

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA**

**IS THE RAWLSIAN UTOPIA OVERLY  
INTELLECTUAL?**

**ARTUR COMIRAN TONON**

Porto Alegre (RS)

2022

**ARTUR COMIRAN TONON**

**IS THE RAWLSIAN UTOPIA OVERLY  
INTELLECTUAL?**

Dissertação submetida ao Programa de Pós-Graduação em Filosofia da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção de título de Mestre em Filosofia.

Orientador: Dr. Paulo Baptista Caruso MacDonald

Porto Alegre (RS)

2022

### CIP - Catalogação na Publicação

Comiran Tonon, Artur  
Is the Rawlsian Utopia Overly Intellectual? / Artur  
Comiran Tonon. -- 2022.  
87 f.  
Orientador: Paulo Baptista Caruso MacDonald.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Instituto de Filosofia e Ciências  
Humanas, Programa de Pós-Graduação em Filosofia, Porto  
Alegre, BR-RS, 2022.

1. Political Liberalism. 2. Ideal Theory. 3.  
Reflective Equilibrium. 4. Realistic Utopia. 5.  
Overintellectualization. I. Baptista Caruso MacDonald,  
Paulo, orient. II. Título.



**ATA PARA ASSINATURA Nº \_\_\_\_\_**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
Instituto de Filosofia e Ciências Humanas

Programa de Pós-Graduação em Filosofia  
FILOSOFIA - Mestrado Acadêmico  
Ata de defesa de Dissertação

Aluno: Artur Comiran Tonon, com ingresso em 04/03/2020  
Título: **IS THE RAWLSIAN UTOPIA OVERLY INTELLECTUAL?**  
Orientador: Prof. Dr. Paulo Baptista Caruso Macdonald

Data: 08/11/2022  
Horário: 14:00  
Local: IFCH/UFRGS

<b>Banca Examinadora</b>	<b>Origem</b>
Nikolay Steffens Martins	UFRGS
Mariana Kuhn de Oliveira	UFRGS
Denilson Luis Werle	UFSC

Porto Alegre, 08 de novembro de 2022

<b>Membros</b>	<b>Assinatura</b>	<b>Avaliação</b>	<b>Indicação de Voto de Louvor</b>
Nikolay Steffens Martins	_____	aprovado	_____
Mariana Kuhn de Oliveira	_____	aprovado	_____
Denilson Luis Werle	_____	aprovado	_____

aprovado

Conceito Geral da Banca: ( ) Correções solicitadas: ( ) Sim ( X ) Não  
Indicação de Voto de Louvor: ( ) Sim ( ) Não

**Observação:** Esta Ata não pode ser considerada como instrumento final do processo de concessão de título ao aluno.

\_\_\_\_\_  
Aluno

\_\_\_\_\_  
Orientador

Programa de Pós-Graduação em Filosofia  
Av. Bento Gonçalves, 9500 Prédio 43322 - 205D - Bairro Agronomia - Telefone 33088220  
Porto Alegre - RS

*Nobody has ever lived without daydreams, but it is a question of knowing them deeper and deeper and in this way keeping them trained unerringly, usefully, on what is right. Let the daydreams grow even fuller, since this means they are enriching themselves around the sober glance; not in the sense of clogging, but of becoming clear. Not in the sense of merely contemplative reason which takes things as they are and as they stand, but of participating reason which takes them as they go, and therefore also as they could go better. Then let the daydreams grow really fuller, that is, clearer, less random, more familiar, more clearly understood and more mediated with the course of things. So that the wheat which is trying to ripen can be encouraged to grow and be harvested.*

– Ernst Bloch, *The Principle Hope*

## RESUMO

TONON, ARTUR. **Is the Rawlsian utopia overly intellectual?** 2022, Dissertação (Mestre em Filosofia) – Instituto de Filosofia e Ciências Humanas – Universidade Federal do Rio Grande do Sul.

Este trabalho visa analisar as exigências intelectuais do ideal da sociedade bem-ordenada de John Rawls. Parte-se do princípio de que um ideal de sociedade visto como um objeto a ser produzido no mundo deve ser tal que desejemos produzi-lo. São consideradas duas objeções que colocam em questão a desejabilidade do ideal de sociedade rawlseano por considerá-lo ‘hiperintelectualizado’. A primeira objeção questiona os critérios estabelecidos por Rawls para que uma doutrina abrangente seja considerada razoável, apontando que, ao incluir critérios racionalistas, Rawls exclui arbitrariamente visões de mundo plenamente compatíveis com a cooperação entre cidadãos livres e iguais. A segunda objeção rejeita a proposição rawlseana de que um estado de equilíbrio reflexivo amplo e geral ocorre na sociedade bem-ordenada, argumentando que tal estado exigiria que os cidadãos em geral tivessem disposições de desenvolver habilidades teóricas características de um filósofo. Considera-se, ao final, possíveis caminhos para evitar a hiperintelectualização. De um lado, argumenta-se que doutrinas abrangentes razoáveis são melhor compreendidas em termos estritamente éticos, bem como é salientado o papel de doutrinas abrangentes menos articuladas no modelo de formação do consenso sobreposto de Rawls. De outro lado, é esboçada uma noção alternativa de aceitação pública que não recorre a um estado de equilíbrio reflexivo pleno, mas cuja caracterização envolve a idealização de um tal estado, mesmo que este último não deva ser produzido no mundo.

**Palavras-chave:** liberalismo político, teoria ideal, equilíbrio reflexivo, hiperintelectualização.

## ABSTRACT

In the present study I aim to assess the intellectual conditions of John Rawls' ideal of a well-ordered society. I assume that such an ideal of society, seen as a conception of an object to be produced, must be one that we desire to produce. I consider two objections that put into question the desirability of the Rawlsian ideal of society as they consider it 'overintellectualized.' The first objection questions whether Rawls' criteria for a comprehensive doctrine to be reasonable are too rationalistic, resulting in the arbitrarily exclusion of views that are perfectly reasonable in an ethical sense. The second objection rejects the Rawlsian statement that a state of general and wide reflective equilibrium is supposed to obtain in the well-ordered society, arguing that such a state would require citizens in general to have dispositions to develop theoretical virtues characteristic of those of a philosopher. I consider, at the end, possible ways out of the overintellectualization charges. On the one hand, I argue that reasonable comprehensive doctrines should be understood strictly in ethical terms, and I emphasize the role of partially comprehensive doctrines in Rawls' model of the formation of the overlapping consensus. On the other hand, I sketch an alternative notion of public acceptance, which do not appeal to a state of full reflective equilibrium, but whose characterization involves an idealization of such a state, even if we should not aim to produce it in this world.

**Keywords:** political liberalism, ideal theory, reflective equilibrium, overintellectualization.

# Contents

<b>INTRODUCTION</b>	<b>9</b>
<b>1. PRELIMINARY REMARKS</b>	<b>12</b>
1.1 Ideal theory	12
1.2 Full publicity and autonomous acceptance	15
1.3 Rawls' justification programme and autonomous acceptance	20
1.4 Stability and overlapping consensus	25
1.5 Overlapping consensus and the three kinds of justification	30
1.6 The role of reflective equilibrium in theory and practice	39
<b>2. OVERINTELLECTUALIZATION</b>	<b>42</b>
2.1 Reasonable comprehensive doctrines as exercises of reason	43
2.2 The philosopher's dream of full reflective equilibrium	50
2.3 Am I missing the point?	60
<b>3. WAYS OUT</b>	<b>63</b>
3.1 The ethical account of reasonable pluralism	64
3.2 The dignity of partially comprehensive doctrines	67
3.3 A promising picture of public acceptance	71
3.4 What is left for the justification programme?	78
<b>CONCLUDING REMARKS</b>	<b>81</b>
<b>REFERENCES</b>	<b>84</b>



## INTRODUCTION

No matter how sophisticated one's interpretation of Rawls' theoretical apparatus is, one who forgets that it serves a practical purpose is not getting it right. Practical reason, says Rawls, is not merely about the "knowledge of given objects" but mainly about "the production of objects according to a conception of those objects" (PL, 93<sup>1</sup>). The object of interest in Rawls' philosophy is an ideal of society, or what he calls a 'realistic utopia'<sup>2</sup>. 'Utopian' and 'realistic' both connect to his practical aspiration: on the one hand, a conception that aims to guide social change cannot be a faithful portrait of reality – it must idealize, picture a place that doesn't exist anywhere, a utopia. On the other hand, that exercise of imagination would not have any practical significance if it weren't possible and desirable, given basic facts about us and our social life, that such a society came about through our political efforts. A perfectly just society that can only exist in our dreams is nothing but wishful thinking; even worse, a perfectly just society that would become a nightmare if it ever came out of our dreams would be at best a realistic dystopia.

