

**What is utopian about the realistic utopia?  
Relocating Rawls in the space of normative proposals**

*Christian Arnsperger*  
*FNRS & Chaire Hoover (Université catholique de Louvain)*  
*arnsperger@etes.ucl.ac.be*

November 2004

Forthcoming in *Revue Internationale de Philosophie*, Spring 2005,  
Special issue on Rawls edited by Catherine Audard

**1. Introduction**

John Rawls's emphasis on the distinction between ideal and non-ideal theory, his preoccupation with impartial justification and the strains of commitment, and his consecutive insistence on differentiating comprehensive and political moral doctrines all testify to his deep concern for offering a theory that, while being fully normative, takes account of the constraints of real-world implementation. This concern is most clearly present in his repeated claim to have constructed a "realistic utopia". The question is whether his theory has succeeded in providing a credible resolution of the tension between realism and utopianism.

This paper argues that it has not. It challenges the idea of a realistic utopia by taking it to what I believe to be its own logical conclusions, and deriving from those conclusions a call for relocating Rawls's justice as fairness (Rawls, 1971, 2001) within the space of philosophical theories of political institutions. As I shall argue, there is an uneasy coexistence between the theory's internal coherence as a fully-fledged normative proposal and Rawls's claim that it is practically implementable in the here-and-now alongside *other* equally fully-fledged normative proposals. Rawls's own theory therefore has to be subjected to a suitably revised version of the constraints he himself suggests in his *Political Liberalism* (Rawls, 1995).

I will build up my argument as follows. Section 2 dissects the internal logic of Rawls's "realistic utopia" project. Section 3 then formulates the basic challenge I address to Rawls from his own premises. Section 4 specifies that challenge further by arguing that it can be traced back to a structural problem within the liberal tradition. Finally, in section 5, I offer my own

suggestion for relocating Rawls's realistic utopia within the overall space of normative political proposals.

## 2. The structure of Rawls's "realistic utopia" project

Rawls repeatedly writes that his work is an endeavor to convey a realistic utopia to those whom he designates as "you and me, here and now". As its etymology indicates, a utopia provides a normative ideal of society that can effectively be "out of this world". Let us denote this utopian society by  $U^*$ . Being out of this world,  $U^*$  may have no connection whatsoever with the truly existing society  $S$  in which you and me live at the moment we read Rawls.

Or does it? Rawls complicates things by terming his utopia a "realistic" one. It is therefore two things at the same time: (a)  $U^*$  does not coincide with the currently existing society  $S$  since it has various otherworldly characteristics, but (b)  $U^*$  is in essential ways connected to "reality"—which means, inescapably, that in order to be intelligible for "you and me, here and now", it has to evidence features which are *also* present in *some society*  $S^*$  which "you and me, here and now" can call "real" while we are reading Rawls. The key issue is how distinct  $S^*$  is from  $S$ .

By "sharing certain key features", Rawls must mean the following:  $U^*$  is such that, all the way through the process of moral and political deliberation which leads to the principles of justice we never violate any feature  $\phi$  present in  $S^*$  whose violation would make the individuals' commitment to these principles in  $U^*$  exceedingly strained. The fact that what we want to avoid is the violation of commitment-related features *present in*  $S^*$ —and not merely in  $U^*$ —is what makes  $U^*$  a *realistic* utopia. The fact that we are concerned with the strains of commitment *in*  $U^*$ —and not merely in  $S^*$ —is what makes  $U^*$  a realistic *utopia*. Conjoining these two sides of the coin implies that in order for  $U^*$  to be this realistic utopia, the features  $\phi$  present in  $S^*$  have to be also present in  $U^*$ . Hence insisting on  $U^*$  being both realistic and a utopia unavoidably establishes a very close connection between  $U^*$  and the reference society  $S^*$  to which "you and me, here and now" can relate as being "real" as we are thinking about  $U^*$ . To make this plain, let me distinguish crucially between  $S$  as being *real* and  $S^*$  as being *real\**. If there is no difference between "reality" and "{real\*}ity", then  $S^*=S$  and  $U^*$  is no utopia at all. In general terms, the realistic utopia can be denoted as  $U^*(\phi, S^*)$ .

