limited moral capacities. According to Garthoff, 'Rawls would sooner abandon the difference principle than abandon his appeals to stability considerations' (Garthoff 2016). But would he be right to do so?

The objection to tailoring principles to the feasibility constraints implied by human nature is that it can affect the quality of social relations. Even if people were selfish by nature and not by social construction, it would be important for us to be able to recognize that this fact is regrettable, because our attitudes matter for the relationships we have with one another. If optimal principles are infeasible, it would be important to recognize that our acceptance of second-best but more stable principles is contingent on whatever the unfortunate facts of human nature are that make the first-best principles infeasible. To see this, consider how knowledge of what would happen in counterfactual scenarios can affect the nature of personal relationships. If we are friends, I would not hurt or take advantage of you even if the opportunity presented itself, and mutual recognition of that fact is part of what it means to have a friendship. Even if I have limitations as a friend, however, we might still be friends so long as those limitations are not too serious and we commonly recognize them as limitations. Perhaps I have a bad temper, and am liable to say hurtful things when angry. I may not be able to change this fact, but if I recognize it as unfortunate and so apologize and make amends, we may be able to maintain our friendship. What matters to our relationship is not just what ends up happening, as things turn out, but what would happen in alternate circumstances, and our mutual recognition of the moral status of these

counterfactuals. Blurring the line between stability and justice could lead us to define justice down, by which I mean labeling as a principle of justice a principle that is only optimal given limitations of human psychology that are unfortunate from the point of view of justice itself.

It might be objected that I am wrong to treat stability as a matter of feasibility. Rawls originally spoke of the ‘inherent’ stability of just institutions (Rawls 1999, pp. 192, 193, 436), and later switched to speaking of ‘stability for the right reasons,’ which he said was not merely a matter of avoiding futility (Rawls 1996, p.142; see also Garthoff 2016). Stability is necessarily inherent, because it is a matter of forces internal to a system that bring it back to equilibrium when perturbed. If coercive penal institutions are counted as part of the system, however, even a political order maintained by purely by force could count as stable. Paul Weithman points out that when Rawls describes the stability of a well-ordered society, he talks about it being able ‘to generate its own support’ (Weithman 2010, p.119). ‘Support’ is not compliance based on fear, but some form of endorsement.<sup>5</sup> Weithman therefore stresses the importance of ‘congruence’ to the case for stability (Weithman 2010, pp. 57-67). Rawls’s congruence claim is roughly that in a well-ordered society, where everyone has such a sense of justice, it is a good *for me*, for me to have a sense of justice, considering only my personal good apart from the intrinsic importance of justice. According to the standard account of the evolution of Rawls’s view, the argument for congruence in *Theory* did not take into account the diversity of reasonable conceptions of the good that would exist in a well-ordered society. The problem was then to show that justice as fairness could be the object of an overlapping consensus of comprehensive doctrines. At this

---

5 Weithman contrasts ‘inherent’ with ‘imposed’ stability in order to remind us that enforcement and punishment are not taken to be part of the core system, and thus do not count as internal. The status of enforcement is actually more complicated than Weithman’s initial presentation suggests, because (as we will see in Section 5), Rawls thought that even in a well-ordered society enforcement would still be needed, in order to provide assurance of compliance on the part of others (Rawls 1999: 237).

point, it became misleading to speak of the ‘inherent’ stability of justice as fairness, since this stability would depend on endorsement of the conception based on different comprehensive viewpoints (Weithman 2010, p. 341).

On Weithman’s account, however, stability for the right reasons looks more like a value or a moral principle than a feasibility constraint. Stability for the right reasons excludes ‘indoctrination and internalization of arbitrary authority.’ Instead, stability for the right reasons is based on ‘transparent’ moral education. ‘The justice of the WOS [well-ordered society] is therefore sustained by the autonomous activity of its members.’ Because each of us affirms our sense of justice from within our own comprehensive doctrine, this affirmation is ‘willing and transparent.’ By showing that a liberal political conception of justice can be stable for the right reasons Rawls has demonstrated that this conception can ‘transform our final ends’ so that we can affirm our sense of justice ‘without violating our freedom’ (Weithman 2010, p. 342).