The well-ordered society – Rawls utopia – is a society governed by a public conception of justice, that is, a society in which "everyone accepts and knows that others accept the same principles of justice" (TJ, 4<sup>3</sup>). Publicity, however, entails not just any acceptance. First, citizens must share reasons for acceptance. Second, it should neither be forced nor the result of indoctrination; citizens are to accept the conception because they are responsive to those shared reasons. Third, those shared reasons cannot be any considerations that we happen to share, rather they are supposed to be right reasons. One might say, in the well-ordered society justification of a conception of justice is intertwined with public acceptance for shared right reasons. Yet acceptance goes even further. Besides acknowledging the shared reasons that make a political conception of justice acceptable to all, citizens of the well-ordered society also accept the ideals of that conception from their own fundamental, and not publicly shareable, view of life – in Rawls' terms, their reasonable comprehensive doctrines. This way, citizens of the well-ordered society have the strongest possible motivation to uphold the terms of social cooperation.

---

<sup>1</sup> *Political Liberalism: Expanded Edition*, 2005.

<sup>2</sup> For the notion of a realistic utopia, see Rawls (1999b; 6-7, 11-23).

<sup>3</sup> *A Theory of Justice: Revised Edition*, 1999.

That stringent form of public acceptance with shared reasons connects with a deep liberal commitment: that each citizen should be able to find good reasons to accept political authority. Now, more than having good reasons, members of the well-ordered society are supposed to effectively judge and act in response to the right reasons to submit to political authority. Therefore, let us say, public acceptance in the well-ordered society yields general autonomous acceptance.

However, I assume that an attractive ideal of a just society must also avoid taxing citizens in general with too many intellectual requirements. There are two points about the well-ordered society that might be taken as red flags of overintellectualization; namely, (i) Rawls' suggestion (in some passages at least) that reasonable comprehensive doctrines are exercises of reason, and (ii) his suggestion that in the well-ordered society a state of wide and general reflective equilibrium occurs. In fact, those seem so intellectually demanding that I am inclined to question, as Bernard Williams (2015) did, whether the well-ordered society "would be held together under that degree of self-conscious ethical understanding, and for what actual societies it represents a reasonable aspiration" (113). I want to unpack those tentative remarks and apply them to Rawls' later developments, especially in *Political Liberalism*.

This is how I approach the matters: In chapter 1, I explain the part of Rawls' machinery, covering (a) his conception of ideal theory and what I call the practical condition, which requires ideals of society (if understood as realistic utopias) to be desirable states; (b) the publicity condition and its connection with the idea of respecting citizens' autonomy; (c) Rawls' justification programme and its possible interpretation as a key to autonomous acceptance; and (d) an image of a well-ordered society in which an overlapping consensus obtains accompanied by a state of general and wide reflective equilibrium. In chapter 2, I lay down two "overintellectualization objections"<sup>4</sup>, one concerning the characterization of a reasonable comprehensive doctrine, while the other questions the description of a well-ordered society as a state of general and wide reflective equilibrium. Both claim that political liberalism pictures an overly intellectual society and consequently fails to meet the practical condition. Lastly, in chapter 3, I consider possible ways out to each of those objections.

In response to the first objection, I outline a less demanding account of reasonable comprehensive doctrines. First, I claim that reasonable doctrines should be defined simply as

---

<sup>4</sup> I take the term, as well as some important insights (especially for the second objection and the proposed way out of it), from Lord (2017). His work addresses a contemporary debate concerning what it is to be responsible for acting in a certain way, or what it is to act *from* (and not merely in accordance with) a reason.

doctrines that reasonable persons affirm; that is, doctrines that can be affirmed without jeopardizing reasonable persons' motive of reciprocity in their political relation. Second, I show that a less intellectual kind of doctrine – the partially comprehensive – have a prominent role in Rawls' political liberalism.

Against the second objection, I begin by sketching an alternative interpretation of the notion of acceptance. I argue that acceptance may be understood in a more dispositional way. That is, as a disposition to judge and act competently in circumstances where the demands of the political conception apply, together with an attitude of opening to others' reasoning when facing disagreement or responsibility of their various social positions. What such a notion does not demand is either the endorsement of explicitly stated principles, or the far-reaching process of self-assessment and reflection that characterizes the search for wide reflective equilibrium. Then I depict what seems to be a deeply satisfactory ideal of society in which full reflective equilibrium never obtains.

Finally, I consider aftermath of the discussion regarding Rawls' justification programme. I maintain that, even though one cannot expect nor wish to obtain anything close to an endpoint of such a programme in this world, it serves a key role in the construction of a less intellectual utopia. That is because, I argue, one can only be sure that everyone is *able* to find justification of their political institutions if one imagines what it would be like if everyone engaged in the search for wide reflective equilibrium. Despite such a "theoretical" role, though, the justification programme is also continuously carried out at various levels of abstraction by citizens' responsible discussion in actual practices of justification.

Discussions about political liberalism seem idle if one forgets, even for a moment, that it is a practical endeavor. That is my main concern in this work. Thus, I aim to provide an attractive picture of what the realization of justice as fairness in this world would look like. What I offer is just an outline of it, not a complete description, but I hope that it proves at least an interesting take on Rawls' project.

## **1. PRELIMINARY REMARKS**

It goes without saying that I owe a lot of clarification and context. There are key ideas of Rawls' project embedded in those questions, each of which is subject to a great deal of philosophical debate. There is no way other than taking a detour and explaining those ideas. My aim in this chapter is to assemble some pieces and build a picture of what the realization of Rawls' utopia would look like, in order to judge whether it "represents a reasonable aspiration," considering its intellectual demands. For the most part, I will limit myself to indicate points where Rawls' claims are subject to criticism, but, as far as my reasoning goes, I will try to avoid relying on controversial assumptions of my own. Let me begin, then, with Rawls' idea of ideal theory.

### **1.1 Ideal theory**

Ideal theorizing is one of the most common sources of uneasiness concerning Rawls' work. Besides the various philosophical critiques<sup>5</sup>, there is a source of uneasiness that hardly any reader fails to experience: people concerned about justice normally want to engage in, or at least provide tools for political action that aims to fight existing injustice. Now, Rawls' ideal theory arises from a division of tasks between theoretical approaches. Ideal theory is "strict compliance theory," in which "everyone is presumed to act justly and to do [their] part in upholding just institutions," while non-ideal theory is "partial compliance theory," which "studies the principles that govern how we are to deal with injustice" (TJ, 8). Thus, Rawls' discussion, restricted as it is to ideal theory, weeds out from the start what motivates most

---

<sup>5</sup> Artillery against Rawls' ideal theory is of various calibers. For instance, Mills (2005) objects that its abstraction blinds us to actual structures of domination, and ends up helping in their promotion – he sees it as ideology, in the pejorative sense. Appiah (2017; 123-126) questions whether aiming at principles to govern a well-ordered society is aiming too high; perhaps, he goes, aiming lower is more likely to yield better results. Sen (2008; 336-340) claims that working out a picture of an ideally just society is of little practical help, since making comparative judgements about the relative justice of existing societies does not presuppose that know what an ideally just society would look like, any more than we need to identify the highest peak on earth in order to judge the comparative height of two small mountains. Besides, he claims that Rawls has a wrong moral epistemology, since we learn about justice, most of the time, confronting injustice, not thinking about it in the abstract. Each of these objections poses a challenge to Rawls' account of ideal theory, but I cannot address those objections here. For a defense of ideal theory, see Simmons (2010).

people's interest in theorizing about justice, namely, existing injustice. However, the exclusion is not gratuitous. Ideal theory does not abstract from injustice merely hoping that one day it ceases to exist. Rather, according to Rawls it serves to provide an ideal of society that informs the way we, here and now, should deal with the problems of partial compliance theory; in fact, he says, it "provides (...) the only basis for the systematic grasp of these more pressing problems" (*ibid.*)<sup>6</sup>.