Rawls has always emphasized that the Original Position (henceforth OP) is a heuristic device not an empirical datum. This, of course, does not do away with the fact that  $U^*$  is a function of the pair  $(\phi, S^*)$ ; indeed, one may say that it is precisely because of this dependence that the OP can be viewed as a *heuristic* rather than an artificial or counterfactual device: “you and me, here and now” are able to *understand* the OP’s content to the extent that it is not disconnected from the pair  $(\phi, S^*)$ . What justifies the principles of justice that govern  $U^*$  is the fact that, according to Rawls, these principles can be shown to be part and parcel of a *reflective equilibrium*. This can be denoted as  $\rho = \{cc, pj, cj \mid I, \phi, S^*\}$ : the reflective equilibrium contains a conception of free and equal citizens ( $cc$ ), a set of principles of justice ( $pj$ ) and a set of considered judgments ( $cj$ ), and these three elements are adjusted to one another in such a way that none of the implications of  $pj$ , evaluated given the free and equal citizens’ objective information ( $I$ ) and {real\*}istic social reference point ( $S^*$ ), are incompatible either with any element of  $cj$  (a judgment itself “filtered” through the constraining features  $\phi$  which the set  $cj$  both depends on and implies) or with any element of  $cc$ . From the reflective equilibrium Rawls derives a theory of social justice in which the basic structure of social interaction  $S^*$  which is {real\*}istic in the sense that it is connected with various constraining features  $\phi$  that ensure that there are no strains of commitment in  $S^*$ .

Once we see the crucial role played by the peculiar notion of reality referred to by the adjective “real\*”, the issue of exactly how  $\phi, S^*$  and  $S$  are connected *in the minds of the Rawlsian citizens* becomes crucial. Are the  $\phi$ -connected constraints implied by “{real\*}ism” not such that  $U^*$  will be unable to differ from  $S$  in certain important ways in which free and equal citizens would like to see it differ? This poses the central question of how the search for a *realistic* utopia can leave any room for explicitly felt social dissatisfaction and for explicit social criticism on the part of rational and reasonable citizens.

The formalism of Rawls’s model imposes a twofold layer of abstraction. Between the currently existing society  $S$  and the OP, there is what can be called the *impartiality barrier* which corresponds to the part of our “sense of justice” that allows us to step outside of reality and enter a hypothetical, reflective position. There is a second barrier *within* the OP, one that can be called the *evaluation barrier*: it is the cognitive separation that allows us to evaluate implications and needed adjustments with reference to {real\*}ity rather than reality—that allows us, in other words, to freely play around with the various hypothetical parameters until we have adjusted

them into a coherent picture of what the {real\*} social world should be. So in fact,  $S^*$  is nothing more an *internally coherent* utopia.

But what are the *reasons* why anyone like “you and me, here and now” would overstep the impartiality barrier in the first place? The move into impartiality, we are aware, is merely instrumental to obtaining a compelling image of an alternative  $S^*$  to the currently existing  $S$ ; and why would “you and me, here and now” want to formulate such an alternative, were it not for the fact that we are *dissatisfied* with  $S$  and want to reflect on ways to *improve* society? However, given this dissatisfaction, is Rawls’s justificatory setup in any way plausible, let alone desirable, as a tool for formulating the alternative?

### 3. Formulating the challenge: Is Rawls an advocate of the *status quo*?

In this section, I shall express the challenge to Rawls’s “realistic utopia” project in its most extreme version. This will then allow us, in the next two sections, to see to what extent Rawls’s theory of justice needs to be saved or reinterpreted.

Clearly, the crucial point lies in the interdependence of  $S^*$ ,  $\phi$  and  $I$  in  $U^*$ . Rawls is adamant that the “objective” information  $I$  available to the free and equal citizens in the OP contains, in particular, general knowledge about human psychology and the general functioning of human societies. These pieces of information are those required to be able to track the implications of the principles of justice in any  $S^*$  from an impartial point of view. The main case in point is Rawls’s well-known discussion of the incentive constraints which, according to him, the Difference Principle needs to heed. Total economic equality, Rawls claims, may not be a reasonable distributive principle because due to incentive mechanisms, the richest and most productive members of society will end up making decisions that will harm the lifetime expectations of the least advantaged more than if some degree of inequality had been accepted. Of course, Rawls is not saying that total equality at a high level is *never* technically feasible—he is, however, suggesting that the cultural, psychological and social data that determine the degree to which it is feasible are not, themselves, part of the theory of justice: they are external parameters which, being present in  $I$ , will eventually shape the set of constraining features  $\phi$  which circumscribe the strains of commitment in a particular well-ordered society  $S^*$ .