If this is the right way of understanding stability, then it may not be true that blurring the line between stability and justice risks defining justice down, because stability in this sense is not a feasibility constraint. When stability is understood as requiring voluntary, transparent endorsability of a conception of justice based on diverse but reasonable comprehensive doctrines, then stability seems to overlap with if not reduce to the concern with public justifiability that came to the fore in Rawls’s later work. The question at this point would be whether it is right that true justice (or what I take to be true justice) should be subordinate to our sharing a conception of justice that is acceptable to (or based on reasons acceptable to), all reasonable religious and philosophical doctrines. The concern with public justifiability can also seem to involve watering justice down – not for the sake of feasibility but for the sake of a competing value (e.g. civic friendship). I do not want to take a position on the correct

interpretation of Rawlsian stability, apart from insisting that the idea was always at least in part about feasibility. If Rawlsian stability also involves distinctive values or moral concerns these should be clearly distinguished from the concern with feasibility. For to whatever extent the concern about stability is a matter of feasibility, the worry about defining justice down will arise.

### **Reciprocity, Justice, and Human Nature**

I have argued that it is important to recognize feasibility constraints rather than building them into principles of justice. This claim raises the question of whether justice itself is a second-best virtue, as Michael Sandel claimed (Sandel 1998, pp.32), following Hume.<sup>6</sup> If we refuse the designation ‘fundamental principle of justice’ to a principle that is only best given feasibility constraints, should we refuse to call principles of justice fundamental moral principles, since we only need principles of justice because of human selfishness? It is a mistake, however, to think that limited benevolence is an unfortunate feature of human nature that limits the extent to which higher virtues are feasible, making justice a ‘remedial’ virtue. Consider what human beings would be like if limited benevolence did not obtain. We would have impartial concern for the well-being of all, with no special concern for our own well-being over that of others. As Bernard Williams suggested, however, this scenario is not a moral ideal, and perhaps not even coherent (Smart and Williams 1973, pp. 108-18). If I care for everyone’s well-being equally I do not have any special projects, purposes or relationships of my own; I, and everyone else, care only about benefitting people who do not care about their own well-being, except in so far as it is someone’s well-being. Why should I care about your well-being at all, if you do not have any special concern for it? Why do we all care about everyone’s well-being when none of us cares about any one person’s well-being specifically? What would well-being consist in if no one had any

---

6 See *A Treatise of Human Nature* Book III, Part II, Chapter 2, ‘Of the origin of justice and property;’ (Hume 1826a: 253-73) as well as *An Enquiry Concerning the Principles of Morals* Chapter III ‘Of Justice’ (Hume 1826b: 253-78).

projects except for this generic project of increasing aggregate well-being?

These familiar objections to strict forms of utilitarianism make the point that the fact that human beings are not perfectly, impartially benevolent is not a flaw of our nature that makes higher values infeasible, forcing us to settle for mere justice. Limited benevolence is a background condition for the intelligibility of ideas such as well-being, happiness, love, friendship, and community, as well as justice. Limited benevolence is not a feature of human psychology that limits realization of justice value but a condition that makes justice relevant or applicable. That Rawls set up the original position to take into account the facts involved in the circumstances of justice is not a problem, for these are not facts that limit the realization of justice but conditions in which the concept applies – facts about justice, rather than facts about the world that constrain the achievement of justice.<sup>7</sup> In short, whether or not principles should take into account a particular fact about human nature depends on whether it is simply a fact of psychology that limits how much justice is currently feasible, or a fact about justice, and the concept's conditions of application – what we might call a constitutive condition.

