

Uwe Steinhoff

## The Uselessness of Rawls's "Ideal Theory"

Over the years a few authors have argued that Rawls's ideal theory of justice is useless for the real world. This criticism has been largely ignored by Rawlsians, but in the light of a recent accumulation of such criticisms, some authors (in particular Holly Lawford-Smith, A. John Simmons, Zofia Stemplowska and Laura Valentini) have tried to defend ideal theory. In this article I will recapitulate the precise problem with Rawls's ideal theory, argue that some of Rawls's defenders misconceive it, and show that recent attempts to rescue Rawls's ideal theory from the charge of being useless fail. While there are useful kinds of ideal theory, Rawls's is not one of them. In addition, Rawls's very tentative suggestions for some kind of bridge between ideal and non-ideal theory are contradictory insofar as they implicitly presuppose the non-existence of the problem they are meant to solve. Thus, Rawls's "non-ideal theory," too, is useless, and not so much a theory at all but a set of *ad hoc* stipulations. Finally, I will show that certain attempts within the global justice debate to use some variation of the original position to *directly* derive guidelines for the real world are misguided and yield no useful results.

## I. The Problem with Rawls' "Ideal Theory"

Rawls limits<sup>1</sup> his "ideal theory" of justice to "the basic structure of society conceived for the time being as a closed system isolated from other societies,"<sup>2</sup> and also conceived as a society in which *everybody* acts justly and does his part in upholding just institutions.<sup>3</sup> That is, he introduces (at least<sup>4</sup>) two "idealizations," namely the closedness condition and the strict compliance condition, and limits his "ideal theory" of justice to societies where these two conditions are fulfilled. In still other words, Rawls's "ideal theory" of justice tells us *this*:

RITJ: Under the conditions of closedness C and of strict compliance SC with the principles of justice that parties to the original position would choose, the basic structure of a society is just if and only if it fulfills the two Rawlsian principles of justice RPJ.

Thus, we can define an "ideal theory" as a theory that tells us what *would* be the case if

---

<sup>1</sup> Or better: he claims to limit it in this way. In fact, however, he also talks about natural individual duties, not only about the basic structure.

<sup>2</sup> John Rawls, *A Theory of Justice* (Cambridge, MA: The Belknap Press of Harvard University Press, 1999), p. 7.

<sup>3</sup> *Ibid.*, e.g. p. 8 and 308-9.

<sup>4</sup> It also seems that Rawls at least implicitly conceives of the society in question as one containing only people capable of co-operation (at least at some point in their lives). However, there is no need to go into this here. I will stick to his two most famous idealizations.

certain *non-realized conditions* were realized. Obviously, for rather straightforward logical reasons (the antecedent of the implication is wrong), RITJ, like *any other* “ideal theory” of justice and morality, be it Rawlsian or not, does not by itself tell us what would be just or how we ought to behave in the *real* world, that is, in our world, where, as it happens, C and SC do not hold.<sup>5</sup> (Note that I do *not* claim that it implies by itself *nothing* about the real world. I do not need to make such a claim, though, and I will return to the issue below.)

This inapplicability argument cannot be overcome by claiming “that it relies on the premise that an ideal theory contains *only* criteria of (ideal) justice” and that “a moral theory such as a theory of justice contains more than the slogan that summarizes its conclusion,” for example, “a theory of value.”<sup>6</sup> Such remarks miss the point of the inapplicability charge, since there is no need for such a premise. An ideal theory of justice, as defined here, has the form of a conditional sentence (like RITJ), and if “a theory of value” appears on the side of the antecedent and is thus “contained” within the

---

<sup>5</sup> Annette Baier, “Theory and Reflective Practices,” in her *Postures of the Mind: Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985), pp. 207-27, at 210, correctly notes: “The real moral guidance comes when we are told not how the law of an ideal community would apply to a concrete case in that ideal world, but when we are told [what] we should therefore do, in this nonideal world.” Simmons and other defenders of Rawls quote from Baier’s article, but it seems that they do not quite appreciate the logical force of her point.

<sup>6</sup> Eva Erman and Niklas Möller, “Three Failed Charges Against Ideal Theory,” *Social Theory and Practice* 39(1) (2013), pp. 19-44, at 30. Erman and Möller are also mistaken in categorizing the inapplicability problem as “epistemological” (*ibid.*, pp. 27 and 31); rather, it is a *logical* problem.

ideal theory, it has not yet been *asserted*; and if it appears on the side of the consequence and is thus “contained” within the ideal theory, it has still not yet been asserted: all that has (at best) been asserted in that case is that this theory *would* be correct *if* the antecedent were correct (but in ideal theory, of course, the antecedent is *not* correct). Thus, the *structure* and the idealized antecedent of an ideal theory make sure that *whatever* theory of value is “contained” within this ideal theory has no normative force as long as it does not escape this containment. *That* is the charge.

Of course, Rawls *book* entitled *A Theory of Justice* contains more than the ideal theory, but pointing out this obvious fact certainly does not save the *ideal theory* that the book contains from the charge in question. The charge, again, is not that a *conjunction* of an ideal-theory-conditional with *something else*, for example with a simple normative assertion, cannot contain something that is directly applicable to the real world. The conjunctions “RITJ and you should not torture babies for fun” or “If there are no planets, then fascism is just, and you should not torture babies for fun,” for instance, do contain an applicable normative statement, and the former conjunction is not superior to the latter. However, the applicable normative statement that the conjunctions contain, namely that you should not torture babies for fun, is quite able to stand on its own. Conjoining applicable and true normative assertions with ideal theories helps neither the latter nor the former.

Thus, the charge that RITJ does not by itself tell us what would be just or how we ought to behave in the real world stands. Consequently, the two claims that in a certain *real* society a basic structure that *fulfils* RPJ is *unjust* and that in a certain other real society a basic structure that does *not* fulfill RPJ is *just* are entirely compatible with RITJ, with Rawls’s “ideal theory” of justice. This means that Rawls’s bold claim (which, as it seems, is endorsed by Simmons and many other defenders of Rawls) that his “two

principles of justice provide an Archimedean point for appraising *existing* institutions”<sup>7</sup> does not make too much sense. Making existing institutions conform *more* to Rawls’s principles might actually make them *less* just. (I will have to say more about this latter point shortly.)

To be sure, at times Rawls recognizes that there might be a severe problem with his “ideal theory”:<sup>8</sup>

We must ascertain how the ideal conception of justice applies, if indeed it applies at all, to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice. The discussion of these problems belongs to the partial compliance part of nonideal theory. It includes, among other things, the theory of punishment and compensatory justice, just war and conscientious objection, civil disobedience and militant resistance. These are among the central issues of political life, yet so far the conception of justice as fairness does not directly apply to them.<sup>9</sup>

Unfortunately, Rawls’s does not at all “ascertain” that his theory applies to the real

---

<sup>7</sup> Rawls, *A Theory of Justice*, p. 456, my emphasis. Compare also pp. 215, 216 and 231.

<sup>8</sup> I use quotation marks because Rawls’s “ideal theory” is anything but ideal. Incidentally, Jürgen Habermas and Karl-Otto Apel have the same problem as Rawls. While they and some of their followers, unlike Rawls, dedicate lengthy discussions to it, they certainly do not solve it. On this point (for Habermas) see Uwe Steinhoff, *The Philosophy of Jürgen Habermas: A Critical Introduction* (Oxford: Oxford University Press, 2010), Ch. 2.5, and (for Apel) *Kritik der kommunikativen Rationalität: Eine Darstellung und Kritik der kommunikationstheoretischen Philosophie von Jürgen Habermas und Karl-Otto Apel* (Paderborn: Mentis, 2006), Ch. 3.6.2.

<sup>9</sup> Rawls, *A Theory of Justice*, p. 309, see also 215-6.

world, whether “directly” or indirectly – he just *assures* the reader that it does. Those assurances, however, remain entirely unsubstantiated.<sup>10</sup>

But can’t “ideal theory” be *made* to tell us what would be just or how we ought to behave in the *real* world? Well, while it is *logically impossible* for Rawls’s or any other “ideal theory” to tell us such things by itself, it can of course do so in conjunction with certain further assumptions. However, first, it has to be noted that those assumptions necessarily *cannot* be part of “ideal theory” anymore; and second, the question arises as to how plausible those additional assumptions are and how they can be *justified*.

An additional assumption that makes “ideal theory” relevant for the real world in the respect at issue can be called a *bridging-principle*. Here is one:

BPLD: If RITJ is correct, that is, if, under the conditions of closedness C and of strict compliance SC with the principles of justice that parties to the original position would choose, the basic structure of a society is just if and only if it fulfills the two Rawlsian principles of justice RPJ, then any society in which one of these conditions is not fulfilled is just if and only if it is a liberal democracy.