The question, now, is what determines the content of the objective-information set  $I$  in the eye of the theorist. Is there really such a thing as social-structure-independent “facts of social life”; and if not, can whatever data one chooses to fix within  $I$  be thought up independently of a *pre-existing theory of society*? The answer, obviously, is no: in order to be able to gradually arrive at  $U^*$  through the complex seven-step algorithm sketched earlier, the theorist needs to assume that the “free and equal citizens” have endowed themselves with a theory of society, which gets deposited in set  $I$  in the form either of certain elements of information or of certain theoretical assertions about how a society works. For the time being—and most certainly in keeping with Rawls’s unanimity-oriented view of “free and equal citizens” in the OP—let us assume that one single theory of society,  $T$ , is shared by everyone so that  $I = \{T, \text{etc.}\}$ . While the citizens in the OP certainly evaluate all implications of the  $(cc, pj, cj)$ -complexes in an impartial way—i.e. not knowing who they might be in any {real\*} society  $S^*$ —they nevertheless draw up that evaluation using  $T$  and not some other theory of society. Therefore, the whole reflective process is contingent upon the social theory deposited in  $I$ . Hence,  $U^*$  itself is contingent upon a particular  $T$ .

Let us now go one step further and ask on what basis “you and me, here and now” are likely to choose our theory of society for  $I$  at the moment we overstep the impartiality barrier. Clearly, we won’t be able to wait for the result of our reflective process in order to know what social theory it will grind out for us, since having a social theory to begin with is a *pre-condition* to the whole reflective process itself. So we are in a bind—unless, of course, we simply take  $T$  to be a theory of what we already know and experience “here and now”, namely the currently existing social structure  $S$ . This is, in fact, the only experiential anchoring point we can have if it is assumed for the moment that we have not been taught any other theory of society. In that case,  $U^*$  and in particular  $S^*$  will be contingent on  $S$ —and this implies that the set  $\phi$  of constraining features congruent with  $U^*$  will be those already implicit in the currently existing social structure! For instance, whatever incentive mechanisms are considered normal in the currently existing capitalist structure will spill over as limits on what can be considered unjust or unacceptable in  $U^*$ . A Rawlsian may reply that  $T=S$  is not required by the justificatory setup. But then, what remains of the purportedly “realistic” character of Rawls’s utopia? We have to remember that up to this point  $U^*$  could only be called an internally consistent utopia; if the congruence of  $\phi$ ,  $I$  and

$S^*$  is to mean more than the mere absence of logical contradictions (which has never been sufficient for realism), i.e. if  $U^*$  is to be in addition a *realistic* utopia, then  $T=S$  is indeed a focal point that is difficult to shrug off.

True enough, none of this implies that  $pj$  will simply leave the existing structure  $S$  untouched. However, it will “touch”  $S$  only on its surface or boundary—that is,  $pj$  will be a set of *S-framing principles of justice*, a set of principles which may mitigate or counteract in an incentive-preserving way some of the systemic effects of  $S$ ; it can never be, however, a set of *S-revolutionizing* principles because the presence of the constraining features  $\phi$  in  $U^*$  would prevent it. The difficulty is made even more daunting if we acknowledge the fact that elements of  $T$  may constrain (perhaps even strongly) the latitude the free and equal citizens have in changing their representation of themselves in  $cc$ . An extreme case may be the one where the adoption of  $T$  fixes  $cc$  once and for all; and in that case, the restrictive character of the set  $\phi$  is likely to be even stronger, implying an ever more likely convergence of  $U^*$  on  $S$  itself if  $T=S$  as suggested a moment ago. Then  $pj$  may in some cases end-up not even *S-framing*, but downright *S-replicating*. This, in turn, means that if the constraints that keep  $U^*$  from being structurally different from  $S$  are that strict, the citizens who stepped into the OP with theory  $T$  in mind will obviously not be impelled to militate for a structurally different society once they have stepped back into  $S$ .