There are different versions of ideal theories in that sense. We may imagine different moral outlooks (restricted to political justice or not) assuming universal compliance. One needs only to take their preferred theory and consider its results as if it were generally complied with. Rawls' version of it gains contours in the form of a well-ordered society, defined as a society "effectively regulated by a public conception of justice" (JAF, 8<sup>7</sup>). The notion of a well-ordered society is extremely complex and involves a number of conditions; I will focus specifically on publicity. By publicity, we saw, Rawls means that each member of the society accepts and knows that others also accept the same conception of justice, that fact being publicly known.

Publicity is not implied by the notion of strict compliance; there is no logical contradiction in a strict compliance theory that violates such a condition. For instance, one may picture in theory an ideal of society governed by, say, the average utility principle, which is universally complied with, "and yet this fact not be widely known or explicitly recognized" (TJ, 115). It becomes clear, then, that Rawls' what distinguishes ideal theory in Rawls' sense (the strict compliance assumption) is by no means the only idealization involved in his project. The question I pursue in this work largely concerns Rawls' idealization of publicity. In actual societies, people hardly accept the same principles of justice, whatever the conditions for such acceptance are; let alone share a public commitment to the same principles. When elaborating his conception, Rawls idealizes away such facts; in their place, he conceives citizens who endorse each the same conception for the right reasons, and who know that others do the same.

In assessing the merits of Rawls' idealization, I take seriously his claim that political

---

<sup>6</sup> Whether ideal theory should have the priority Rawls assigns to it is a matter of philosophical controversy. At some points, Rawls claims that, in the absence of ideal theory, "there is no rational basis for continually adjusting the social process so as to preserve background justice, nor for eliminating existing injustice" (PL, 285). That claim is stronger than the one quoted in the text body. For it does not follow from the fact that ideal theory provides the only basis for a *systematic* grasp of non-ideal theory that ideal theory provides the only *rational basis* for dealing with non-ideal theory. The latter would require the very controversial claim that the continuous effort to tackle injustice lacks rationality if one does not have the systematic grasp produced by ideal theory. Although I believe that ideal theory can dispense with the later claim, I will not pursue the issue further.

<sup>7</sup> *Justice as Fairness: A Restatement*, 2001.

liberalism is an exercise of practical reason, that is, it concerns not merely the knowledge of given objects, but primarily the “production of certain objects in accordance with a conception of those objects” (PL, 93) . The relevant object here is an ideal of society, or a realistic utopia, which contains the satisfaction of the publicity condition. And ‘production,’ obviously, is meant to be in this world. Taking Rawls by his word,

There is no necessity to invoke theological or metaphysical doctrines to support its principles [those of justice as fairness], nor to imagine another world that compensates for and corrects the inequalities which the two principles permit in this one. Conceptions of justice must be justified by the conditions of our life as we know it or not at all (TJ, 398).

Connecting Rawls’ ideal theory with the practical character of his philosophy is a step towards defusing the mentioned ‘uneasiness’ that almost everyone who reads him feels. However, it also imposes a burden on liberals of a Rawlsian influence, which often use ideal theory as a crutch to avoid committing to anything of practical importance. Such an avoidance, I think, tends to make ideal political philosophy more of an obsession with purity<sup>8</sup>. Therefore, I draw the following practical condition for ideal theories (as long as they claim to be exercises of practical reason):

**Practical condition:** an ideal conception of justice must be rejected in case its full realization in this world is undesirable<sup>9</sup>.

---

<sup>8</sup> A possible criticism of ideal theory is related to this. One might resist ideal theory not because it is incompatible with nonideal theory, but because too much time and effort is spent pursuing the first in detriment of the second.

<sup>9</sup> One might want to add that it must also be possible to realize such an ideal society in this world, however that would bring unnecessary difficulties. Even though Rawls claims that a realistic utopia must establish that an idealized society “can exist somewhere and at some time,” (RAWLS, 1999b; 127) it is not clear whether it must be possible to attain its *full* realization. Perhaps it suffices that it establishes an ideal to which we can get sufficiently (and continuously?) closer, even though we can never reach an endpoint. Although some notion of possibility must enter, I need not discuss which one it is, for the objections I consider question only the desirability of the ideal.

## 1.2 Full publicity and autonomous acceptance

With the practical condition at hand, let us flesh out briefly the idealization of publicity. In TJ, it is introduced as a formal condition for political conceptions of justice. It excludes from the start conceptions that become self-defeating once publicized, imposing that we judge a conception according to how well it fares once it forms a shared public basis of justification. It imposes that we judge principles of political justice as if they formed a “publicly acknowledged and fully effective moral [constitution] of social life” (TJ, 115), which “provides a mutually recognized point of view from which citizens can adjudicate claims of political right” (JAF, 9). What appears as a condition for us in judging the merits of a given conception of justice comes to life in the well-ordered society. This contrast between perspectives is characteristic: the ideal anticipates the realization of something we should be trying to achieve now. Thus in PL, Rawls describes three conditions of publicity, which are satisfied in the well-ordered society.

- (i) “Citizens accept and know that others likewise accept those principles, and this knowledge in turn is publicly recognized.” Furthermore, everyone accepts it “on the basis of commonly shared beliefs confirmed by methods of inquiry and ways of reasoning generally accepted.”
- (ii) They agree on the general beliefs “in light of which first principles of justice themselves can be accepted” (eg. facts about human nature and the working of institutions) because “these beliefs are supported by publicly shared methods of inquiry and forms of reasoning.” Thus citizens agree on how social institutions work in general, and they agree that their political association is subject to a number of conditions, such as moderate scarcity, of reasonable pluralism, and so on.
- (iii) The whole justification of the public conception of justice as presented in its own terms, is publicly available. It includes the argumentative path from a conception’s fundamental ideas to its specific principles mediated by the argument from the original position<sup>10</sup> (PL, 66-67).

No doubt, the full publicity condition is a very strong one. Certainly it is not a constraint

---

<sup>10</sup> I will return to this condition later on (3.3), since it provides some textual evidence to the effect that Rawls is aware of overintellectualization issues. He stresses that such a full justification may be only publicly available, meaning that it does not require every citizen to engage in such an intricate reasoning; thus he acknowledges that there is a limit to what can be reasonably expected of citizens’ moral reflection.

without which principles of justice are unintelligible, or incompatible with moral discourse, or the like. There are other moral outlooks that dispense with publicity. Sidgwick's version of utilitarianism, for instance, is openly incompatible with it<sup>11</sup>. In Rawls' view, merely pointing that such a view does not satisfy the publicity condition, as if it were a condition prior to moral argument, does not work. Instead, one must work out a substantive defense of publicity as a reasonable constraint<sup>12</sup>. As in every other part of Rawls' theory, the adequacy of it should be assessed according to whether it, and the resulting conception as a whole, fit our considered judgments in reflective equilibrium. So, what is Rawls' substantive case for publicity?

To begin with, a Rawlsian defense does not need to argue that publicity fits moral theories irrespective of their content and scope. His political principles are not designed to cover every aspect of our moral life<sup>13</sup>. They are restricted in scope, applied only to the basic structure of society<sup>14</sup>, a framework of basic institutions within which every social interaction occurs. The relationship of citizens within the basic structure is called "political relationship." It has two peculiar features: (i) adherence to it is not voluntary – we all enter in it by birth and cannot leave it as we wish<sup>15</sup>, and (ii) it subjects those involved to political power exercised in the form

---

<sup>11</sup> "On Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world" (SIDGWICK, 1962; 489).

<sup>12</sup> Williams' (2015; 120-122) critique of Sidgwick's position is one such defense, although he would hardly accept Rawls' strong publicity condition (defending that a doctrine is morally defective if it is collectively self-defeating is not the same as requiring that a doctrine be fully public). I am discussing moral outlooks generally, but one has to bear in mind that Rawls' discussion only concerns principles that apply to basic political institutions. This means that Rawls' critique of utilitarianism is primarily the critique of (some version of) the utilitarian principle as a principle to govern basic political institutions, not as a standard for individual conduct. That means, if one wants to advance a critique of Sidgwick based on political liberalism, one should focus only on the political facet of it.