If Rawls’s theory of justice is correct, then this is not an ideal principle, since *ex hypothesi* its antecedent (namely RITJ) is correct. While, however, this bridging principle makes RITJ *indirectly applicable* to the real world, it does not succeed in making it *useful*. For it is “natural to assume” – to borrow one of the expressions Rawls uses when assuring us of the relevance of “ideal theory” – that for any justification of BPLD there would be a *much simpler* justification for the justice of liberal democracies that would explain their justice under real circumstances in a *straightforward* manner, that is,

---

<sup>10</sup> A. John Simmons, “Ideal and Nonideal Theory,” *Philosophy and Public Affairs* 38(1) (2010), pp. 6-36, e. g. at pp. 10 and 12, quotes several of Rawls’s assurances to this effect. It probably goes without saying that quoting them is not the same as showing that they are correct.

without taking the awkward detour through “ideal theory.” What, then, do we need “ideal theory” for?<sup>11</sup> (And a *justification* of BPLD would be direly needed; after all, there is an infinite supply of other possible bridging principles that could compete with BPLD, for example BPBA, “BA” for “benevolent absolutism.” )

The situation does not change if one develops a bridging principle by elevating the *idealizations* contained in “ideal theory” to a *political ideal*, into a utopia one should strive for. This is what Rawls sometimes suggests, and it is definitely Simmons’ strategy. He states:

To this ... complaint about counterfactual assumptions and “significant falsehoods,” the appropriate Rawlsian response seems simple and straightforward. Rawls’s ultimate target—the ideal of his ideal theory—is the most just institutional structure that can be achieved within the constraints set by the more or less intractable aspects of human nature and the nature of the world. Fleshing out and arguing for such an ideal requires the use of counterfactual assumptions ...<sup>12</sup> So while having identified our ultimate objectives is hardly “of little use” in addressing real societal problems of injustice, it is perfectly fair (as we have also seen) to observe that practical political prescriptions cannot be expected to be easily

---

<sup>11</sup> For a fine discussion of the failure of methods of “bridging the gap between a set of political principles and morality appropriate to ideal conditions and a set of political principles and a morality appropriate to the imperfect circumstances within which we live out our lives,” see Michael Phillips’ seminal article, “Reflections on the Transition from Ideal to Non-Ideal Theory,” *Nous* 19(4) (1985), pp. 551-570, esp. pp. 561-66. Curiously, Simmons nowhere mentions Phillips (nor, to my knowledge, does Rawls).

<sup>12</sup> Simmons, “Ideal and Nonideal Theory,” p. 31.

extracted from ideal theory.<sup>13</sup>

Thus, the bridging principle implied here is this (UTC stands for “Utopia to come”):

BPUTC: If RITJ is correct, that is, if, under the conditions of closedness C and of strict compliance SC with the principles of justice that parties to the original position would choose, the basic structure of a society is just if and only if it fulfills the two Rawlsian principles of justice RPJ, then we should strive for a society that satisfies conditions C and SC.

The problem is that neither Rawls nor Simmons provide any justification for this principle, and that Simmons overlooks a few things in his “simple and straightforward” response.<sup>14</sup> First, he paints a somewhat rosy picture of the ideal of Rawlsian “ideal theory.”<sup>15</sup> Second, he also ignores the fact that Rawls offers us at least two “ideals”: one on the level of domestic theory (*The Theory of Justice*) and another on the level of international theory (*The Law of Peoples*).

Let me explain. Regarding the first point, it should be noted that the first ideal depicts a *closed* society – a detail Simmons does not mention – in which everyone abides by the

---

<sup>13</sup> Ibid., p. 32.

<sup>14</sup> Simmons, “Ideal and Nonideal Theory,” p. 6, claims, incidentally, that “those who criticize the Rawlsian approach [to ‘ideal’ and ‘non-ideal theory’] (in the remarkably thin literature that actually addresses the distinction) seem mostly to dismiss it as flawed in relatively simpleminded ways.” Leaving aside the fact that the Rawlsians’ failure to address this issue reflects rather badly on Rawlsians and “ideal theorists,” not on the critics, my argument here is of course that it is Simmons’ (and other authors’) attempt to show that Rawlsian “ideal theory” is relevant for the real world that is relatively “simpleminded.”

<sup>15</sup> Ibid., p. 32, the passage starting with “Ideal theorizing says ...”



Rawlsian principles of justice. Yet how anyone, in our times, can in all seriousness offer a *closed* society as political ideal (the German Democratic Republic was comparatively closed, thanks to the Berlin Wall, and so is North Korea) is difficult to understand. A closed society, irrespective of its other merits or demerits, certainly does not qualify as “the most just institutional structure that can be achieved.”

The second ideal (which contradicts the first one, so that Simmons and Rawls should at least clarify *which* ideal it is that we should strive for) describes an open international society, which comprises so-called “liberal” and so-called “decent” national societies (none of them closed). The decent ones express their decency among other things by treating part of their citizenry as second-class citizens, denying them an equal standing. The liberal societies, on the other hand, show their liberality among other things by their disrespect towards the second-class citizens in the “decent” societies – they show this disrespect by deeming an international society *just* that comprises such “decent” national societies and hence the corresponding maltreatment of certain innocent human beings.<sup>16</sup> Is this really supposed to be “the most just institutional structure that can be achieved” on an international level?

I have come across the objection here that neither Simmons nor Rawls suggest elevating a closed society to an ideal, and that they therefore do *not* need to justify BPUTC. The first part of this charge is true – but misses the point – the second is not. To wit, my point here is that if you defend the idealizations in Rawls’s “ideal theory” by claiming that they (at least partly) constitute an ideal you should strive for, then you cannot just let one of those idealizations fall under the table only because it is convenient. That would simply be cheating.<sup>17</sup> And therefore it is indeed quite appropriate to point out

---

<sup>16</sup> Rawls, of course, develops his international utopia in *The Law of Peoples* (Cambridge, MA: Harvard University Press, 2001), esp. pp. 11-58.

<sup>17</sup> I have also come across the objection that closedness is a “methodological” idealization

that Simmons conspicuously fails to mention the closedness condition when he describes the “ideal.” However, once the closedness condition comes into the picture again – where it belongs – BPUTC *is* the principle you would need to justify.

But, for the sake of argument: assume we just drop (without any justification, of course) the closedness condition anyway. Would that give us a worthy ideal? No, it would not, for the problem Rawls himself mentioned would just return in a different form: why should a society that fully complies with principles of justice that have been designed for *closed* societies be a worthy ideal in an *open* world?

The answer is, of course, that it is *not* a worthy ideal. To see this, let us take a more critical look at Rawls’s formulation of the two principles of justice. Rawls formulates those principles as referring to *persons*. But *why*? (Characteristically, this question has neither occurred to Rawls nor to Rawlsians.) Given that the parties to the original position labor under the full-compliance and closedness assumptions<sup>18</sup> and are, moreover, not guided by *moral* concerns, but instead by their own rational self-interest,<sup>19</sup> and given further that they know that they are to decide on the principles of justice for *their own*

---

while full compliance is not. The fact of the matter, however, is that, first, it is irrelevant for my argument what kind of idealization the two idealizations are, and second, Rawls nowhere says that the two idealizations are different in kind.

<sup>18</sup> I readily admit, however, that dropping the closedness assumption in the original position would not help much either. See Uwe Steinhoff, “Unsavoury Implications of *A Theory of Justice* and *The Law of Peoples*: The Denial of Human Rights and the Justification of Slavery,” *The Philosophical Forum* 43(2) (2012), pp. 175-196, esp. at 181-183.

<sup>19</sup> Rawls (1999), p. 12.

“association” or society,<sup>20</sup> that is, given that they know that they are insiders, that they are members of their society – given all this, there is absolutely no reason for them to grant *outsiders* any moral standing whatsoever.<sup>21</sup> To do so would simply be irrational.<sup>22</sup>

Thus, the principles that the parties to the original position would actually choose would have to be formulated, for example, as follows (the important changes are in italics):

(1) Each *citizen* of a society is to have an equal right *against his society* to the most extensive total system of equal basic liberties compatible with a similar system of liberty *for all members* of the society.

(2) Social and economic inequalities *within a society* are to be arranged so that they are both:

(a) to the greatest benefit of the least advantaged *members* within the society, consistent with the just savings principle, and

(b) attached to offices and positions open to all *members* of the society under conditions of fair equality of opportunity for all *members* of the society.

Of course, the principles of “the law of peoples” would also have to be reformulated accordingly, namely in such a way that they give only liberal and decent states any rights. After all, for liberal and decent peoples in the international original position it would

---

<sup>20</sup> Ibid., p. 10, 102, 119.

<sup>21</sup> Thus, Rawls’s conception amounts in effect to what Allen Buchanan calls and criticizes as “justice as self-interested reciprocity” in his “Justice as Reciprocity versus Subject-Centered Justice,” *Philosophy and Public Affairs* 19(3) (1990), pp. 227-252, at 229. In the past, Buchanan has denied this (ibid., p. 230, n. 6), but meanwhile he seems to agree (personal communication).