So here is the challenge: Rawls’s theory appears as a rather neat device for normatively consolidating the *status quo*—on one single sufficient condition: that all rational and reasonable citizens enter the OP with a social theory that is a clone of the existing social structure, or that is so able to obscure the problematic aspects of  $S$  that the reflective utopia  $U^*$  will keep on obscuring them. And how could it be otherwise in a Rawlsian framework since citizens are, by assumption, convinced that it is only *after* the reflective equilibrium has been established that they may have a yardstick to criticize the existing social structure—which, of course, they will not in the end, given the peculiar nature of the Rawlsian justificatory setup?

#### **4. Focusing the challenge: How dependent is the reflective equilibrium on social reality and on “unrealistic” social theory?**

That the above is not an empty claim is evidenced by the way in which Rawls conceives of social criticism in his Kantian framework: “Viewing the theory of justice as a whole, the 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 and held to be unjust to the extent that they depart from it *without sufficient reason*. (...) Thus *as far as circumstances permit*, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of the deviation from perfect justice. (...) If we have a reasonably clear picture of what is just, our considered convictions of justice may fall more closely into line even though we cannot formulate precisely how this greater convergence comes about. Thus while the principles of justice belong to the theory of an ideal state of affairs, they are generally relevant” (Rawls, 1971: 246, my italics).

The italicized conditionalities correspond to the features  $\phi$  of the real society  $S$  which make  $U^*$  unattainable. However, since  $U^*$  has been arrived at through impartial deliberation, this must mean that these constraining features of “realism” must have been available as information already in the OP. How could they be so available, if not in the form of empirical data or parts of social theories, i.e., conceptions endowed with a significant degree of abstraction or generality about how a society functions? We must realize that, had the free and equal citizens no such information *before* they start deliberating, they would be utterly unable to reach a reflective equilibrium since they would have to perform the impossible task of comparing various sets of principles of justice *for themselves*, without any reference to a society on which these various sets of principles can have distinct and variably desirable impacts. This, in turn, would mean that these citizens would actually, behind the veil of ignorance, have to design not only principles which are best for you and me, here and now, but in addition the very institutional and anthropological structures—all the way from the production and distribution of goods to the functioning of families—required for any human being to be able to know what “you and me, here and now” means.

In other words, if we were to assume that the citizens do not base their deliberation about the principles on certain pre-established general views of how society functions, the reflective equilibrium would have to be written as  $\rho = \{cc, pj, cj, \phi^* \mid \hat{I}\}$ , where the information set  $\hat{I}$  would now include only the most basic data about human nature—if any—that can be considered universally valid regardless of the individual’s social context. Consequently, the OP would now serve not only to stabilize principles, considered judgments and conceptions of citizenship, but

also to determine what are the *absolutely right strains of commitment*  $\phi^*$  given the most basic data about human nature—if any—that can be considered universally valid regardless of the individual’s social context. If realism could be confined to simply respecting the most basic, perhaps merely biological, aspects of human nature expressed in  $\hat{I}$ , such a modified theory of justice could indeed be legitimately termed a realistic utopia: one in which all but the most basic, a-social features of humanity—most notably, what can legitimately count as a strain of commitment—are themselves subjected to deliberation and are *made part of the principles of justice* (see Arnsperger, 1998).

Such a stripped-down OP, in which almost no empirical element is taken to constrain deliberation about what is just, corresponds to the classical idea of utopia. Many otherworldly designs in the utopian literature rely precisely on the fact that they lie outside of our *social* world, and that features of social interaction usually thought to embody “realistic” constraints are really only the products of an insufficiently daring imagination—with the underlying assumption that, once presented in a sufficiently suggestive way, the lack of realism could be converted into a message of genuine feasibility. Thus, making all but the most basic, biological features of humanity part of what a theory of justice has to shape would make the realistic-utopia project suddenly much more utopian than realistic: the distance from  $S^*$  to  $S$  would be significantly increased.

Clearly, Rawls is not attracted by this classical utopianism—quite probably because of his strong anchoring in both Kantian rationalism and Scottish empiricism (see Rawls, 2000) which, when combined, create a quite strong conviction as to why human reason and human passions cannot be tailored to fit just any social, let alone biological, design. Might this not, however, leave him with rather scant equipment to perform what has, after all, always one of the main implicit functions of utopian reflection—namely, social criticism? I want to claim it does not leave him unarmed for social *criticism*, but it does make him quite weak for *world-changing* social criticism because the link he establishes between social reality and social theory is a rather conservative one. And this is because he is reluctant to pull most of the features contained in the information set  $I$  inside the compound of things about which the theory of justice has something categorical—rather than merely hypothetical—to say.