<sup>13</sup> More than that, his principles are not even the application of a wider moral doctrine; they are supposed to be independent of comprehensive doctrines, because one of its aims is to be absorbed into radically different doctrines. See JAF (181-182).

<sup>14</sup> The basic structure of society is defined as follows: it is "the way in which the main political and social institutions of society fit together into one system of social cooperation, and the way they assign basic rights and duties and regulate the division of advantages that arises from social cooperation over time. (JAF, 10)". The reasons for such a restriction are of two kinds: first, as the basic structure provides a basic framework within which every social interaction occurs, it is responsible for maintaining "background justice." Second, due to its deep effects in citizens' life from the very start, of which citizens individually have little control, the justice of its functioning not only brings a sense of fairness to them, but also promotes the internalization of the ideals of person and society on which a conception of justice is based. Unfortunately, I cannot discuss these matters in detail.

<sup>15</sup> People can, in fact, leave the political society; that is what the right to emigration is for. Although emigration is an important matter, and the right to it an essential feature of a democratic citizenship, it hardly makes our subjection to political authority 'voluntary.' Brazilian dictatorship propaganda, which used to say "Brazil, love it or leave it" provides a regrettable illustration of the contrary thought. The fact that we are born in a specific social



of law, which is always coercive. For even reasonable individuals are not expected to agree with every reason used to justify the form of political authority, nor to agree with every piece of legislation, leave alone every legal decision (PL, 135-6). Moreover, the way the basic structure of society functions has great influence in citizens' life prospects<sup>16</sup>, and shapes importantly their characters and aims<sup>17</sup>.

In this context, one reason for publicity relates to its role in promoting an ideal of person. Rawls' ideal of person is that of an autonomous and fully cooperating member of society conceived as a fair system of cooperation over time. It implies that a person be regarded as free, in the sense of being capable of having a conception of the good that they are to advance in the cooperative scheme. They regard themselves and others as capable of changing their conceptions in light of considerations that they see as relevant (be them prudential or moral or even some experience of conversion). The ideal of person, as described, also implies that they regard themselves and others as equals, that is, as capable of advancing their conception of the good in the cooperative scheme in ways that are compatible with a fair system of cooperation to all (and everyone is capable of giving and receiving justification of that system, acknowledging that it is equally justified to all). That is what being a citizen is about. Now, according to Rawls, the existence of a public culture that expresses those ideals is a driving force toward the realization and maintenance of them in society. Thus according to Rawls,

Acquaintance with and participation in that public culture is one way citizens learn to conceive of themselves as free and equal, a conception which, if left to their own reflections, they would most likely never form, much less accept and desire to realize

---

position, which is inevitably affected by such a relationship, already hints that adherence to it is not voluntary. Furthermore, even if we manage to leave the political society, we will probably end up in another one, subject to the same kind of political power.

<sup>16</sup> Think of the life prospects of women in a society that does not officially recognize gender equality, say by excluding women from assuming certain political positions. Or think of the life prospects of someone who's born in a poor family in a profoundly unequal society where there's no distribution policy.

<sup>17</sup> Think of a society that officially recognizes distinct levels of citizenship, determined, say, by the family in which one is born. Such a society promotes a conception of person according to which there are citizens by birth worthier than others. And if it is a stable society, arguably many of the citizens from the lower strata will identify themselves as less worthy citizens, if not because the political culture excludes from them an important social basis of self-respect.

(JAF, 56).

It is Rawlsian motivational thesis that, when citizens grow under institutions that are known to live up to ideals of social cooperation, they tend to acquire themselves the desire to live up to the ideal of a cooperating member. Therefore, publicity in the well-ordered society serves an educational role.

Another reason for full publicity connects with the purpose of a conception that aims to govern the political relationship: as a conception that grounds institutions that rely on coercive power, it “should stand up to public scrutiny.” When it satisfies the publicity condition, it serves as a public basis for citizens to justify their conduct before one another “confident that this avowed reckoning itself will strengthen and not weaken public understanding.” In addition, it ensures “that citizens are in a position to know and to accept the pervasive influence of the basic structure that shapes their conception of themselves, their character and ends” (PL, 68). Hence, publicity allows citizens, as a collective body, to exercise coercive power upon themselves as individuals in ways that are consistent with them regarding themselves as free and equal. They regard themselves as equals because they “regard one another as having an equal right to determine, and to assess upon due reflection” (RAWLS, 1980; 309) the principles to govern the exercise of political power so that it is justified through their common reason. They regard themselves as free because, even though adherence to the scheme is non-voluntary, nothing about its coercive power needs to be hidden from each citizen in order for them to accept it.

But of special importance to me is the peculiar form of acceptance that is public in the well-ordered society of justice as fairness, which relates to a key liberal desideratum. It consists in the idea that respecting citizens’ autonomy requires that each citizen be able to find, by their own (but, of course, through discussion with others), good reasons to accept their coercive institutions. Call this the desideratum of autonomous acceptance.

Justice as fairness achieves that by picturing a state where every reasonable citizen accepts its principles because they are responsive the shared reasons that make them justified. This comes from the assumption that its citizens have an effective sense of justice, characterized as “the capacity to understand, to apply and to act from (and not merely in accordance with) the principles of justice” (RAWLS, 1980; 312). Now, I take it that acting from the principles of justice means that they act in virtue of the reasons that give such principles their normative content, and that the *in virtue* relation is strong enough to satisfy the following condition: when

one acts from the principles of justice, it is not an accident that one performs the action that is commanded by those principles.

Non-accidentality in acting for reasons allows one to draw the line between an agent acting in a way that happens to match what they have reason to do from an agent acting that way because that's what they have reason to do<sup>18</sup>. Note that even if a person steadily acts in accordance with some principle, that by itself does not warrant the claim that their action relates to the principle non-accidentally. For instance, if I am coerced to act in accordance with some principle with which I disagree, I may consciously act in accordance with it, but my reasons to act that way do not relate in any way to what is supposed to justify the principle.

In the well-ordered society citizens' acceptance is not accidental in that deep sense. It is not the result of coercion or indoctrination and it is not influenced by arbitrary considerations; rather it is acceptance for the reasons that make justice as fairness justified. Rawls connects such reasons with his argument from the original position; he states that the members of the well-ordered society acknowledge the principles of justice as those that "would be adopted in the original position," that is, those that "they would give to themselves when fairly represented as free and equal persons" (PL, 77).

Although it is a crucial feature of justice as fairness, I will not discuss the justificatory role of the original position, nor the merits of the argument from it<sup>19</sup>. Instead, I will draw some relations between Rawls justification programme as a collective search for wide reflective equilibrium and the referred idea of autonomous acceptance, at least as it is strongly suggested in Rawls' text.

---

<sup>18</sup> For an illustration, take the widely discussed "Kant's grocer" case. Suppose there are two grocery store owners, Karl and Immanuel. Each is currently finishing a transaction where a \$1.00 change is required. Both give the correct change. However, our characters are crucially different. Karl does not care about morality; he gives the correct change because he believes that this policy is good for business, while Immanuel does so because he believes that justice requires giving his customers the correct change. When we look at the reasons for which Karl and Immanuel acted, it seems that in both cases it is the fact that \$1.00 is the correct amount. However, it is clear that only Immanuel has acted for a moral reason, since Karl's action is conditioned by the fact that giving the correct change is good for him in the long run. That Karl's action is in accordance with a moral reason is, in this sense, accidental. This example appears in the first section of *The Groundwork of the Metaphysics of Morals*. I take the formulation given by Lord (2017). To more contemporary discussions of the case, see Way, (forthcoming); Schroeder (2018).

<sup>19</sup> Moreover, I will not discuss what it means to say that every citizen acknowledges the principles of justice as those that would be agreed to in the original position, nor whether this in itself amounts to an overintellectualization.