<sup>22</sup> I have argued for this at length in “Unsavory Implications.”

likewise be irrational to give other peoples (let alone “outlaw states”) any moral standing.<sup>23</sup> But this has the consequence that the “ideal” of a society (or of a world-society) that is *fully compliant* with the principles of justice that would *actually* be chosen in the original position is the ideal of a society (or world-society) that is completely *compatible* with the members’ of one society enslaving or exterminating the members of another society.<sup>24</sup> That is hardly a worthy “ideal.”

(At this point, incidentally, I have met with the further objection that criticizing the content of Rawls’s idealizations seems to be irrelevant if I want to show that Rawls’s theory is inapplicable. But as I already said, I do not want to show that. I want to show that it is *useless*. I first claimed that his theory is by itself not *directly* applicable to the real world. This means that by itself *it provides no practical guidance whatsoever*. This is the first sense of “useless.” Second, the indirect application of the theory, as we will see below, does not provide us with anything that other theories could not also provide, and that much more efficiently. Thus, the theory is *superfluous*. This is the second sense of “useless.” Third, if we combine the theory with a bridging principle like BPUTC – or even with a bridging principle that omits the closedness condition – then this combination actually *is* directly applicable. However, it has utterly absurd and thus *clearly unacceptable implications*. This is the third sense of “useless.” To show, however, that the theory *cum* bridging principle has absurd implications obviously requires that we discuss “content,” and this is precisely what I am doing at the moment.)

Moreover, even the second idealization, strict compliance, might be far less appealing than it might first appear. Simmons talks about what “can be achieved within the

---

<sup>23</sup> Ibid., p. 190-193.

<sup>24</sup> Again, for the elaborate version of this argument, see *ibid.* By the way, being indignant about and dismissive towards this argument, as I have experienced Rawlsians to be, is not quite the same as refuting it.

constraints set by the more or less intractable aspects of human nature and the nature of the world.” Respecting those constraints is a feature of what Rawls calls a *realistic* utopia.<sup>25</sup> However, the assumption that there can ever be a society where *everyone* abides by the Rawlsian principles of justice *all the time* is certainly *not* compatible with the intractable aspects of human nature<sup>26</sup> (therefore, incidentally, the parties to the original

---

<sup>25</sup> Rawls (2001), pp. 7 and 12-13.

<sup>26</sup> As Colin Farrelly, in “Justice in Ideal Theory: A Refutation,” *Political Studies* 55 (2007), pp. 844-864, at 845, rightly notes: “... by assuming full compliance, ideal theorists violate the constraints of a *realistic* utopia.” A Rawlsian critic has commented that one cannot show that the strict compliance assumption is not compatible with human nature by quoting Farrelly. Actually, I do not quote Farrelly as evidence but to give him credit for his observation. The actual *evidence* both Farrelly and I rely on, in contrast, is the long, brutal history of humanity, which teaches some very obvious lessons. This evidence cannot be made to disappear by quoting Rawls. Another critic claimed that a “charitable” interpretation of Rawls would relax the full compliance assumption. However, in reply, such “charity” would come at the price of accuracy: it would ignore what Rawls is explicitly saying in *A Theory of Justice*. Of course, in *Justice as Fairness: A Restatement* (Cambridge and London: The Belknap Press of Harvard University Press, 2003), p. 13, Rawls unceremoniously – as if this were not a major deviation from his earlier statements and in need of some explanation – qualifies the requirement that “everyone strictly complies” with the principles of justice by putting, in brackets, a “(nearly)” before it. This cavalier attitude of his is quite inappropriate: the full compliance assumption cannot be “relaxed” without thereby running into the insurmountable problem that I diagnose below in the context of Brock and Moellendorf.

position, far from being rational, must suffer from severe cognitive shortcomings: they make precisely this utopian assumption of strict compliance while *simultaneously* knowing “the general facts about human society” and “the laws of human psychology”<sup>27</sup> – this just doesn’t add up).

One of the facts of human psychology that is not compatible with the assumption of everyone behaving justly all the time is people’s partiality towards their friends and loved ones. Sometimes people will commit injustices, however slight, to promote their own friends and family members. Thus, striving for complete justice seems to involve striving for complete impartiality. Complete impartiality, however, is incompatible with love and friendship. (As many theologians, philosophers, poets and novelists have noted, some evils cannot be eradicated without also eradicating certain goods.) One is *necessarily* partial towards ones friends and beloved ones – this is partly what friendship and love *mean*.

Yet, it is dubious at best that the ideal of a fully just society can justify the sacrifice of love and friendship. To be sure, Rawls claims that justice “is the first virtue of social institutions, as truth is of systems of thought.”<sup>28</sup> But whether this is so or not, *complete justice* is not the first virtue of *societies* or *persons*, and justice is certainly not the only virtue. An *ideal* society, to deserve the name, has to balance different virtues in an adequate, human and humane way instead of sacrificing all other principles and values to the apotheosis of one.

A further problem with the bridging principle BPUTC is one I have already mentioned above in another context: since the basic structure of a society where we do not have full compliance might be just only when it does *not* conform to Rawls’s principles of justice, striving towards a Rawlsian ideal might actually move us *away* from justice. While this, I

---

<sup>27</sup> Rawls, *A Theory of Justice*, p. 119.

<sup>28</sup> *Ibid.*, p. 3.

submit, is a rather obvious logical point, and one that is well documented by the literature on “the problem of the second best,”<sup>29</sup> an example might help to make it still clearer.

I have repeatedly pointed out that Rawls does not offer us anything more than mere unsubstantiated assurances for the practical relevance of his “ideal theory.” One such unsubstantiated assurance is this: “It is clear from the preceding remarks that we need an account of penal sanctions however limited even for ideal theory.”<sup>30</sup> At least Rawls admits that such an account would be *limited*,<sup>31</sup> but he is not willing to admit what is obvious: “ideal theory” cannot provide *any* room for penal sanctions. After all, the parties to the original position, at all four stages of the four-stage sequence,<sup>32</sup> “assume that the principles they acknowledge, whatever they are, will be strictly complied with and followed by everyone.”<sup>33</sup> However, if they assume that, they obviously must also thereby assume that sanctions or even as much as the mere threat of sanctions are superfluous.<sup>34</sup> If, for example, they consider “Driving on the right side is just and imposing sanctions on anyone or threatening to do so is unjust” to be a principle of justice, then they will

---

<sup>29</sup> The *locus classicus* is R.G. Lipsey and Kelvin Lancaster, “The General Theory of Second Best,” *The Review of Economic Studies*, 24(1) (1956-58), pp. 11–32.

<sup>30</sup> Rawls, *A Theory of Justice*, p. 212.

<sup>31</sup> See also *ibid.*, p. 277.

<sup>32</sup> *Ibid.*, pp. 171-5.

<sup>33</sup> *Ibid.*, p. 308-9. Incidentally, Rawls’s concern about stability is due to the strict compliance assumption a red herring, as far as ideal theory is concerned.

<sup>34</sup> Incidentally, the idea of actual sanctions within a fully just society is incoherent. If everybody behaves justly, any sanction imposed would necessarily be imposed on an innocent person and therefore constitute an *injustice*. Then, however, the society would not be fully just anymore.

assume that in case they choose this principle *everybody* will drive on the right side while *nobody* will threaten to impose any sanctions on anyone. Moreover, if a penal system is superfluous, it is also *unjust*, for it would be a waste of public funds and thus clearly not compatible with Rawls's principles of justice. Rawls is not willing to fully admit this state of affairs because his claim that his "ideal theory" of justice is a standard for the assessment of real institutions looks particularly tenuous if in "ideal theory" there is not even room for a penal system – the latter, after all, being a rather important constituent of just basic structures in the *real* world.

As regards the "preceding remarks" Rawls mentions, they are entirely beside the point. At best one remark he makes in the "preceding remarks" might at least at first glance have a connection to Rawls's assurance "that we need an account of penal sanctions however limited even for ideal theory." He claims:

It is reasonable to assume that even in a well-ordered society the coercive powers of government are to some degree necessary for the stability of social cooperation. For although men know that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. ... By enforcing a public system of penalties government removes the grounds for thinking that others are not complying with the rules. For this reason alone, a coercive sovereign is presumably always necessary, even though in a well-ordered society sanctions are not severe and may never need to be imposed.<sup>35</sup>

In fact, however, these remarks do not lend any support to his claim that penal sanctions play a role even in "ideal theory;" on the contrary, they lend support to my claim that taking guidance from "ideal theory" leads one astray in the real world. First of all, a well-ordered society is not necessarily a society of strict compliance: the claim that

---

<sup>35</sup> Rawls, *A Theory of Justice*, p. 211.



sanctions *may* never need to be imposed suggests that it is conceptually possible that sanctions sometimes do need to be imposed in a well-ordered society – which, of course, is incompatible with the assumption of strict compliance. Second, that penal sanctions and the threat thereof might *make* more people abide with just principles and thus lead to more justice is precisely my point and certainly does not contradict my claim – which I justified above – that parties to the original position would oppose a penal system when, in conformity with “ideal theory,” they simply *assume* strict compliance. Only people who *try to secure* a high degree of compliance will opt for penal sanctions, not people who take strict compliance as a given.