The division between “perfect” theory and “imperfect” reality seems to suggest that Rawls locates his utopia mainly at the intellectual and not the practical level: whatever one’s considered

judgment about perfect justice, it can *never* be allowed to spark political action that would seek to overturn the “circumstances” of the real world. Perfect justice always needs to show deep respect for the imperfections of actual social life, for the human frailty inherent in all cooperation. This is a moral position whose nobility is strongly evidenced by the pathos of the last paragraphs of Rawls’s Introduction to *Political Liberalism*, and which I had myself strongly endorsed in earlier work on Rawlsian pluralism (see Arnsperger, 2000). It is, however, a position that leaves Rawls with little to say when the “circumstances” that make perfect justice unfeasible are deep power divides, structurally abusive social relationships and the like; in such circumstances, saying that we have a duty to correct injustices “if we can” is synonymous to saying that we have a duty towards ourselves to protect our lives—since social brutality or oppression quite clearly signal that “we can’t” change reality in our quality as Rawlsian citizens. In other words, the dimension of *social dissatisfaction*—in this case, social indignation and utopian aspiration as a tool for guiding vehement protest and (possibly violent) political action—seems to be much too absent in the way in which Rawls links the OP, the reflective equilibrium, and social reality.

But what does “much too absent” mean? A Rawlsian might shrug off the objection by saying that Rawls never saw his philosophical task as that of, say, a neo-Marxist thinker. This reply would only be adequate to the extent that the aspiration to world-changing social criticism could be seen as an exclusive province of Marxism—something which conservative philosophers from Hayek to Fukuyama have always had a political interest in claiming. If this claim were correct, then not only Hume but also Kant, Smith, and the whole liberal tradition would have to be classified as thinkers without any major aspiration to world-changing social criticism. Surely, no one would want to claim such a thing.

The question, however, is how a mixture of Kantian and Humean enlightenment can produce—and not only *intend* to produce—a political philosophy that criticizes the social world in such a way as to change it in any substantial way. And the answer, in a nutshell, is the following: These traditions already have, historically speaking, changed the social world immensely by affecting the very way in which liberal democracies have conceived their institutional design. A born-again Karl Marx would say that liberal democracy has undeniably been the dominant progressive force from the mid-1700s to the 1980s—it pulled societies out of the *ancien régime* and legitimized market interactions instead of inherited status, as well as

constitutional democracy instead of autocratic rule. During that period, political liberalism and social-democratic ideals had strong utopian force, and Rawls can be seen as having taken up *ex post* that force and made it into a theory of justice that legitimizes that history using the best, humanistic conceptions of democratic citizenship that arose during that 230-year period. It seems to me, however, that Rawls's justice as fairness, in the way it is constructed, has little or no *pro-active character*. In fact, being mainly a theory of what had already accumulated, it simply arrived too late to affect the future, as evidenced by at least to events which took place in the wake of the publication of the *Theory of Justice*: first, barely eight years later, the crackdown on social-democratic ideals with the simultaneous election of Ronald Reagan and Margaret Thatcher in the very Anglo-Saxon sphere in which Rawls has always been claimed to have had the most impact; and second, another ten years later, Fukuyama's *The End of History* which claims that in effect, the whole market-liberal tradition had now become a tool for managing and justifying an essentially unchanging set of institutions. Reality and {real\*}ity had finally merged, so to speak.

Now, a defender of Rawls can always get out of the quagmire by saying that such an end of liberal-democratic history is not at all what Rawls's theory aims at. This might be correct, but it simply confirms that the actual course of history has now overtaken Rawls's theory as it stands: it cannot help you and me, here and now, as a tool to understand what we have a duty to do *from now on*—except, of course, if you and me and all of us unanimously share the mixture of Kantian and Humean philosophy that would make us adhere to the market economy and the institutions of liberal democracy as embodiments of the “right” theory of society. But if that were so, we would all by assumption be in agreement with Fukuyama's verdict—and, not surprisingly, we would tend to confine our ambitions for social reform to something like Anthony Giddens's (1998) “Third Way”, which merely repeats in a new guise a set of ideas taken from the two-century old liberal tradition—ideas which were once deeply subversive but which are now nothing new, and which divert our attention to what is still wrong with market democracy: the false universalism that hides beneath Kantian impartiality, and the genuine inequality of power that hides beneath supposedly “civilizing” Humean or Smithian market interactions.