### 1.3 Rawls' justification programme and autonomous acceptance

So the task in hand is to account for the satisfaction of the desideratum of autonomous acceptance in the well-ordered society, and I have said that it is satisfied in the well-ordered society because citizens are responsive to the reasons that make justice as fairness justified in a way that it is no accident that they follow its principles. However, a great deal hinges on how to convey the idea that citizens are “moved” by the right reasons.

Consider the following predicament. The desideratum of autonomous acceptance imposes that the demands of justice be not wholly external to reasonable citizens' judgements. At the same time, ‘reasons for acceptance’ should not be understood as any consideration that actually motivates an agent. Political liberalism does not wish to give in to whatever existing motives. That would either fail to produce agreement, since people have conflicting opinions about political justice, or produce a kind of agreement that rests on the wrong reasons, a kind of compromise that would hardly express the ideal of cooperation among equals<sup>20</sup>. The fact that Rawls refuses to go beyond the domain of political values does not entail that he ceases to treat justice as a moral virtue, indeed a very demanding one. But then again, how could we distinguish right from wrong reasons and yet be committed to some non-external standard, one that appeals to each citizen's judgement? I argue that there are reasons to think that Rawls conceived of the search for wide reflective equilibrium as the key to reconcile those two demands. At some points, Rawls seems to link the satisfaction of the liberal desideratum of autonomous acceptance to the achievement of wide reflective equilibrium. But I need to explain the notion of reflective equilibrium first.

Search for wide reflective equilibrium is Rawls' quintessential method of justification for political justice. In fact, the method is all over the place now; no doubt it is one of Rawls' many philosophical successes<sup>21</sup>. But my discussion does not take reflexive equilibrium as a general method in philosophy (although it can be), nor will I try to defend it as the best/the only method for the justification of moral theories (although I would be in good company if I did<sup>22</sup>). Instead, I show that Rawls' appeal to wide reflective equilibrium seems especially promising

---

<sup>20</sup> For instance, as far as citizens' basic right of liberty of consciousness goes, political liberalism should deem irrelevant how strongly some religious groups wish to make their practices universal (see Scanlon (1977; 29).

<sup>21</sup> For a statement of the method's success, see Knight (2017).

<sup>22</sup> See for instance Daniels (1996a, 1996b), Scanlon (2003, 2014), Laden (2014).

as an idea of justification applied to political liberalism, which addresses reasonable persons and requires autonomous acceptance<sup>23</sup>.

The method of reflective equilibrium, as applied to political liberalism, begins with the articulation of considered judgements about political justice. These consist in well-formed beliefs or rules articulated from one's convictions about the subject-matter. Conviction comes in many forms; it might appear as a gut feeling (not articulated at all), or as a feeling accompanied by a wild guess or a more or less articulated opinion. It might be about particular types of action (eg. an act of racial discrimination is unjust) or about a social practice (eg. the institution of slavery is unjust) or even about the subject matter of morality itself (eg. causing unnecessary suffering is wrong). It must, however, have practical force, that is, it must be capable of moving the agent. Moreover, a conviction is eligible to become a considered judgement only if its articulation is made under conditions that are conducive to reliable judgements and exclude common sources of distortion. Thus judgements "given when we are upset or frightened, or when we stand to gain one way or the other (...)" are excluded since they "are likely to be erroneous or to be influenced by an excessive attention to our own interests" (TJ, 42). For the same reason, there should be no time pressure and the agent should be in a position to know the relevant fact of the matter<sup>24</sup>.

One is asked then to articulate at least those considered judgements in which one has

---

<sup>23</sup> As in many other respects, political liberalism does not affirm nor deny that appeal to wide reflective equilibrium is the only, or even the deepest form of justification. Think of some kind of intuitionism, according to which there is an external order of values that we can assess via some special faculty. Although the idea of justification associated with that kind of intuitionism might prove philosophically weak, it might be regarded as a deeper form of justification for a reasonable person who, in the public domain, acknowledges that appeal to wide reflective equilibrium is the only publicly shareable idea of justification.

<sup>24</sup> An interesting objection to such an account of considered judgements questions whether such conditions are always more conducive to reliable judgements. According to it, there are a number of circumstances where having distressful feelings is actually conducive to good judgement, and perhaps failing to experience those feelings gets in the way of good judgement. Think of Williams' (WILLIAMS, 1981) case of a person who must choose between saving their partner's life or that of a random stranger. The sense of urgency that the person is likely to experience plays a key role in forming the judgement that they should save their partner (if there is a judgement to be made; perhaps they should just do it). In fact, if that person stopped to look at things from an impartial perspective to consider what is the right thing to do, we would probably judge it as a lack of character. Likewise, the sense of indignation that comes from experiencing an act of injustice is sometimes crucial in judging it as wrong, and moving too fast to a moment of cool reflection might distort things. While I think that the objection touches on a valuable insight, namely that good judgement sometimes requires one to be emotionally shaken, it appears to misunderstand the role of articulating considered judgements. Articulating considered judgements is a work of abstraction, and it is attached to a specific mode of reflection, which consists in putting things into perspective. The latter is not meant to be present in one's whole ethical experience, since more often than not life demands one to engage one's feelings when thinking about what to do. However, insofar as one is willing to put things into perspective, it seems that those conditions for considered judgements apply. Now, when, and to what extent, a reasonable person is required to put things into perspective seems to me a far more troubling question (which relates to the second overintellectualization objection, as I develop in 2.2).

the most confidence and systematize them in an abstract conception of justice. There are reasons to think that one's existing judgements won't form, at least initially, a coherent set of beliefs or rules. One's moral sensibility is formed in different contexts and at different times without any awareness of the place of one's various convictions into their implicitly affirmed moral conception (if there is one to be found). The upshot is that one's existing considered judgements and any proposed systematic conception of justice are bound to conflict. In the face of such a conflicting relation, one hopes that by repeatedly either dropping judgements incompatible with the conception, or modifying the conception to account for an important judgement, one eventually finds a satisfactory match between their judgements and a conception. That state of affairs is called reflective equilibrium.

So far, the process concerns an individual reflecting on their own conception of justice, as it is expressed initially in a non-systematic fashion by their considered judgements, and finally as a coherent whole. By the end of it, the individual's moral sensibility is expected not to be in conflict with itself anymore; it might be said that reflective equilibrium produces integrity. However, Rawls' treated justification in political philosophy not merely as an attempt to improve one's view regardless of the reasoning of others. In a word, justification is not to be merely justification to self; rather, it is "addressed to others who disagree with us" (TJ, 508; JAF, 27), and it should aim at what we can affirm together. Hence, besides such a profound self-assessment, one should test one's affirmed conception considering what conceptions others would affirm if they tried to reach reflective equilibrium themselves, comparing those conceptions with one's own.

Ultimately, one is required to compare every possible alternative conception, as well as every other combination and every argument for them (1999b; 289). The ideal endpoint of such a process is called wide reflective equilibrium. For obvious reasons, no one can actually achieve wide reflective equilibrium. Even if we discount limitations of any role-taking exercise, achieving wide reflective equilibrium, for any particular individual, is an ideal in itself. It requires that one constantly put their convictions to proof against possible considerations, and one can never be sure that considerations are exhausted (especially, one cannot assure that important considerations won't appear in the future). In addition, one would have to conduct such an endless process from the others' perspectives as well. That is why Rawls restricts the task to what he sees as the "next best thing": uncovering and further developing the structure

of the doctrines familiar from the philosophical tradition<sup>25</sup>.

It seems plausible to say that a person who accepts a conception of justice in wide reflective equilibrium accepts it with the best justification available to that person. For what else could one possibly demand of anyone beyond the careful reflection on their considered judgements in comparison to the considered judgements of others (expressed in the alternative conceptions of justice)? By the same token, a conception of justice that is accepted by *us* in wide reflective equilibrium is accepted with the best justification available to *us*. Perhaps the truth of the matter lies somewhere outside such a state, but then truth lies beyond the point of political liberalism. As Rawls puts it,

What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us (RAWLS, 1980; 306-307)<sup>26</sup>.

Now, I agree with Scanlon (2003) in that the first person singular perspective has primacy in the search for wide reflective equilibrium. It is possible to conjecture about other persons' convictions, and how critical reflection might change them. In fact, search for wide reflective equilibrium requires one to make such conjectures. However, as an exercise of role-taking, it has its obvious limitations. Moreover, the test for any conjecture is that of the sincere reflection of the person whose judgements are in question. In Scanlon's words, "the process of seeking reflective equilibrium is something we each must carry out for ourselves, and it is a process of deciding what to think, not merely one of describing what we do think" (149).