These considerations show that people attempting to make the basic structure of their society more just under *real circumstances* should not take their lead from “ideal theory” – getting rid of penal sanctions, after all, does not make society more just under the circumstances of the real world.

Perhaps Simmons might object here that one has to strive for all aspects of the ideal at once: improving compliance while simultaneously, perhaps somehow proportionately to the degree of heightened compliance, “reducing” (whatever that means exactly) the penal system. However, this is obviously a goal conflict: if you want to improve compliance, reducing the penal system is not a good idea.

Giving the trade-offs involved, one will have to rely on some intuitionist assessment of what changes in reality actually *count* as coming closer to the ideal. But this makes Rawls’s “ideal theory” *doubly* intuitionistic: it is, first, intuitionistic with regard to the weighing and balancing of certain side-constraints Rawls imposes on the methods one uses to achieve “the perfectly just basic structure” (Simmons), and second, also “the measure of departures from the ideal is left importantly to intuition” (Rawls<sup>36</sup>).<sup>37</sup>

---

<sup>36</sup> Ibid., p. 216.

<sup>37</sup> Simmons knows this, of course. See “Ideal and Non-Ideal Theory,” pp. 12 and 18-30.

Simmons thinks that this still doesn't hinder "ideal theory" itself from being *non-intuitionist*,<sup>38</sup> but so what? Rawls wants to reduce the necessity to appeal to intuitions with his theory; he wants to deliver something that is better than intuitionism.<sup>39</sup> However, intuitionism does not need any appeal to "ideal theory" in the first place. It is *directly* applicable to the real world. Rawls's theory, in contrast, first sets up an ideal, then needs intuition to measure how far the real world is away from that ideal, and then also needs intuition in order to balance the side-constraints operating on methods used to come closer to the ideal. Rawls's theory, far from being an improvement over intuitionism, is only an entirely unnecessary complication.

In any case, since we cannot achieve the dubious ideal of full compliance and the appalling ideal of closedness anyway, we are well advised, morally and prudentially, to only try to achieve what we indeed *can* achieve and therefore to give up on the "ideal" of closedness and strict compliance altogether. Rather, if anything, we ought to try to make the basic structure fully just under the actual conditions of openness and partial compliance. (A critic commented on this point that "a society in which there is only partial compliance will never be 'fully just' because it contains non-compliers." That, however, is as true as it is irrelevant. I talked about the basic structure here, not about society, and since Rawls constantly talks about the basic structure *of* society, he obviously distinguished the two things. As a penal code or a law against rape can be fully just even if – or perhaps precisely because – there are non-compliers, so a basic structure can be fully just although society does not live up to its regulations.)

In that context it also needs to be noted that Simmons's constant talk about "the most just institutional structure that can be achieved" or about "the perfectly just basic structure" shows that he does not really understand the "complaint about counterfactual

---

<sup>38</sup> *Ibid.*, p. 28.

<sup>39</sup> Rawls, *A Theory of Justice*, pp. 30-40.

assumptions” he himself is complaining about. The point (or one of the points) is that Rawls’s “ideal theory” is entirely incapable of offering a description of *the* perfectly just basic structure or of *the* perfectly just society: it at best offers, to repeat it once again, a description of what a perfectly just basic structure would be like *under the conditions of closedness and strict compliance*. Provided the right background conditions, societies or basic structures can be perfectly just in *a million other ways* besides in living up to Rawls’s principles of justice.<sup>40</sup> Some of those conditions are actually *real*. Thus, there is no need, and certainly no reason, to rely on Rawls’s unrealistic dystopia(s).

## II. Is there a point to unrealistic assumptions?

Yes, of course there is, and therefore there can be useful ideal theories, as we will see by the example of Fritz Allhoff’s theory below (their use, however, will never lie in telling us what is just or moral in the real world: *no* ideal theory can do this without a bridging principle). However, that does not help Rawlsian “ideal theory” at all. Indeed, Rawls’s theory completely misses the point that wrong assumptions can have. Before returning to that, however, let us have a look at what defenders of Rawls take to be the benefits of the wrong assumptions in Rawls’s theory. Simmons thinks that this is “fairly plain”:

First, of course, if we compare the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles, the

---

<sup>40</sup> See in this context also Robert E. Goodin’s suggestion of “indexing our political prescriptions to socio-economic circumstances” in “Political Ideals and Political Practice,” *British Journal of Political Science* 25(1) (1995), pp. 37-56, at 56.

different effects we observe can reasonably be taken to be wholly the responsibility of the different ordering principles themselves. So our comparison turns out to be quite strictly a comparison only of the principles of justice. If instead we try to evaluate principles in terms of how societies governed by them would operate with a “normal” amount (or a certain percentage) of noncompliance with them (supposing we can even make sense of that hypothesis), we will likely find both that our evaluations yield quite indeterminate results and that the results depend on more than simply the different ordering effects of the principles being compared.<sup>41</sup>

Actually, full compliance is not the only idealization in Rawls’s theory. Thus, at best we would have a comparison of principles under certain conditions, in this case under the unrealistic and unappealing condition of closedness. Second, we would have a comparison not just of the principles by themselves, but of full compliance with different principles under different conditions. And what does this tell us about the real world? Very little. To repeat the point again: even if fully complying with Rawls’s two principles of justice under conditions of closedness would produce a fully just society, this does not mean that implementing those principles of justice in the real world would lead to more justice. It might well lead to much less justice.

Besides, it seems that experts in the field of comparative politics are quite able to compare “principles in terms of how societies governed by them would operate with a ‘normal’ amount (or a certain percentage) of noncompliance.” Of course this might often need extrapolation and perhaps a certain amount of speculation, but it is certainly nothing one cannot make sense of.

---

<sup>41</sup> Simmons, “Ideal and Non-Ideal Theory,” p. 8.

Zofia Stemplowska also thinks that false assumptions can be useful. She claims:

Notice then that tackling broad problems ... inevitably requires making some false assumptions. For example, to theorize at a certain level of generality, we will have to make some approximations that, being approximations, will be false, but that will also be necessary to make any progress at all. In addition to making approximations, theorists may have to pretend that certain difficulties do not arise or they will never be able to say what justice, on the whole, requires ...<sup>42</sup>

However, first, Stemplowska only claims, but does not show, that there is anything “inevitable” about this. Second, Rawls’s “ideal theory” of justice does not say what justice is “on the whole,” but only what justice is under the conditions of closedness and full compliance. Third, while it is true that theories and models can be quite useful although they only approximate the truth – they can be useful if other theories and models do not even manage to accomplish that – one is well advised to keep in mind that the following is not correct: “If a theory of justice T correctly claims that under conditions X a basic structure B would be just, then under conditions that approximate X a basic structure that approximates B would be approximately just.” It might well be the case that the slightest deviation from the conditions X would require a dramatic deviation from B in order to keep the basic structure even approximately just (as, again, the “theory of the second best” confirms<sup>43</sup>). If Rawls or Rawlsians think that such a dramatic

---

<sup>42</sup> Zofia Stemplowska, “What’s Ideal About Ideal Theory?”, *Social Theory and Practice* 34(3), (2008), pp. 319-340, at 327.

<sup>43</sup> See n. 28. Erman and Möller interestingly claim that “as shown by Goodin, the problem of ‘the second best’ is not a criticism of ideal theory but concerns the inevitable

deviation is not required in the case of Rawls's theory, they would have to provide an argument for that. They don't. In fact, they don't even notice the problem. Last but not least, the conditions of closedness and strict compliance do not approximate reality; rather, they are entirely removed from reality.

Stemplowska's further remarks on the significance of "ideal theory" simply presuppose that Rawls's "ideal theory" has provided us with an ideal worth striving for, indeed with *the* ideal (with the ideal of justice "on the whole"). I have already dealt with these mistaken assumptions when discussing Simmons above and shall say no more about them.

Yet, Stemplowska also tries to defend the full-compliance assumption in particular.

She states:

---

problem of tradeoffs that appear if we attempt to realize several competing or conflicting ideal conditions simultaneously." See their "Three Failed Charges against Ideal Theory," p. 31. However, this problem clearly *is* a problem for ideal theory, and this is, *pace* Erman and Möller, exactly what Goodin *says, precisely* using the example of Rawls's ideal theory. See Goodin, "Political Ideals and Political Practice," pp. 54-55. His two suggested remedies to this problem of ideal theory do not *help* ideal theory but obviously make it superfluous: first, searching for principles that are "relatively insensitive to their setting," and second, "indexing our political prescriptions to socio-economic circumstances" (*ibid*, p. 56). This idea of "indexing" is of course nothing more than the idea that what is just under some circumstances, for example under those of the idealized world, is not yet just under other circumstances, for example under those of the real world. Thus, looking to the former will not give us guidance for the latter.