In the previous section, we considered the possibility that people were more selfish and less motivated by reciprocity than Rawls thought, with the result that his principles would not be stable, forcing us to revise Rawls's principles of justice downwards in order to achieve stability. This case was hypothetical, because the evidence from experimental economics and psychology suggests that people are in fact motivated by reciprocity (Fehr et al. 2002). The concern about watering justice down to fit human nature can still be raised, however, given that motivation by

---

7 This analysis goes in the same direction as David Estlund's argument that facts about human nature do not limit the content of justice. Justice only applies where there is limited altruism, but the content of justice is what it is regardless of how altruistic people are; (Estlund 2011: 225-29)

reciprocity in some ways limits what justice demands of us. Whether Rawlsian principles are watered down will turn out to depend on whether we classify motivation by reciprocity as a constraint to which conceptions of justice must accommodate themselves (i.e. a limiting condition), or as one of the so-called circumstances of justice, the conditions under which principles of justice apply (a constitutive condition).

I explained above how the fact of motivation by reciprocity helps support the claim that justice as fairness is more stable than utilitarianism. Yet this fact also generates an assurance problem. Rawls introduced the issue of assurance as part of a discussion of mechanisms for providing public goods. He denied that the need for enforcement derived solely from ‘men’s propensity to self-interest and injustice’ (Rawls 1999, p. 237). Although just persons would not cheat on a rule requiring a given level contribution, they would still want assurance that others would do their part. ‘[E]ven with a sense of justice men’s compliance with a cooperative venture is predicated on the belief that others will do their part’ (Rawls 1999, pp. 296, 238). Whereas on a small scale people might know each other well enough to have this mutual confidence, in a populous society people will be unsure about the motives of their fellow citizens. Sanctions are therefore necessary to provide assurance about the conduct of others, even where everyone is *in fact* motivated by reciprocity (and so would not cheat even if they could get away with it). I will comply even if I know I could get away with shirking, but only if most others are complying. I am not a cheat, but I am not a sucker either.

This attitude may seem perfectly reasonable. However, motivation by reciprocity can have a darker side as well. Reciprocity can involve repayment of harms as well as benefits; tit for tat can lead to cycles of revenge, not just stable cooperation. Suppose that one thought that in an ideally just society the justice system would have no retributive element. It might be that such a society would not be stably just, given the fact of motivation by reciprocity, forcing us to modify

our principles of justice. In this case, insisting that principles of justice be stable given the psychology of reciprocity might affect the content of our principles, leading to a dilution of justice that we could not recognize as such.

Setting aside retribution, there is still a question about the moral status of negative reciprocity, the unwillingness to produce benefits for non-reciprocators. As Stuart White points out, one might think that people ought to be more unconditionally altruistic (White 2003, 65), in which case one would view motivation by reciprocity as an unfortunate limitation of (contemporary) human character. While some duties are no doubt conditional on reciprocity, others are presumably not, leading to an important question about where we draw the line. I have a duty not to kill you except in self-defence even if you would kill me if given the chance. Similarly, even if it is permissible to torture someone to avert a moral disaster, it will only be permissible to torture a torturer in order to avert a moral disaster, not for just any minor benefit; the fact that the torturer *would* torture us, were our situations reversed, does not undercut our duty not to torture. Other duties are plausibly seen as being conditional on reciprocity. For example, one might think that I have a duty to share fairly with others only so long as I can reasonably expect that they would share fairly with me, were our situations reversed. There is room for reasonable disagreement about where to draw the line between the two kinds of duty. In particular, some people will think that the duties associated with distributive justice have a more unilateral character, and that the Rawlsian concern for consistency with the psychology of reciprocity unduly limits the requirements of justice. Demanding stability given motivation by reciprocity restricts the duties individuals can have beyond the reach of coercive institutions, as well as within their interstices, where assurance of compliance on the part of others is lacking.