Rawls's justificatory device is not up to the task of uncovering these hidden flaws because the citizens become critical of society only *after* they have been in the OP and have made up their minds “impartially”, and because the “impartial” information with which he endows his free and equal citizens in the OP is not itself already anchored in a critically-minded dissatisfaction with

the existing society. If you are not critically-minded before you start reflecting on your society, you are likely not to be critically-minded at the end of the process because you will not have been looking for those theoretical ideas that could rationalize and articulate your dissatisfaction with society (see, e.g., Geuss, 1981). For reasons that have to do with Rawls's deep commitment to the victories of older liberals over obscurantist conceptions, his free and equal citizens are *constructed in such a way* that they already live in a society which they have no strong reason to be dissatisfied with.

If instead—and remaining true to Rawls's own *most fundamental* ethical gesture—we were to conceive citizens as crossing the “impartiality barrier” because they want to make sense of injustice they have already theorized and want to combat, the outcome of the Rawlsian deliberation process would be likely to be quite different. Citizens would, for instance, have several theories as to what incentive claims on the part of the currently rich and powerful can be accepted as legitimate, but contrary to what Cohen (1992) suggests in his model of the “justificatory community”, they would not all *already* base their critical discussions upon a unanimous agreement with Rawls's difference principle—that principle would be only one outcome among many others of the discussion among citizens initially endowed *in the OP* with different, and possibly opposed, critical theories of society, i.e., theories *T, T', T''*, etc. that will differ strongly from *S* so as to seem to some to be “unrealistic”, and whose confrontation will be part of the deliberation in the OP. I believe that Rawls actually sensed the necessity for such a modified framework when he modified his own theory and went from *A Theory of Justice* to *Political Liberalism*, replacing comprehensive liberalism with a more strictly political conception of liberalism. However, it seems to me he did not quite spell out all the consequences of such a move.

## **5. Citizens' theory choice as an issue of reasoned solidarity**

The reason why Rawls failed to spell out these consequences is that he remained trapped in the idea that there were some fundamental features of “how a society works” which none of the comprehensive doctrines facing each other in the OP would question—features such as the organization of social cooperation through markets which (as Cohen, 1992, correctly points out)

Rawls does not criticize with a view to how they restrict the possibilities for redistribution and solidarity. *But what conception of solidarity can the free and equal citizens invoke in the OP, since at that point in the deliberation they have no critical theory of what is wrong with current society?* In fact, it is Cohen's own Marxist past which allows him to feel moral indignation at the idea that, in some market societies where the rich have significant social and economic power, the difference principle might be compatible with very little redistribution. Rawls, who sees the incentive constraints as part of the objective information reasonable citizens are endowed with in the OP, cannot see the problem—and as I suggested in section 3, he cannot see the problem because, as a Kantian-Humean liberal, he simply shies away from assuming that reasonable citizens could harbor any social theory *T* that does not rationalize private property and market interactions.

However, if this is so, Cohen's suggestion of implementing a "justificatory community" outside of the OP will be of no avail: if all citizens really do share market-oriented conceptions of how society works and should work, even the poor and weak will view their situation as unfortunate but justified and might, among other things, flee into religion (as does Cohen, 2000, in the epilogue to his book)—while the rich and powerful, while perhaps dimly aware that on some interpretations the Gospels do indeed condemn their wealth, can nevertheless *de facto* get away with whatever incentive arguments they want to invoke. The only significant way of making the picture less bleak is to accept that each citizen, or at least each of a number of citizen groups, holds a distinct theory as to what is wrong with society *when entering the OP*, uses it to diagnose what ought to be changed, and is prepared to act upon that theory outside of the OP, too.

This alone, however, would suffice to jeopardize the whole stability of Rawls's political anthropology: there would be one reflective equilibrium for each citizen or each category of citizen—something which smacks in the face of the contractualist idea that citizens enter the OP seeking an *agreement* on *shared* principles of justice. The cognitive and conceptual difficulties would indeed be immense, since in evaluating the implications of his personal principles of justice each free and equal citizen would have to effectively envision the just society *S\** itself as a game in which all kinds of conceptions of social justice including his own interact, clash, negotiate, and so on—a very true-to-life picture of a lively democratic society, but one which

would induce such an amount of indeterminacy that the theory would simply collapse from a logical point of view.