But if people do not always act as they should, then the assumption of full compliance guards us from making our demands too permissive towards those who are simply unwilling to comply with reasonable limits on the permissibility of their actions. What motivates the assumption of full compliance, then, is the worry that without it the normative requirements we come to acknowledge will be normatively inaccurate. The assumption of full compliance guards us from accepting, as David Estlund has put it, the implausible dictum that “ought implies reasonably likely.”<sup>44</sup>

What is this supposed to mean? That, for example, without *assuming* that there is no rape or murder on Earth we would be more lenient towards rapists or murderers (who, by the way, is even *capable* of making such unrealistic assumptions in the real world – *we* are not parties to the original position, after all)? Isn’t rather the opposite the case, that is, if we assumed there is no rape, our attitude towards rapists would be like our diluted attitudes towards fictional characters? Certainly we would not consider it a societal problem. We would consider it, thanks to our other-worldly assumption, to be no problem at all.

But probably Stemplowska means something else: without an *idea* of what an ideal world without rape (or murder) would be like, we would be more lenient towards rapists (or murderers). But that is clearly wrong, too. People tend to be against rape because they have some idea about what rape *is* like. Because they know that, they are against it. They do not need “ideal theory.”

---

<sup>44</sup> Stemplowska, “What’s Ideal About Ideal Theory?”, p. 332-3. Stemplowska is quoting David M. Estlund, *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2007), p. 265.

Laura Valentini also tries to save “ideal theory,” or at least some “ideal theory.” She thinks there is a “principled way of distinguishing between ‘good’ and ‘bad’ uses of idealisations.”<sup>45</sup> Her way of trying to do this, however, is unconvincing, particularly so with regard to Rawls’s “ideal theory.” She first correctly identifies the most serious charge against “ideal theory” as this:

ideal theory ‘is at best morally irrelevant, at worst morally destructive’. This charge can be summarised as follows: if we apply principles developed under ideal conditions to real-world circumstances—namely those circumstances for which they have allegedly been designed—we are bound to obtain morally counterintuitive results.<sup>46</sup>

However, she then goes on to argue that Rawls’s domestic “ideal theory” engages in “good” idealizations while his international “ideal theory” engages in “bad” ones. Yet, in arguing the former she suddenly misrepresents what she had previously identified as the most serious charge and now claims that it states that certain “omissions are not harmless: by neglecting to discuss racial and sexual discrimination, ideal theory loses its capacity to guide action in the real world.”<sup>47</sup> And she then tries to refute this charge by arguing, quoting Rawls, that an omission is not as such a fault.<sup>48</sup>

But of course omissions are not as such a fault, and nobody has denied this. Theories

---

<sup>45</sup> Laura Valentini, “On the Apparent Paradox of Ideal Theory,” *Journal of Political Philosophy* 17(3) (2009), pp. 332–355, at 352.

<sup>46</sup> *Ibid.*, p. 341. Valentini quotes here Baier, *Postures of the Mind*, p. 225.

<sup>47</sup> Valentini, “On the Apparent Paradox of Ideal Theory,” p. 343.

<sup>48</sup> *Ibid.*, p. 343-7.



need not be applicable to everything. In addition, a theory does not even need to discuss certain things *explicitly* in order to *be* applicable to them. Rawls's theory of justice, for example, is straightforwardly applicable to racial and sexual discrimination *under conditions of closedness and strict compliance*. Rawls's theory flatly implies that if racial and sexual discrimination does not conform to Rawls's two principles of justice, then racial and sexual discrimination are *wrong* under conditions of closedness and strict compliance. But *this is* precisely the charge: Rawls's theory implies such things about a situation of strict compliance and closedness, but these implications cannot tell us what is just or unjust in the *real* world; they are *irrelevant* for the real world. And there is certainly nothing in Valentini's discussion to show otherwise. Thus, her discussion of the domestic case simply misses the point of the criticism.

When discussing Rawls's international theory of justice, however, she again presents the point correctly: Rawls's description of the idealized world "does not apply, i.e., it is, technically, *irrelevant* to existing circumstances, which is precisely what [the most serious charge] claims."<sup>49</sup> And she agrees that this charge is correct. But of course, as we saw, it is no less correct with regard to Rawls's domestic theory.

That her distinction between good and bad Rawlsian "ideal theory" is mistaken can also be seen by examining her attempt to show how exactly the difference comes into being. She claims that in Rawls's domestic "ideal theory" "the *subject* of justice," namely the basic structure, does not "end up being idealized." But, allegedly:

Things look different when we turn to *The Law of Peoples* ... As for Rawls's international theory, this asks how almost self-contained political communities

---

<sup>49</sup> Ibid., p. 348.

ought to behave towards one another, on the assumption that, since they are well-ordered, a just (fair) background is in place. But because real-world societies are not well-ordered, such a background does not exist, and it is precisely its absence that gives rise to the question of international justice in the first place.<sup>50</sup>

Actually, the *subject* of the respective theory of justice is in *both* cases, in the domestic and in the international one, a “basic structure.” That is what the parties to the original position deliberate about, and that is what the theory deals with. Clearly, I do not need to remind anyone of this with regard to Rawls’s domestic theory. But with regard to his international theory, it might be useful to quote Rawls saying this:

Thus, people’s representatives are ... deliberating about the correct subject, in this case the content of the Law of Peoples. Here we may view that law as governing the basic structure of the relations between peoples.<sup>51</sup>

Thus, there is no difference. In *A Theory of Justice* the basic structure is governed by the two principles of justice. In *The Law of Peoples* the “basic structure” is governed by the principles of the Law of Peoples. In neither of the two theories are the principles “idealized” (whatever it may mean to idealize a principle). In both theories the background conditions are idealized: full compliance and closedness there; well-ordered societies that fully comply here. The idealizations in both theories, being idealizations, are unrealistic, and therefore both theories are inapplicable to real societies and basic structures. Hence Valentini’s attempt to save one of those theories from this charge fails.

---

<sup>50</sup> Ibid., p. 351. Incidentally, I am not aware of Rawls stating anywhere that those communities are “almost self-contained.”

<sup>51</sup> Rawls, *The Law of Peoples*, p. 33.

Holly Lawford-Smith's attempt to save Rawls's "ideal theory" is not successful either. She in fact and very surprisingly argues "that there are no idealized assumptions in Rawls ... , there are just counterfactual statements that are either true or false of individuals *as they actually are*. If the counterfactuals are true, a certain world is justified as a good one."<sup>52</sup> How so? She explains:

Consider an analogue. If orchestras had conducted their auditions blind (e.g. with those auditioning obscured from the view of those selecting) before 1970, there would have been a greater proportion of women in the world's best orchestras at that time. The antecedent of the counterfactual is false, because orchestras did not begin the practice of blind auditioning until fairly recently. But it seems remarkably likely that the conditional itself is true, because the proportion of women in orchestras has increased dramatically since the introduction of blind auditions. And something seems to follow from the truth of that conditional, namely that blind auditioning is something we should be committed to. It's just the same with the counterfactuals in Rawls and Dworkin. If their counterfactual conditionals are true—if we are in a position to choose principles of justice, if we are on the desert island—then they have some normative force. If people would choose certain principles of justice were they not made blind and selfish by their psychological limitations, then perhaps those principles of justice should be in place.

What I'm suggesting is that far too much attention goes into attacking the antecedents of the counterfactuals, with objections like 'yes, but people aren't

---

<sup>52</sup> Holly Lawford-Smith, "Debate: Ideal Theory—A Reply to Valentini," *The Journal of Political Philosophy* 18(3) (2010), pp. 357-68, at 363.

really like that, so nothing follows for real people!’<sup>53</sup>

However, the fact of the matter is that Rawls’s ideal (domestic) theory of justice, summarized in RITJ above, is not a counterfactual about persons at all. It is a counterfactual about basic structures; and it is designed to be exactly that.

More importantly still, Lawford-Smith misses the point of the critics of “ideal theory.” They do not necessarily criticize the *veil of ignorance*; they criticize the fact that Rawls’s “ideal theory” of justice is limited to basic structures under the conditions of closedness and strict compliance, conditions that do not hold in the real world and thus make the theory inapplicable to the real world. To be sure, there is a connection between those idealizing conditions and the original position: after all, the parties to the original position assume that their societies will be closed and will be societies of full compliance. The limitations are made operational in the original position by precisely these assumptions. And precisely *these* assumptions are the ones the critics of “ideal theory” criticize: idealizations *about the societies* for which principles of justice are chosen. They do *not*, however, necessarily criticize that *the parties to the original position* are idealized. Lawford-Smith is right in suggesting that the parties to the original position *are* idealized; they are, as she says, “not blinded by envy, not risk-takers, not imperfectly rational.”<sup>54</sup> However, the parties themselves do not assume that there will be no envy, no risk-taking and no imperfect rationality *in the societies* for which they choose the principles, as

---

<sup>53</sup> Ibid.