For example, Cohen maintains that if individuals were motivated by a properly

egalitarian ethos, they would not need such big incentives in order to work in ways that aid the least well off.<sup>8</sup> The correct choices are not mandated by law, however, nor can they be very clearly identified by social norms (where to work, how much to work, how to conduct negotiations over one's salaries and benefits, etc.). Even if with full information it would be possible to determine the demands of social justice in relation to these myriad situations, other people will not generally have enough information to assure themselves about the conduct of others. Therefore everyone will be faced with doubts; am I sacrificing my personal goals for the sake of others who are not willing to do likewise for me (Williams 1998)?

A number of people have pointed out it is misleading to speak of Cohen's ethos as being egalitarian, if by 'egalitarian' one means 'ascribing intrinsic value to equality' in a given dimension (Quong 2010, pp. 327-28; Tomlin 2010, pp. 237-40). My refusal to pursue a lucrative medical career in order to be a starving artist does not generate economic inequality. It fails to promote the well-being of others, but a Paretian ethos cannot be a requirement of *justice*, as Cohen has defined the term, since justice consists in the absence of arbitrary inequalities. Cohen's ethos can still be described as egalitarian in the broader sense of the term that he described in the opening pages of his Tanner Lectures. In order to count as an egalitarian in this sense one need not ascribe intrinsic value to equality, one need only believe that the better off have duties of justice to aid the worse off, and that these duties are more demanding the greater the discrepancy between them, and the worse off the worse off are in absolute terms (Cohen 1992, p. 267). For my purposes, it does not matter how we label Cohen's ethos. The criticism I

---

8 I do not want to exaggerate the differences between Rawls and Cohen on the ethical as opposed to institutional demands of justice. On the one hand, Cohen recognizes an individual prerogative to pursue personal goals at the expense of the goal of benefitting the worst off (Cohen 2008, pp. 10-11). On the other hand, Rawls recognized individual ethical duties that would not be enforced by law, including the duty of mutual respect (Rawls 1999, pp. 56, 297, 470). Still, the demands that Rawls's principles place on us are limited where assurance of compliance on the part of others is lacking.

am considering is that via the requirement of stability Rawls ends up underestimating what justice requires of us. Whether the resulting conception shows insufficient concern for equality or for the absolute position of the worst off, the claim is that Rawls's method defines justice down.

The need for assurance of reciprocity is also one of the reasons for the lack of stronger duties of global distributive justice, in Rawls's scheme. Outside of an institutional scheme, the duty of justice includes no duty to share wealth or resources fairly, but only the duty 'to further just arrangements not yet established, at least when this can be done without too much cost to ourselves' (Rawls 1999, pp. 99, 257-8). We have a limited duty to help establish just global arrangements, but no duty to share the benefits of cooperation fairly in the absence of institutional arrangements sufficient to ensure that others will share fairly with us (or would, were our economic positions reversed). If one thinks that justice places greater unilateral demands on us than Rawls is willing to countenance, one may be inclined to see the institutional focus of Rawls's theory of justice as a limitation dictated by considerations of stability. Rawls will have been guilty of defining justice down, in this case, not in the principles of institutional assessment he identifies, but in the limitation of these principles to the design of institutions, to the exclusion of individual conduct within or beyond these institutions.

In order to meet this objection, we need to anchor the reciprocity condition in justice itself, rather than making it simply a fact about human psychology. One way to do this is to say that social justice is at least in part a relational value, in the sense that its point is to affect the nature of our social relationships. We saw earlier that Cohen distinguished fundamental moral principles from rules of regulation. The latter are rules justified by the consequences of their general observance; the former are criteria by which we assess such consequences. There is an