The other option, therefore, would be to continue assuming that all free and equal citizens meet in the OP carrying one identical social theory. But then, we know that there will be (at the very best) one {real\*} utopia  $U^*$  for each such identical social theory: another instance of indeterminacy which, moreover, would be utterly incapable of explaining, from the cognitive point of view, how people from widely differing walks of life and social positions in  $S$  would coordinate in the OP on any social theory other than  $S$  itself—which, at least, has the advantage of being the theory rationalizing the current system, and therefore of being equally accessible to all equally alienated participants in  $S$ ... Indeed, how would the state and the educational system in  $S$ , when teaching individuals how to be impartial, manage to teach them a theory of the social structure that would clash profoundly with the existing rationale for the prevailing mechanisms of interaction in  $S$  itself?

Therefore, the OP can hold up neither under the *radically pluralistic* assumption that citizens enter it endowed with a variety of social theories nor under the *critically monist* assumption that all citizens have a single social theory that differs from  $T=S$ . Now, any normative analysis of modern society has to defend pluralism as an expression of cognitive and reflective freedom of the individuals, so what alternatives are there? Since there is no straightforward way of linking impartiality with citizens holding particular theories, and since it is a social fact that there exists a plurality of social theories, the domain of theory choice is, in a sense, “up for grabs”.

Here is my own suggestion about how one ought to choose the critical social theory you would want to disseminate among as many people as possible. I believe one should choose one's social theory on the basis of *reasoned solidarity*. This is not quite the same thing as impartiality, although it keeps some aspects of it: reasoned solidarity is an attitude whereby one scans the currently existing social world for tangible instances of *abusive social power impeding certain individuals' or groups' opportunities for free self-realization*. It is actually possible to write down an analytical theory of reasoned solidarity, and it has been done recently (see Arnsperger and Varoufakis, 2003). Such a scan of social reality must combine emotional engagement and clear-headed empathy (see Livet, 2002); its result can be reasonable only if one does not blindly follow one's own inclinations, revolts and dislikes (or one's mere communitarian sense of belonging, as

in Walzer, 1988) but rather makes use of all tools of historical, sociological and political-economic knowledge-gathering. It will not be impartial knowledge in the sense of Rawls's "objective-information" set  $I$  in the OP; but neither will it be a mere reflection of your inborn impulses or of your personal antipathies. Thus, reasoned solidarity provides a crucial realistic alternative to Rawls's realism: instead of focusing on supposedly insuperable and quasi-natural incentive constraints, it focuses on structural power and wealth unbalances within  $S$ , and these data serve as the anchoring of  $S^*$  in  $S$ , while pointing towards what has to be overcome through political action so as to reach  $U^*$ .

Viewed from this angle, Rawls's justice as fairness is a very particular case, giving market-related incentives and formally democratic rights logical precedence over, say, redress of asymmetries in economic power or radical economic egalitarianism. Rawls's proposal is, of course, not at all incoherent: it stands as it is and can be embraced on a specific conception of liberal impartiality—but it can hardly be used to disqualify other, liberal or less liberal, but nevertheless reasonable conceptions of what impartiality and solidarity command.

Once one has decided on the basis of reasoned solidarity which philosophical theory of political institutions to convey to whomever wants to listen or act along, there will be no choice but to enter the political arena and its processes of struggle, negotiation and possible defeat. In such a political world of constant and unstable compromise (see Arnsperger and Picavet, 2004), how philosophical doctrines change the world is less clear than when we assume unanimous adherence to some theory of justice; Rawls's justice as fairness can be helpful tools for self-clarification and focus within a subset of citizens when they interact with another subset of a different persuasion, but each of these tools will always remain partial—and I believe, in Rawls's case, nowadays largely ineffectual—in a real-time social process which, though in no sense dialectical, radically escapes the grasp of any one philosophical theory of political institutions. Thus, Rawls's theory of justice as fairness is relocated within social and political space as one tool among others—along with other political-liberal theories, but also along with very different ones, each with its own conception of impartiality and, hence, of reasoned solidarity.