<sup>54</sup> Ibid. Of course, one might doubt that being fanatically risk-averse is “ideal.”

Rawls makes perfectly clear.<sup>55</sup> Thus, this is no idealization at all in the sense intended by the critics of “ideal theory.”

Thus, to avoid distorting the actual issue Lawford-Smith would at least, if she insists on talking about the original position, have to identify the *right* counterfactual in that context, which would be this one: “If, behind a veil of ignorance, rational people not blinded by envy and risk-taking were to choose principles of justice for their society and *if they also made the entirely unrealistic assumption that their society will be closed and blessed with strict compliance*, then they would choose the principles of justice X.” Yet, unsurprisingly she does not explain why *this* counterfactual should be able to justify anything with regard to *real* societies.

Thus we have to modify her orchestra “analogue” a little in order to get a real analogy to the counterfactual that is actually at issue. This is one possibility that suggests itself: “If orchestras would conduct their auditions blind (e.g. with those auditioning obscured from the view of those selecting) *and also made the entirely unrealistic assumption that whomever they would choose would fully comply with their standards*, they would just randomly choose whoever shows up for the audition.” What does this show or “justify” with regard to how we should conduct auditions in the real world? Nothing.

Finally, allow me to say something about the curious comparison both Valentini and

---

<sup>55</sup> “The motivation of the persons in the original position must not be confused with the motivation of persons in everyday life who accept the principles of justice and who have the corresponding sense of justice. . . . Thus, more generally, the motivation of the parties in the original position does not determine directly the motivation of people in a just society.” Rawls, *A Theory of Justice*, pp. 127-8.

Lawford-Smith (the latter following the former) draw between Galileo's Law of Falling Bodies (or his frictionless plane thought experiment) and Rawls's "ideal theory."

Valentini states:

Famously, Galileo's Law of Falling Bodies is elaborated under the idealised assumption that friction does not exist. In spite of it being 'false' of existing falling bodies, such an idealised assumption is widely regarded as appropriate for theoretical purposes. Galileo's Law is not charged with being unsound because it is arrived at by assuming friction away. ... In fact, friction can be taken back into account when such a Law is employed to make predictions about falling bodies in the real world.<sup>56</sup>

And Lawford-Smith concurs:

I think Galileo's thought-experiment idealizes its subject and yet is none the worse for it. It assumes away friction, then makes a prediction about how bodies should behave. When we apply the theory, we build real-world complications like friction back into the story. One way of understanding what we are doing is that we specify an ideal, the way things behave in simplified or ideal conditions, such as in worlds without gravity, and then we *relax* the assumption of ideality by reintroducing real complications, like friction. This seems to apply *equally* to Rawls's ideas about international justice ...<sup>57</sup>

A few comments are in order. First, Galileo does not develop his Law of Falling Bodies under the idealized assumption that friction *does not exist*. He rather says that if

---

<sup>56</sup> Valentini, "On the Apparent Paradox of Ideal Theory," p. 354.

<sup>57</sup> Lawford-Smith, "Debate: Ideal Theory," p. 365.

there is no friction, the bodies will behave as he says they will. In some parts of the world (namely those we call vacuum) there *indeed* is no friction. Thus, his Law, even if formulated as a conditional (“If there is no friction, then...”) – which is not even necessary in modern physics, where you can get rid of the idealization by using equations about forces – is directly applicable to large parts of the real world. That is more than can be said about Rawls’s theory.

Second, if we made the idealizing assumption that there is friction *everywhere* in the real world, then Galileo’s law, formulated as a conditional with a now wrong antecedent, would *by itself* be *equally inapplicable* to the real world as RITJ. For example, it would not tell us whether objects falling in an environment with friction would fall faster or slower than in a vacuum, would come to a grinding halt and then go on falling, or even reverse course. It would be useless.

Fortunately for Galileo Galilei, however, he differs from Rawls by providing more in his theory than just that one conditional law. Instead, he described a lot of real world cases and gave *plausible, detailed and quite accurate explanations* of how to factor friction in again, stating that friction would slow down objects, have a greater effect on light objects than on heavy ones, etc. – thus enabling his readers to use his theory to make concrete predictions about the *real* world. Rawls, in contrast, does not offer *anything* plausible *or* detailed *or* quite accurate by way of explaining how his “ideal theory” of justice is to be brought to bear on the real world.

Lawford-Smith actually seems to admit that much, albeit only in the last footnote of her article:

This claim that ideal standards can be relaxed relies to some extent upon there

being a bridge from ideal to non-ideal theory, and there seems to be only one proposal on the table about how that bridge might work. See (Murphy 2000). I don't have in mind a formal algorithm for the relaxation from ideal standards, so if that is necessary then my proposal will be vulnerable until a plausible bridge is established.<sup>58</sup>

First, Lawford-Smith not only has no formal algorithm for the “relaxation from ideal standards” in mind; she offers *no* bridging principle at all, whether in the form of an algorithm or a metaphor. Second, Murphy's proposal is not so much a bridging principle but in need of a bridge itself. His central principle, the compliance condition, goes:

Agent-neutral principles should not under partial compliance require sacrifice of an agent where the total compliance effect on her, taking that sacrifice into account, would be worse than it would be (all other aspects of her situation remaining the same) under full compliance.<sup>59</sup>

This principle fails to be action-guiding as long as it is not explained what full compliance requires (and Murphy does not explain that; thus, he has developed his “non-ideal theory” before his “ideal” one), it has counter-intuitive implications in many realistic scenarios,<sup>60</sup> and, most importantly in our present context, it is certainly not a

---

<sup>58</sup> Ibid., p. 367, n. 21.

<sup>59</sup> Liam B. Murphy, *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press, 2000), p. 80.

<sup>60</sup> See, for example, Richard J. Arneson, “Moral limits on the demands of beneficence?”, in Deen K. Chatterjee (ed.), *The Ethics of Assistance: Morality and the Distant Needy* (Cambridge: Cambridge University Press, 2002), pp. 33-58, esp. at 36-7.



bridging principle for Rawls's principles of justice at all: it does not tell us how Rawls's principles of justice should be "relaxed," to use Lawford-Smith's expression, to apply to the real world. Rawls's principles of justice are simply not Murphy's concern.

However, Michael Phillips (whom Lawford-Smith does not mention) has at length considered strategies to bridge the gap between Rawlsian "ideal theory" and the real world, and rightly discarded them all.<sup>61</sup> Perhaps in an idealized world where we *would have* a bridging principle "ideal theory" might be applicable, but, again, we are not living in an ideal world; we are living in the real one. We have now been waiting for 40 years to be presented with a viable bridging principle, but nothing remotely plausible has been offered – in fact, Rawlsians and ideal theorists have quite simply ignored the problem. Most will continue to do so, and it is to be expected that none of them will solve it. For those interested in the real world, it is time to move on.

### III. Good "Ideal Theory" and Bad "Ideal Theory"

Thus, the fact remains that Rawls's "ideal theory" is useless. This statement is quite compatible with the admission that making wrong assumptions and philosophizing about idealized or even outrageously unrealistic circumstances can make very good sense. Philosophers, after all, (or at least quite a number of them) use hypotheticals all the time. To use one example, Fritz Allhoff has developed what I will call here Allhoff's "ideal theory of torture" (that is not his choice of terms). Allhof develops an highly idealized

---

<sup>61</sup> Phillips, "Reflections on the Transition from Ideal to Non-Ideal Theory."

ticking bomb example (there are ticking bomb examples that are not idealized, incidentally<sup>62</sup>) and states that this case shows or at least intuitively strongly suggests “that torture is permitted in highly idealized cases and is *completely* silent about whether torture would be permitted in other (less idealized) cases.”<sup>63</sup> Thus, he is well aware of the limitations of “ideal theory.” Yet, he thinks that the idealized case is still relevant for practice, namely in this way:

I directly argue against Kantianism (and other views with similar commitments) on the grounds that, if they cannot accommodate the intuitions in [idealized] ticking time-bomb cases, they simply cannot be plausible moral views ... I postulate that, even if this paper has dealt with idealized cases, it paves the way for the justification of torture in the real world by removing some candidate theories (e.g., Kantianism) and allowing others that both could and are likely to justify real-world torture.<sup>64</sup>

This is a useful indirect application of “ideal theory” to the real world, whether or not one agrees with Allhoff’s conclusions as to the justifiability of torture. Unfortunately, though, there is no such useful application of Rawls’s theory. First of all, while Allhoff’s

---

<sup>62</sup> See Uwe Steinhoff, “Defusing the Ticking Social Bomb Argument: The Right to Self-Defensive Torture”, *Global Dialogue* 12(1) (2010), <http://ftp4.dns-systems.net/~worlddiag/index.php>, the section entitled “Straw Man Idealisations.”