intermediate possibility, however, which is that we adopt a principle to regulate our conduct not in view of the causal consequences this adoption will have, over time, but in view of its constitutive effect on the nature of the relationship we have with each other. In this case, the outcome the regulation seeks to advance is not brought about by a causal process; rather, the regulation of our conduct by the principle is itself the valuable outcome, because this regulation puts us into a particular relationship with each other. The reason we should not want to profit from good fortune in the genetic or social lotteries (except in ways that benefit all), is that when made manifest and reciprocated, this motivation affects the character of the relationship between citizens. From within the original position, expression of mutual respect in institutions can only have instrumental significance (if we assume that self-respect depends on being respected by others). More fundamentally, however, distributive justice permits us to live on terms of mutual respect with those engaged in social cooperation. If the point of distributive justice is that it gives our relationships the right character, then it is natural to think that our commitment to distributive justice will be attenuated where assurance of compliance on the part of others is lacking. To continue to share fairly with you when you are refuse to do likewise with me would be to treat myself simply as means for satisfying your needs. As White argues, ‘scaling down one’s efforts on the non-reciprocator’s behalf can then be understood as an assertion of one’s own dignity (“I am not a doormat”),’ and so indirectly an affirmation of the dignity of others’, not simply a ‘brute fact’ of psychology (White 2003, 65-6).

The situation is more complicated in multi-person settings. Is A required to fulfill a duty with respect to B and C, if B but not C will do likewise, and the only options are to fulfill the duty with respect to everyone or noone? This can be the situation with respect to contribution to public goods. The solution is of course enforcement. However, the duty to vote for and otherwise support just political decisions cannot be enforced. So long as the decisions made will be

complied with to the requisite level, it is presumably not too much to ask citizens to aim for justice in their political decision-making, even if others are not doing likewise. I'm not going to pay my taxes if no one else is, but if I can count on enough others paying their taxes, I have a duty to support just tax law even if the political conduct of others is purely self-interested. These matters obviously require further discussion.<sup>9</sup> The point I want to insist on is that if the purpose of the duties associated with distributive justice is to create or constitute a relationship of mutual respect, unilateral compliance will not be successful (unless it elicits compliance, in which case it is not unilateral). This way of thinking grounds the concern for reciprocity in facts about justice (that social or distributive justice is a relational concern) rather than in facts about human psychology that generate feasibility constraints.

## **Conclusion**

Why should we ever be interested in questions about what would be just under factual conditions that do not presently obtain? Why should not we focus solely on what is just here and now? Various answers are possible. We might make faster intellectual progress by figuring out what is just in simple unrealistic cases before turning to more complex realistic cases. We might need an account of what justice requires when everyone is acting justly in order to answer the question of what is the just way to respond to injustice. We might need to have a long-term goal of justice in mind, to avoid getting stuck at local improvements that block further progress. In this paper, I have set out a fourth possible justification for ideal normative theory. Beliefs about

---

9 Moreover, when institutions are needed but not yet in place, I have a natural duty to help create them, a duty that is not conditional on others reciprocating, but which is limited in the extent of costs it can impose on me. It is for this reason that I say that our duties of distributive justice are only 'attenuated' and not eliminated when assurance of compliance on the part of others is lacking. I discuss the relationship between reciprocity and justice at more length in Lister 2013.

what would be just under conditions that do not obtain can affect the social relations that do obtain here and now. From this point of view, Rawls's theory could be criticized for being too realistic, rather than for being too ideal. Fact-sensitivity limits the generality of normative principles without necessarily making them something less or other than normative principles, I claimed. Yet the fact-sensitivity of Rawls's principles includes the requirement that they be stable given motivation by reciprocity. If principles depend on facts of psychology that are unfortunate from the point of view of justice, it would be important for us to recognize the traits in question as regrettable, even if they cannot (presently) be changed. Revising principles downwards in order to achieve stability threatens to make such recognition impossible. However, I also claimed that reciprocity can be construed as an inherent aspect of justice, rather than a de facto limit that would make more demanding principles infeasible. Whether or not one believes that the principles of justice as fairness have been wrongly watered down due to considerations of stability depends on whether one views motivation by reciprocity as (i) a fact of human nature limiting how much justice is feasible, or (ii) a feature of justice itself, which is a relational value, at least in part, and which therefore includes a reciprocity condition.