---

<sup>25</sup> Rawls does not elaborate on why this is supposed to be second best thing to do. Perhaps it is based on the thought, borrowed from Sidgwick (1962; 6), that philosophers are able to express common people's views in a more systematic fashion. Even though I am not fully convinced, I will leave this issue aside.

<sup>26</sup> There are a number of possibilities here. First, what we affirm in wide reflective equilibrium might coincide with an antecedently given order of values, in which case it does not matter (practically speaking) that such an order be uncovered. Second, what we affirm in wide reflective equilibrium might be in conflict with an antecedent order of values, meaning that truth is not accessible through reflection. Thus, truth is unavailable to us through our powers of reflection, hence it cannot have practical significance. Third, such an order of values might be constituted by our affirming it in wide reflective equilibrium. Fourth, it might be that no conception of justice (or a family thereof) would be collectively affirmed in wide reflective equilibrium, in which case political philosophy (at least of Rawlsian influence) can only hope to reduce disagreements. Hence, it seems that the notion of truth of an antecedently given order of values does not have any practical significance that is not achieved by collective acceptance in wide reflective equilibrium.

This is why Rawls' justification programme in political philosophy consists in the continuous collective effort to reach agreement while searching for wide reflective equilibrium. Rawls' theses in political liberalism, I argue, should be read as conjectures about what reasonable persons would affirm in wide reflective equilibrium. This way everyone is called to 'take the driver's seat' and advance the justification programme, if they will. Such an interpretation is confirmed by Rawls' assertion that it is from the point of view of "you and me" (not from an artificial or ideal point of view) that we assess any conception of justice (PL, 28). And from that point of view

The test is that of reflective equilibrium: how well the view as a whole articulates our more firm considered convictions of political justice, at all levels of generality, after due examination, once all adjustments and revisions that seem compelling have been made. A conception of justice that meets this criterion is the conception that, so far as we can now ascertain, is the most reasonable for us. (*ibid.*)

Ok, great, but what does this have to do with autonomous acceptance? Well, it seems that appeal to reflective equilibrium is able to eschew the earlier predicament of how to distinguish right from wrong reasons and yet avoid commitment to a standard of justification that is not external to citizens' reasoning. First, since the search for wide reflective equilibrium is something that we each must carry out for ourselves, its results are not external to each of us. Second, it is clear that we, flesh and blood persons, are not in a state of wide reflective equilibrium; indeed, we are a long way from it. Therefore, appeal to wide reflective equilibrium does not bow to whatever existing motives. If one adds that the best justification available for us at any given time is reached by the occurrence of shared acceptance in wide reflective equilibrium, there lies a way to distinguish right from wrong reasons without appealing to an external standard.

Furthermore, Rawls seems to suggest at times that the idea of autonomy requires one to be searching for wide reflective equilibrium, as he assumes that "free and willing agreement is agreement endorsed by our considered convictions on due reflection" (RAWLS, 1987; 425, n.8). That is, the expression of a citizen's will to act according to a conception of justice is non-accidental if it is the result of a process of reflection, and I suppose that the more far reaching the reflection the less subject to contingencies and arbitrariness is the citizen. This seems to explain Rawls' suggestion to "think of each citizen [in the well-ordered society] as having



achieved wide (...) reflective equilibrium” (JAF, 31). Realizing that state amounts to reaching the endpoint of the justification programme<sup>27</sup>, and at that point the liberal desideratum of autonomous acceptance seems to have been achieved at the highest level. However, if political autonomy commits one to the search of wide reflective equilibrium, it comes at a very price, as I explain in 2.2.

#### **1.4 Stability and overlapping consensus**

The first three sections have dealt with very general features of Rawls’ project. Now, I want to explain how the deepest form of acceptance would come about in his utopia, that is, I want to describe the occurrence of the overlapping consensus. My aim is to articulate the connection between the overlapping consensus and Rawls’ notion of justification. But before that, let me contextualize it a bit and explain why the idea is so important.

The idea of an overlapping consensus appears as part of a response to the problem of the stability of a conception of justice. Stability is a persistent concern in Rawls work, and it has far greater importance than meets the eye. More than just addressing the problem of avoiding major disturbances in the political order (a difficult problem in itself), stability is central to his “defense of reasonable faith” (PL, 101, 172). In short, Rawls believed that our efforts to produce a political society oriented toward the cooperation between free and equal citizens would be frustrated from the start unless we believed that such a society is a real possibility. Merely providing principles that, if effectively regulative of society, would ensure

---

<sup>27</sup> Strictly speaking, we cannot reach an endpoint of the justification programme. One cannot ever ascertain that the relevant considerations have permanently run out, for even if the totality of existing conceptions of justice, and every possible combination of them, were considered (which is impossible in itself), it wouldn’t follow that no other relevant consideration could arise in the future. Besides, one cannot ascertain that the generally agreed facts about human nature and society that underlie reasonable conceptions today would not prove wrong tomorrow (JAF, 5). Therefore, the justification programme as a collective search for wide reflective equilibrium cannot ever reach an endpoint, as long as political society continues to exist. Rawls says, “it is a point at infinity we can never reach, though we may get closer to it (...) (PL, 385)”.

Nonetheless, I believe that this fact is compatible with picturing how such an endpoint would look like in a very specific sense. Following Rawls, one should distinguish two different points of view: that of “ourselves – of you and me –,” who are advancing the justification programme by searching reflective equilibrium, and that of the members of the well-ordered society (PL, 28). Now, the well-ordered society is an idealization of what we see, here and now, as the most reasonable conception of society; it aims to depict what the endpoint of the justification programme would look like if the search stopped right now. Since it aims to be the conception that we, here and now, have the most reason to accept, we should indeed try to realize it. However, we are not stuck to that conception; as reflection goes on, it may undergo radical change.

fair cooperation is not enough to cheer us up as to our future social world. In order to redeem reasonable, or practical faith, one needs to show that our psychology, together with other facts about us and our political world, are conducive, or at least not averse, to such principles. Moreover, our concern for justice cannot be that of providing a snapshot of just institutions; political institutions are an ongoing scheme, which supersede the life of any particular citizen. Hence, Rawls takes great pains to show not only that a just scheme might come about, but mainly that a just society can remain stably just over time. Thus, defense of reasonable faith for Rawls takes the form of a “defense of the possibility of a just constitutional regime” (*ibid.*)<sup>28</sup>. Note that the stability problem is designed for ideal theory. It asks not how we, under our unfortunate non-ideal circumstances, may arrive at a stable just society. Rather, it asks whether people would maintain over time a just scheme when they found themselves already in it.

Stability, in this context, predicates a conception of justice, that is, its capacity to “generate its own support” when effective regulative of society’s basic institutions (JAF, 186; TJ, 398). In other words, a conception of justice has to produce a society in which people grown in it acquire sufficient motivation to uphold the scheme of cooperation without the need for excessive ‘external intervention,’ such as the constant and widespread use of legal coercion. A lot of things may undermine stability, for instance, the propensity to feel envy in some contexts, the limits of our altruistic behavior, and temptations to give priority to other cherished values in particular occasions. The structure of Rawls’ case for the stability of his conception of justice is the following: (i) showing that citizens grown in the well-ordered society of justice as fairness would normally acquire the corresponding sense of justice (a sense of justice informed by justice as fairness); and (ii) showing that the acquired sense of justice will normally provide reasons that outweigh conflicting reasons in the relevant circumstances. Rawls hopes to defend (ii) by showing that, in the well-ordered society, living up to the ideals of individual conduct and association that are part of their sense of justice is, in some way, part of citizens’ good.