<sup>63</sup> Fritz Allhoff, “A Defense of Torture: Separation of Cases, Ticking Time-bombs, and Moral Justification,” *International Journal of Applied Philosophy* 19(2) (2005), pp. 243-64, at 246.

<sup>64</sup> *Ibid.*, p. 243, abstract.

idealized ticking bomb case really tests our intuitions and it is an empirical fact (acknowledged by absolutist anti-torture opponents and confirmed by unpublished empirical research of Allhoff's) that most people will find torture justifiable under *those* extreme idealized conditions, Rawls does not test our intuitions at all, despite all his talk about "reflective equilibrium." Instead, he argues that the parties to the original position will choose the principles of justice he proposes if they are to choose those principles for a closed society blessed with strict compliance. However, he has *designed* those parties for the purpose of doing that in the first place, so the whole enterprise looks like legerdemain (a point made by many critics of Rawls). If, however, we are talking about "us" or most people (even if we are only talking about most people in liberal democracies), it is simply not true that "we" or they find the Rawlsian principles of justice compelling principles for a situation of closedness and strict compliance.<sup>65</sup> Rawls does not clarify or reveal our intuitions in the least.

Second, and even more importantly, let us suppose Rawls's "ideal theory" of justice were correct (which it is not). Where would its indirect applicability come from? Let us have a look at the structure of Allhoff's argument again. Roughly, it goes like this: if Kantianism is right, torture is prohibited under all circumstances.<sup>66</sup> The idealized ticking bomb case shows that torture is not prohibited under all circumstances. Thus, Kantianism is wrong. If Kantianism is wrong, however, it cannot be used as an argument against

---

<sup>65</sup> See Gillian Brock, *Global Justice: A Cosmopolitan Account* (Oxford: Oxford University Press, 2009), pp. 54-7, where she reports empirical evidence to this effect gathered by, among others, Frohlich and Oppenheimer.

<sup>66</sup> Just for the record: I disagree with Allhoff on this issue.

torture anymore – neither in the idealized world, nor in the real world. Thus, certain arguments for torture that could only be stopped by Kantianism would now have to be accepted – and this is obviously a very relevant result for the real world.

What would the parallel argument in the case of Rawls’s theory be? What theory or view does it refute? Well, it would refute utilitarianism and libertarianism – and this would make it indirectly applicable to the real world and thus useful, wouldn’t it? Not quite. It would indeed make it indirectly applicable, but it would not make it useful if “useful” is supposed to be incompatible with being superfluous. For while, on the assumption that Kantianism indeed implies an absolute prohibition of torture, Allhoff’s idealized ticking bomb example is an extraordinarily strong hypothetical if it comes to undermine the conviction that torture is prohibited under *absolutely all* circumstances, there are certainly much better and widely available thought experiments than the one provided by Rawls’s voluminous “ideal theory” to undermine libertarian and utilitarian intuitions (thought experiments, for example, about torturing babies in order to maximize happiness or about letting babies drown in the puddle in front of you although you could save them by slightly moving your toe).

Thus, Rawls’s theory of justice, if correct, would, for example, undermine real-world *utilitarian or libertarian arguments* in support of totalitarian dictatorships or Manchester capitalism – something other thought experiments achieve much more economically, efficiently and convincingly. It would not, however, show that other arguments for totalitarian dictatorships must be wrong, that totalitarian dictatorships are unjust in the real world, that we should strive for one of the Rawlsian “ideals” in the real world, or that we should make our basic structures in the real world conform more to Rawls’s principles

of justice. It would, in short, show that some justificatory doctrines or arguments are *invalid* in the real world, but it would not tell us what is just and what we should do in the real world. It is and remains useless: an unnecessary complication that offers absolutely nothing to make it superior to intuitionism.

#### IV. What “Non-Ideal Theory”?

I have used the terms “ideal theory” and “non-ideal theory” in quotation marks throughout the paper. The reason for this is that I think that a useless theory is anything but ideal; and while I think that Rawls’s “non-ideal theory” is indeed not ideal, it is also not a *theory* – it is merely a collection of dogmatic ad-hoc stipulations.

Before going into this, however, let us have a brief look at the merits and demerits of the distinction itself. Simmons states:

In its barest outline, of course, the idea behind the distinction is simple and appealing, no doubt accounting for the widespread acceptance of (or indifference toward) Rawls’s version of it: “the intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances.” [TJ, p. 245] This “ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception” (TJ, p. 246). “Nonideal theory asks how this long-term goal might be achieved, or worked toward, usually in gradual steps. It looks for courses of action

that are morally permissible and politically possible as well as likely to be effective.”<sup>67</sup>

However, this distinction is anything but appealing if it is all there is with regard to morality in Rawls – and it is all there is. Rawls and Simmons overlook a very big part of morality. Morality cannot be reduced to the question of what ideal we should strive for and of how we can do so permissibly. Very often, at the very least, the reasons why we ought to do something have nothing whatsoever to do with whether the course of action in question advances the ideal or not. To be sure, perhaps Rawls and Simmons think this insight can be accommodated by referring to the impermissible – some things, they can admit (and they seem to admit), can be impermissible in and of themselves, for example torturing babies, even if they advance the ideal. Yet, this strategy does not work with what is *permissible* in itself (as opposed to what is permissible *as a means to promote the ideal*). It seems, for example, that people have a *right* to give their own interests in certain circumstances priority over advancing the ideal. That is, they may at some times go on a vacation, even if not doing so would not be impermissible in itself, and even if doing so does not advance the ideal. Of course, probably in the *ideal* world Rawls’s theory would also give individuals these rights, but given how he (and Simmons) distinguish “ideal” and “non-ideal theory,” this is not the case for the real world. Here “non-ideal theory” seems to succumb to a monistic fanaticism not too different from the one of utilitarianism: unless doing so would mean doing something that is impermissible

---

<sup>67</sup> Simmons, “Ideal and Non-Ideal Theory,” p. 8. With “TJ” Simmons refers to the *first* edition of *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971). The last quote is from Rawls, *The Law of Peoples*, p. 89.

in itself, there is *one and only one thing you ought to do*, namely promote the ideal.<sup>68</sup>

I said at the beginning of this section that Rawls's "non-ideal theory" is merely a collection of dogmatic ad-hoc stipulations. For Rawls does not explain how we tell the permissible means to achieve the ideal from the impermissible ones. After all, he cannot rely on the guidance of "ideal theory" there, for, as he rightly notes, "the parties [to the original position] are choosing a conception of justice suitable for favorable conditions," but then, of course, "it is possible that [the chosen principles] no longer hold" in the real world.<sup>69</sup> Thus, whatever principles, prohibitions, permissions or rights are valid in the ideal world, it cannot be presupposed that they also apply to the real one.

The consequence of this logical fact is that the permissions, rights, duties and side-constraints Rawls envisions for the real world pop up out of the blue, like Cain's wife in

---

<sup>68</sup> A critic claimed here that Rawlsian ideal theory is allegedly "primarily" interested in the policies that *governments* should pursue and that therefore individuals will be left free to pursue all kinds of projects against the background of policies that aim to bring the society as a whole closer to the ideal. Actually, however, first, Rawls's theory is not primarily interested in governments but in basic structures. Second, it is *also* interested in the behavior of individuals. *A Theory of Justice* has two chapters entitled "Principles for Individuals," and Rawls puts much emphasis on the "natural duty of justice." Besides, in the quote Simmons adduces, Rawls says that "we" are to achieve the ideal. Thus, my point stands: it is entirely unclear where the freedom, that is, the permission to pursue all kinds of projects is supposed to come from if ideal and non-ideal theory in the sense of Simmons and Rawls is all there is.

<sup>69</sup> Rawls, *A Theory of Justice*, p. 215-6.

the Bible, who seems to have been produced by some competing Genesis the reader is left in the dark about (the pious reader, of course, doesn't care). This can most clearly be seen in Rawls's *The Law of People*, where no attempt is made to actually *justify* the principles for the non-ideal world (admittedly, the "justifications" of the ideal principles are not particularly elaborate either). There is simply no discernible connection of the non-ideal principles with the ideal ones.

While the second part of *The Law of Peoples* looks more like revelation than philosophical argument, in *A Theory of Justice* Rawls does make some rudimentary efforts to explain where the principles for the real world come from. He suggests two different strategies that contradict each other. As regards the first one, he states (and curiously Simmons does not discuss this statement in explaining Rawls's stance towards "non-ideal theory" – other Rawls commentators do not pay much attention to it either):

Non-ideal theory ... is worked out after an ideal conception of justice has been chosen; only then do the parties [to the original position] ask which principles to adopt under less happy conditions.<sup>70</sup>

The less happy conditions are "natural limitations and accidents of human life" as well as "injustice ..., either in the social arrangements or in the conduct of individuals."<sup>71</sup> Thus, we have here a variation of the original position *where the assumption of strict compliance is dropped*.

There are two ways to interpret this quote. First, the emphasis on the sequence – first "ideal theory," *only then* "non-ideal theory" – suggests that the parties to the non-ideal, as

---

<sup>70</sup> Ibid., p. 216.