But how, then, are these various *political* doctrines to coexist in the actual, empirical field of political *practice*? I believe that at that meta-political level the only viable solution lies somewhere between Habermas's strongly procedural conception of democracy, in which various normative conceptions of politics have to provisionally agree on norms of rational *political* (not

just moral or infra-political) disagreement, and a revised version of Rawls's political liberalism in which the *political* coexistence of *comprehensive moral* doctrines is replaced by the *meta-political* coexistence of *political-liberal and other political* doctrines. It may well be that—somewhat along the lines of his last attempts in *The Law of Peoples* (Rawls, 1999)—an OP can be conceived of where the holders of various political utopias, some liberal in Rawls's own sense and some not, some realistic in Rawls's own sense and others less so, can get together and devise rules of meta-political coexistence. It may even be that—contrary to what Rawls claims for an OP between peoples—such varied utopias can settle on a revised version of Rawls's principles as those that could optimally regulate *political disagreement* and ensure an equal participation of all political philosophies in the deliberation on the conditions of their coexistence. Indeed, this seems quite possible: the principle of Equal Liberties would now ensure that all circulating political utopias—Rawls's justice as fairness being merely one of them—would legally have an equal and maximal access to the public sphere; the principle of Fair Equality of Opportunity (where “fair” may no longer be an adequate adjective) would now ensure that all utopias get taught and disseminated on equal footing and that no holder of a political utopia could be banned if he—contrary to the holders of an anti-democratic or fascist utopia—accepted the same conditions of equal access for all others; and the Difference Principle would now ensure that the real (as opposed to merely legal) opportunities for all political utopias of having at least a minimal weight in the public discussion over time, and of occasionally tilting the political compromise in their own favor, is as high as possible.

Perhaps this is, in the end, what *Political Liberalism* was aiming at. But if so, its focus on consensus between comprehensive moral doctrines, rather than on disagreement and compromise between fully-fledged political utopias, was maladjusted due to Rawls's antiquated preoccupation with “realism” as reflecting and protecting the core economic and democratic values of eighteenth- and nineteenth-century liberalism. Unless, like most conservatives, we remain obsessed with communist and fascist totalitarianism as the *only* alternatives to the competitive market economy and to private-property-protecting representative democracy inherited from the German and Scottish enlightenments, there is no reason why the scope of “realism” in utopian reflection cannot, and should not, be made significantly than what Rawls warrants.

## References

- Arnsperger, Ch. (1998), “John Rawls et l’engagement moral”, *Revue de métaphysique et de morale* **102**: 237-57.
- Arnsperger, Ch. (2000), “Au-delà de la raison et des pouvoirs: Le fondement ‘trans-raisonnable’ de la raison libérale”, *Revue philosophique de Louvain* **98**: 83-106.
- Arnsperger, Ch. and E. Picavet (2004), “More Than *Modus Vivendi*, Less Than Overlapping Consensus: Toward a Political Theory of Social Compromise”, *Social Science Information* **43**: 167-204.
- Arnsperger, Ch. and Y. Varoufakis (2003), “Toward a Theory of Solidarity”, *Erkenntnis* **59**: 155-188.
- Cohen, G.A. (1992), “Equality, Incentives, and Community”, in G.B. Petersen (ed.), *The Tanner Lecture on Human Values XIII*: 262-329, Salt Lake City, University of Utah Press.
- Cohen, G.A. (2000), *If You’re an Egalitarian, How Come You’re so Rich?*, Cambridge, Harvard University Press.
- Geuss, R. (1981), *The Idea of a Critical Theory*, Cambridge, Cambridge University Press.
- Giddens, A. (1998), *The Third Way*, Cambridge, Polity.
- Livet, P. (2002), *Émotions et rationalité morale*, Paris, Presses Universitaires de France.
- Rawls, J. (1971), *A Theory of Justice*, Cambridge, Harvard University Press.
- Rawls, J. (1995), *Political Liberalism*, New York, Columbia University Press.
- Rawls, J. (1999), *The Law of Peoples*, Cambridge, Harvard University Press.
- Rawls, J. (2000), *Lectures on the History of Moral Philosophy*, Cambridge, Harvard University Press.
- Rawls, J. (2001), *Justice as Fairness: A Restatement*, Cambridge, Harvard University Press.
- Walzer, M. (1988), *Interpretation and Social Criticism*, Cambridge, Harvard University Press.