## References

- Arneson, Richard. 2008. Justice is Not Equality. *Ratio* 21 371-91.
- Cohen, G. A. 1992. *Incentives, inequality, and community*. Vol. 13, *Tanner lectures on human values*. Salt Lake City, Utah: University of Utah Press.
- Cohen, G. A. 2003. Facts and principles. *Philosophy and Public Affairs* 31 211-45.
- Cohen, G. A. 2008. *Rescuing justice and equality*. Cambridge, Mass.: Harvard University Press.
- Corak, Miles. 2013. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives* 27 (3): 79-102.

- Estlund, David. 2011. Human nature and the limits (if any) of political philosophy. *Philosophy and Public Affairs* 39 (3): 207-37.
- Estlund, David M. 1998. Liberalism, equality, and fraternity in cohen's critique of rawls. *The Journal of Political Philosophy* 6 (1): 99-112.
- Farrelly, Colin. 2007. Justice in ideal theory: a refutation. *Political Studies* 55 844-64.
- Fehr, Ernst, Urs Fischbacher, and Simon Gächter. 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature* 13 1-25.
- Garthoff, Jon. 2016. Rawlsian stability. *Res Publica* forthcoming.
- Hume, David. 1826a. *A treatise of human nature*. Vol. 2, *The philosophical works of David Hume*. London: Adam Black and William Tait and Charles Tait.
- Hume, David. 1826b. *An inquiry concerning the principles of morals*. Vol. 4, *The philosophical works of david hume*. London: Adam Black and William Tait and Charles Tait.
- Lister, Andrew. 2013. Reciprocity, relationships, and distributive justice. *Social Theory and Practice*, 39 (1): 70-94.
- Mills, Charles W. 2005. 'Ideal theory' as ideology. *Hypatia* 20 165-84.
- Mills, Charles W. 2009. Rawls on race/race in Rawls. *Southern Journal of Philosophy* 47: 161-84.
- Pogge, Thomas. 2008. Cohen to the rescue! *Ratio* 21: 454-75.
- Quong, Jonathan. 2010. Justice beyond equality. *Social Theory and Practice* 36 (2): 315-40.
- Rawls, John. 1996. *Political liberalism*. New York: Columbia University Press.
- Rawls, John. 1999. *A theory of justice*. Cambridge, Mass.: Harvard Univeristy Press. Original edition, 1971.
- Sandel, Michael. 1998. *Liberalism and the limits of justice, 2nd Edition*. Cambridge: Cambridge

University Press.

Smart, J. J. C., and Bernard Williams. 1973. *Utilitarianism: for and against*. Cambridge:

Cambridge University Press.

Stemplowska, Zofia, and Adam Swift. 2012. Ideal and Nonideal Theory. In *The Oxford*

*handbook of political philosophy*, ed. David Estlund, 373-92. Oxford: Oxford University

Press.

Tomlin, Patrick. 2010. Survey article: internal doubts about Cohen's rescue of justice. *The*

*Journal of Political Philosophy* 18 (2): 228-47.

Valentini, Laura. 2009. On the apparent paradox of ideal theory. *The Journal of Political*

*Philosophy*, 17 (3): 332-55.

Valentini, Laura. 2012. Ideal vs. non-ideal theory: a conceptual map. *Philosophy Compass*, 7/9:

654-664.

Weithman, Paul. 2010. *Why political liberalism? On John Rawls's political turn*. Oxford: Oxford

University Press.

White, Stuart. 2003. *The civic minimum: On the rights and obligations of economic citizenship*.

Oxford: Oxford University Press.

Wiens, David. 2015. Against ideal guidance. *The Journal of Politics* 77:433-446.

Williams, Andrew. 1998. Incentives, inequality, and publicity. *Philosophy and Public Affairs* 27:

225-47.