<sup>71</sup> Ibid., p. 215.



it were, original position already have the ideal principles of justice in hand and now apply them to the real world (the strict compliance assumption dropped) while still behind a veil of ignorance. However, with such an approach Rawls simply begs the question. It is on the very same page, after all, that he acknowledges that the principles that have been chosen for ideal circumstances “might no longer hold” for the real world. But then to ask the parties to the non-ideal original position to happily apply them to the real world anyway will certainly not do. Besides, some explanation would be due as to *what* such an “application” would look like. Applying them directly would lead to counter-intuitive results. And since even we (Rawls and Simmons included, as I have argued), who are not burdened by a veil of ignorance, do not have a clue as to how to “indirectly apply” the principles of justice in any plausible way to the real world, the parties to the original position behind the veil of ignorance will not know, either. In short, Rawls does not solve the problem here, he only moves the bump in the carpet.

The second interpretation relies on the fact that the quoted passage does not really semantically and logically *imply* that the parties to the non-ideal original position are guided in their decisions by the ideal principles. But if they are not, of course, the embarrassing question – a very good question, incidentally<sup>72</sup> – arises as to why we shouldn’t do away with “ideal theory” altogether and let the parties behind the veil of ignorance make their decisions without making any idealizing assumptions in the first place. In other words, this second interpretation effectively does away with “ideal theory.” It would no longer have any discernible function. (Incidentally, you can choose

---

<sup>72</sup> See also Phillips, “Reflections on the Transition from Ideal to Non-Ideal Theory,” p. 554-5, and esp. 568, n. 2.

an ideal to strive for even without making idealizing assumptions. I suppose, for example, that Rawls, when devising his “ideal,” did not really assume that we are living in a world of strict compliance.)

The second strategy Rawls employs ignores the logical fact I was referring to a few paragraphs ago, namely the fact that you cannot import principles justified for ideal circumstances into the real world in order to then “bridge” the gap between ideal principles and the real world: after all, if there is a gap, there is a gap, so that the attempt to take bridging principles from the ideal world of all places ignores the very problem it is meant to solve.

This is precisely what Rawls does when he discusses the alleged duty to comply with an unjust law and civil disobedience in the real world and in that discussion constantly relies on an alleged “natural duty of justice.”<sup>73</sup> He claims that this duty would be chosen in the original position,<sup>74</sup> but even if that were true, it does certainly not follow that it applies to the real world. Charles Beitz, incidentally, commits the same mistake. He claims: “Ideal justice ... comes into nonideal politics by way of the natural duty to secure just institutions where none presently exist.”<sup>75</sup> How that “natural duty” itself, however, “comes into non-ideal politics” he does not explain – nor does Rawls.

---

<sup>73</sup> Rawls, *A Theory of Justice*, pp. 308-12 and 319-43.

<sup>74</sup> *Ibid.*, pp. 293-301.

<sup>75</sup> Charles R. Beitz, *Political Theory and International Relations* (Princeton: Princeton University Press, 1999), p. 171. He gives this explanation in a chapter entitled “Applications to the Nonideal World.” The possibility that there might be no applications of Rawls’s theory to the real world is one he never considers.

Even more explanations would be needed – and again are not provided – if one combines this mistake with the use of the second interpretation of the first strategy. Thus, two “cosmopolitans,” Darrell Moellendorf and Gillian Brock, defending *Cosmopolitan Justice* and *Global Justice*, respectively, in their books of these titles, devise an original position whose parties clearly do not work under the assumption of strict compliance (otherwise they would not deliberate about military intervention or terrorism).<sup>76</sup> And both authors assume throughout their books that a natural duty of justice would be justified by such a procedure (along with a lot of other things).

Their assumption is unwarranted. First of all, Rawls emphasizes that “the veil of ignorance excludes all knowledge of likelihoods”<sup>77</sup> and Brock agrees, saying “that you should not have access to information that could lead you to deduce the odds of your being in some circumstances as against others”<sup>78</sup> (and Moellendorf nowhere disagrees). The reason for this is that justice is supposed to be “blind.” Now, in Brock’s account the parties to her original position “are to participate in deciding what would be a fair framework for interactions and relations among the world’s inhabitants.”<sup>79</sup> Yet, when the parties to the original position know that there will not be strict compliance and do not know any probabilities, a first question they will ask Brock is: “What difference does it

---

<sup>76</sup> Darrel Moellendorf, *Cosmopolitan Justice* (Cambridge: Westview Press, 2002), pp. 16-18; Gillian Brock, *Global Justice: A Cosmopolitan Account* (Oxford: Oxford University Press, 2009), pp. 48-54.

<sup>77</sup> Rawls, *A Theory of Justice*, p. 134.

<sup>78</sup> Brock, *Global Justice*, p. 49.

<sup>79</sup> *Ibid.*

make what framework or principles we choose?” Rawls has an easy answer to that question: “If you choose, for example, a principle that says that just people need to cut off their ears when they are 18, then you will cut off your ears when you are 18 since *ex hypothesi* there is strict compliance. That’s the difference.” What can Brock say? “If you choose, for example, a principle that says that just people need to cut off their ears when they are 18, then ...” Yes, then *what?* Since there is no strict compliance and no knowledge of probabilities, absolutely nothing follows. In other words, Brock cannot even provide the parties to the original positions with a *motivation* to choose anything. And she can certainly not explain why people would choose cosmopolitan principles rather than National Socialist ones. Given the very assumptions made in the original position, it makes no difference what the parties will choose. The reasoning applies to Moellendorf, too.

One might wish to change the original position by postulating that while choosing certain principles does not ensure full compliance, it makes compliance more likely. Yet, if you choose a natural duty to promote just institutions, then that will also make it more likely that you will fight for democracy in a dictatorship, and thus it will make it more likely that you will be killed by the dictator. Since, however, parties in the original position are constitutionally risk-averse, as it were, and moved by the worst case scenarios,<sup>80</sup> the prospect of freedom will not outweigh the prospect of death, and so, it seems, you will not choose the natural duty of justice. On the other hand, if the dictator does not get overthrown you might be killed anyway, and other people’s trying to overthrow him might make this *less* likely. So you have a reason to vote for the natural

---

<sup>80</sup> Rawls, *A Theory of Justice*, pp. 148-9.

duty of justice. Yet, on second thought, you might *be* the dictator, and so other peoples' trying to overthrow you might get you killed. Etc. The upshot is: without probabilities and without the strict compliance assumption, there is not enough data to make *any* decision. If, however, you allow the parties in the original position – throwing all of the Rawlsianism that was left overboard – to *know the actual probabilities* of their being in one position rather than in another and to know the actual changes the choice of certain principles will make to those probabilities, you will get completely counter-intuitive results and certainly not the results “cosmopolitans” and liberal egalitarians are looking for. In any case, the proposal of Moellendorf and Brock is simply incoherent.

I conclude that both Rawlsian “ideal theory” and Rawlsian “non-ideal theory” are useless. There are no proposals in sight that are capable of changing that and certainly none are to be expected. The fact that all that Rawlsians offer us are articles claiming that Rawls's theory *can* be applied to the real world but no articles that actually *do* apply it to the real world (in a logically conclusive way, not by means of *ad hoc* stipulations) is quite revealing in this respect. There is no reason to wait another forty years. Any moral theory relevant for the real world will not be Rawlsian. It is time to move on.

A final remark: does all this teach us something about (normative) ideal theory (as defined here) in general, not only about Rawls's ideal theory? Frankly, I am not particularly interested in this question, for the risk that in the absence of knock-down arguments against ideal theory *new* expansive ideal theories will mushroom is close to zero. The real danger is that *Rawls's* ideal theory will be pursued further (and still more time will be wasted), and it is this unappealing prospect that inspired this paper. However, for what it is worth, I think the above considerations clearly show (and of

course we already knew this) that limited ideal theory, that is, hypotheticals and thought experiments, are very useful for testing our intuitions and weeding out incorrect principles. They also strongly suggest, however, that all (normative) ideal theories are useless when it comes to directly telling us what to do, or what to regard as just, in the real world. Of course, I haven't *proven* that with respect to ideal theory in general (and I had not the slightest intention to do that: this paper is directed against Rawls). But the burden of proof is on the other side anyway (as Rawls's own remarks suggest), and the only way proponents of ideal theory can convincingly shoulder this burden is by *actually* using an ideal theory (again: in a logically conclusive way, not by means of *ad hoc* stipulations) for the purpose of directly telling us what to do in the real world. As already said, however, proponents of ideal theory simply do not do that. One wonders why.<sup>81</sup>

---

<sup>81</sup> I presented a first draft of this paper at the School of Law of the Torcuato di Tella University, Buenos Aires. I thank the audience for useful comments. I also thank Allen Buchanan, Holly Lawford-Smith, and Eduardo Rivera López for helpful written comments on previous drafts of this paper.