Although the structure of Rawls’ case for stability is roughly the same throughout his work, significant parts of his argument underwent radical changes – so radical as to demand modifications in the whole theory. I cannot account for either his whole argument or the reasons for and content of its modifications. Rawlsian scholars run into deep disagreements on that matter, and I’m not in a position to take a side in that debate. I intend to ‘stay on the surface’ of

---

<sup>28</sup> Paul Weithman (2009) provides a rich account of Rawls on practical faith, relating it to philosophical trends of his time, as well as his Kantian influence. See also his discussion of Rawls’ theodicy in this connection (2010; 11-14).

those disagreements, so to speak. I focus on Rawls' later works, particularly on the overlapping consensus argument as giving content to the second step of Rawls' case for the stability of a political conception of justice. Thus, my remarks will assume some acquaintance with Rawls' work. I hope, though, that they end up clear enough and sufficient for the present purpose of clarifying his notion of public acceptance in the well-ordered society.

Focusing on the second step of Rawls' case for stability, he asks: how can we ensure that, in citizens' deliberations, the demands of the conception of justice will have, in sufficiently many cases, enough weight to prevail against conflicting demands of other values? Otherwise, their motivation to uphold the scheme of cooperation would be too easily overridden, meaning that social cooperation based on fair terms would be in constant threat of collapsing. In order to overcome those risks, we saw, Rawls has to show that citizens' sense of justice in the well-ordered society is regulative of their political lives. The argument from the overlapping consensus is introduced as an answer to that problem. I will sketch a version of it in what follows.

Let me begin with Rawls' characterization of the two aspects of a reasonable person. First, a citizen, insofar as they are reasonable, will be willing to propose, accept and abide by fair terms of social cooperation when they judge that it is reasonable for everyone to accept those terms. Second, a reasonable person also recognizes the burdens of judgments – “the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life” (PL, 56) –, which make it unlikely that citizens' common reason will always bring convergence to their judgments<sup>29</sup>. Consequently, they acknowledge a space of reasonable disagreement.

The first aspect of a reasonable person sets the bar below which, Rawls would say, no project of political justification is fruitful. For the first disposition of the reasonable is a condition of openness to other people's reasons, without which one is not moved by any justification, except perhaps the one they give for themselves. In other words, Rawls' political liberalism addresses citizens who, in their political relations, have the moral motive of reciprocity towards others, which drives them into the public world, demanding from them justification to others (PL, 49-50; 53-54).

---

<sup>29</sup> Rawls gives a list of some of those sources of disagreement: complex and conflicting evidence; the weight of the relevant considerations; vagueness of our concepts; the influence of life experience in our judgment; the multiplicity of kinds of normative considerations that bear on a single issue, and the impossibility of realizing all the society's cherished values at the same time (PL, 56 - 57).

The second aspect acquires special importance in light of what Rawls calls the fact of reasonable pluralism. Roughly, it is the fact that, in our society, people who are sensitive to the motive of reciprocity nevertheless disagree over what is fundamentally valuable in life – they affirm different comprehensive doctrines<sup>30</sup>. Moreover, that fact is not conceived as a sad accident that happens to plague modern societies; rather, Rawls puts it as an expected consequence of living under free institutions over time (PL, 129).

Now, imagine that Bill, a reasonable guy, affirms a comprehensive doctrine that consists in believing a more or less coherent set of claims about values. Let us suppose that Bill's affirmed political values are subordinate to a domain of nonpolitical values, say, a view of what makes for a well-lived and excellent human life. He cherishes political values such as equal basic liberties, equality of opportunity, and distributive justice and gives them due priority because he sees them as conditions for the development of human excellencies. He believes that the political values are so conditioned, and we may suppose that believing something is the same as believing that it is true. Besides, usually we want to defend the claims we believe are true all the way, at least until we find reasons to stop believing them. But, if this is so, does Bill have any reasons at all for refraining from pressing his claims about human flourishing when proposing what he sees as the right ordering of political values to govern political institutions? At first sight, it seems that he does not. Given that the value he attaches to political values is subordinate to other nonpolitical values, perhaps one could even find it shady if he refuses to advance his 'true' reasons for accepting one rather than other arrangement. If so, acknowledging the fact of reasonable pluralism would amount to acknowledging endless disagreement about the ordering of political values.

Here's the catch: once Bill (as a reasonable guy) recognizes the burdens of judgement, he realizes that other people who are equally well informed, capable of sound reasoning, and disposed to reach agreement may arrive at different, but reasonable, practical conclusions. It is only when Bill acknowledges such burdens that he is in a position to distinguish his belief that something is true from the public recognition of it being true – as Waldron (1999) says, truth about justice “never appears in politics *in propria persona*, but only – if at all – in the form of somebody's controversial belief” (111). What lacks, then, is a common shared basis for distinguishing merely believed judgements about fundamental values from true ones. But

---

<sup>30</sup> I discuss in length the notion of reasonable pluralism in connection to that of a reasonable comprehensive doctrine in the following chapter.

because of the fact of reasonable pluralism, none such basis is to be found<sup>31</sup>.

As a result, Bill has the further motive to refrain from relying on his particular comprehensive doctrine when discussing fair terms of social cooperation. Political liberalism appeals to people like Bill as it tries to settle the issue by offering a conception the construction of which is restricted to values entrenched in the history and culture of a democratic society. It avoids values of any particular view of the fundamental values in life. Instead, it looks at the public culture as “the shared fund of implicitly recognized basic ideas and principles.” (PL, 8-9) From there, it works out a ‘freestanding’ political conception of justice designed to apply to a well-ordered society. Political liberalism then hopes to provide an ideal of society in which citizens make good their motivation of reciprocity toward others by subjecting their basic political institutions to a conception that is independent of any particular comprehensive doctrine.

However, staying on the surface of political ideas has the consequence that, as such, political liberalism doesn’t make any direct claim as to the content of people’s other, deeper reasons and motives (be them moral, philosophical, or religious). These include one’s particular conception of the good, but also various judgements related to justice that fall outside the domain of political values. At this point, it could be asked: do the members of society have reasons of their own to accept the political conception of justice? And what assurance do we have that they would generally give the priority needed for the support of just institutions in light of their other important values? How can we say that people would not be tempted to give priority to other cherished values that might conflict with the demands of political justice in particular cases?

Rawls’ conjecture that an overlapping consensus obtains in the well-ordered society plays a central role in answering those questions. To summarize, the overlapping consensus is the state of affairs in which citizens have accepted the political conception of justice because their reasonable comprehensive doctrines have incorporated it from their own particular point of view. Thus, it delivers the reconciliation between citizens’ moral motive of reciprocity expressed in a conception of justice drawn from a general perspective and their deeper motives. In addition, it does so without presupposing adherence to a single, or a restricted set, of comprehensive doctrines, since every reasonable comprehensive doctrine is supposed to have incorporated the political conception from its own perspective. Therefore, when the overlapping

---

<sup>31</sup> This paragraph draws on Nagel (1988).

consensus obtains, citizens are not normally torn between their political and non-political values – their acceptance of the political conception is wholehearted. And if such an overlapping consensus is publicly known to obtain, then citizens have the strongest assurance that others will also be motivated to uphold the scheme, which enhances their mutual trust.

Defending the conjecture that an overlapping consensus obtains is the last step in the construction of Rawls' realistic utopia. It aims to show that a society effectively regulated by justice as fairness would generate the deepest possible public acceptance compatible with a democratic framework. There, each citizen wholeheartedly accepts justice as fairness, even if the content and form of what makes acceptance wholehearted is not the same for every citizen.

However, reading Rawls we get many colors of the overlapping consensus, some more intellectual than others. In what follows, I lay down what, in my view, is the most exacting picture of an overlapping consensus, which connects it with ideas of justification and reflective equilibrium.

### **1.5 Overlapping consensus and the three kinds of justification**

Let me flesh out those thoughts using Rawls' Reply to Habermas as a guide (PL, 385-395). I am especially interested in the response to Habermas' question "whether an overlapping consensus adds to the justification of a political conception of justice already taken to be justified as reasonable." Does it, or is it just "a necessary condition of social stability"? (PL, 385) Rawls claims that the overlapping consensus has a role of justification in that it is a necessary condition for the best possible justification of a conception of justice in a democratic society. He defends that claim by distinguishing three kinds of justification. In the remainder of this chapter, I reconstruct Rawls' argument to that effect<sup>32</sup>.

He begins with a kind of justification that is restricted to political values – *pro tanto* justification. It aims to justify the conception of justice in public reason by showing that

---

<sup>32</sup> My main interest in that reconstruction is the relation between the overlapping consensus as the deepest form of public acceptance in political liberalism and Rawls' idea of justification. Rawls' response to Habermas includes an account of legitimacy that is crucial to his argument. I am more interested in the way Rawls exposed political liberalism as a response to Habermas than in whether the response is sound.

The political values specified by it can be suitably ordered, or balanced, so that those values alone give a reasonable answer by public reason to all or nearly all questions concerning constitutional essentials and basic justice (PL, 386).

This kind of justification covers the case of citizens of the well-ordered society wanting to justify themselves to every other citizen in public, in a word, citizens using public reason. Reasonable citizens, acknowledging the fact of reasonable pluralism, do not expect convergence that is originated in anyone's particular perspective; hence they feel the need of a suitably general perspective for reasoning in public. Reasoning from such a perspective is a condition for them to offer reasons that they expect others to accept, thus it is a condition for them to express a disposition of reciprocity. Rawls thinks that only what he calls the special domain of political values can be detached from citizens' particular perspectives and form a common basis for public reason. Thus, *pro tanto* justification is called for when it is crucial to reason from a perspective that is independent of anyone in particular, or "freestanding." There are two main applications of that kind of justification.

First, it is the justification that citizens of the well-ordered society give each other in the political forum when further specifying the political conception of justice in discussing its application to matters of basic justice using the guidelines of public reason. For instance, it is the justification in play when they exercise constituent power, in a constitutional convention, to flesh out the political conception of justice into applicable basic rights and liberties. It is also the relevant kind of justification appealed to by judges of the well-ordered society when deciding matters that touch on constitutional essentials by appeal to the political conception.

Second, it is the justification relied on by Rawls himself in the construction of the justice as fairness<sup>33</sup>. Rawls asserts that the content of public reason is given by a political conception of justice (PL, 223). It is the conception of justice that specifies the two relevant kinds of political values: (i) values of political justice, "the values of equal political and civil liberty; equality of opportunity; the values of social equality and economic reciprocity" etc. and (ii) values of public reason, "the guidelines for public inquiry, which make that inquiry free and public" (PL, 224). But if the political conception of justice is to play that role, it must itself be constructed from political values alone. That is, if we assume that only political values can be

---

<sup>33</sup> For simplicity I've been talking only of Rawls' favored liberal conception. But, as I have already averted, political liberalism allows for the possibility of an overlapping consensus around a family of reasonable liberal conceptions of justice.

suitably detached from citizens' individual perspectives and form a common basis for agreement, the general acceptance of the conception of justice depends on it being founded exclusively on political values. Rawls' argumentative strategy is in line with that thought. He starts from certain familiar political values, expressed in basic ideas to be found in the public political culture<sup>34</sup>. Then, he attempts to further specify those ideas so that he ends up with a progressively more complex ordering of values that might address the present disagreement on what is the best way to "satisfy the fair terms of cooperation between citizens regarded as free and equal" (PL, 4).

So far, I have only stated Rawls' claim that only appeal to the special domain of political values might succeed in forming a common basis of justification for public reason. I have also listed a couple of political values cited by Rawls. Yet, what are the so called political values, and what is so special about them? In 1.2 I have described the political relationship of citizens within the basic structure of society. Recall, that relationship has two relevant features: adherence to it is not voluntary, and everyone individually is subject to the (coercive) political power that is exercised in the name of all. Political values, then, are those that are expressed in that particular relationship – the relationship of citizens as such –, or the values of their association, if you will. But why are these values 'detachable' from anyone's individual perspective, and in what sense are they detachable?

That issue is connected to the much discussed Rawlsian claim that the basic ideas of justice as fairness are drawn from "the public political culture of a democratic society, and to the traditions of interpretation of its constitution and basic laws" (JAF, 5). To address the issue of how to specify the basic rights and liberties, opportunities and rules of distribution in a way that do justice to the relationship of citizens as free and equals, Rawls avoids resource to the considered judgements of anyone in particular. Instead, he turns directly to the institutional tradition of a democratic society and aims to identify there some familiar ideas that convey relevant political values. If we grant that Rawls is allowed to do such a move, it seems that he has found a great spot to base his construction, since he is then able to spell out the content of a conception of justice without appealing to anyone's particular perspective<sup>35</sup>.

---

<sup>34</sup> Roughly, the core ideas that form the basis of justice as fairness are that of society as a system of social cooperation between persons who are seen as advancing a conception of the good in that system and who deserve (and are able to provide) an equally good justification of it.

<sup>35</sup> I do not claim that Rawls is able to avoid anyone's particular perspective at any point. That would not be consistent with the importance he gives to the conjecture that the freestanding political conception will form an overlapping consensus. What he would be able to do is to spell out *the content* of his doctrine without such an appeal.



Why should we think that Rawls is allowed to do such a move, though? Part of the Rawlsian answer lies in the special form that institutionalization of political values takes, together with Rawls' view that shared institutions have a major role in shaping citizens' self-conceptions. As Joshua Cohen (1994) puts it, "political ideas are institutionalized in a democratic society in ways that comprehensive moral – or religious or philosophical – ideas are not." While comprehensive doctrines are "institutionalized – if at all – in (...) associations that are not shared" by all citizens<sup>36</sup>, political ideas are "acquired in part through shared associations" (1533). Rawls' move relies on that fact together with the conjecture that

Citizens who grow up within a reasonably stable democracy will find [the basic political ideas] familiar and attractive: the political ideas "expressed" in common, public institutions and appealed to in the culture to justify those institutions will shape citizens' moral-political education" (*ibid.*, 1532).

Since Rawls' intended audience is primarily a people that publicly professes democratic government<sup>37</sup>, he thinks that the institutional forces that inculcate democratic ideals are, at least imperfectly, actually at play. I say imperfectly because there is always a gap between what actual societies publicly profess and the extent to which they, through their institutions and their individual actions, live up to their public identities, so to speak. Rawls does not claim that any actual society is even close to realizing their professed democratic ideals. Instead, he hopes to find in the tradition of democratic thought, to the extent that it is institutionalized now, a germ through which 'we' (his audience) might work out a deeper form of agreement. Such an approach relates to Rawls' view of political philosophy as reconciliation, that is, as providing an account that helps us "accept and affirm our social world positively, not merely to be

---

<sup>36</sup> The basic case is that of association in different churches, each with different comprehensive doctrines.

<sup>37</sup> I do not consider the controversy of whether the claims of political liberalism could be translated to peoples who do not live in publicly professed liberal democracies (see Nussbaum, 2015; O'Neil, 2003). I do think that it is extremely important that political liberalism be presentable to societies with different institutional histories than liberal democracies (also liberal democracies with a different institutional history than that of the United States). But Rawls' choice of audience is not necessarily a problem in itself, as long as he has identified a real practical problem that faces liberal democracies. Perhaps finding a *locus* of shared ideals and principles that convey some idea of reciprocity in societies without institutionalized liberal ideals will prove more difficult, but that does not mean that one shouldn't try it.

resigned to it” (JAF, 3).

Therefore, the fact that the content of justice as fairness is worked out without appeal to anyone’s particular perspective by no means imply that Rawls is indifferent to citizens’ perspectives, or, better put, to their considered judgements. He does not take political values to be detachable because they are external, or indifferent, to citizens’ judgements. Quite the contrary, he only starts in the public political culture because he thinks, in part due to the alluded institutional pressures, that we may look at it as “the shared fund of implicitly recognized ideas and principles” (PL, 8-9). Thus, the special domain of political values is “detachable” from individual perspectives for two reasons: (i) it is a domain of shareable values, and (ii) public political culture embodies such values in a way that it is possible to construct the content of justice as fairness without direct appeal to anyone’s perspective<sup>38</sup>. Borrowing an expression of Rawls himself, the “detachable” character of the political values is better understood as a mode of presentation. He says, the “political conception of justice is presented as freestanding” (PL, 12), but it certainly does not appear as freestanding to anyone that comes to affirm it.

Important for our purposes at the moment is the following. From the identification of those core ideas in the public political culture to the specification of the principles and ideals of justice as fairness, as well as the further specification when discussing constitutional essentials and matters of basic justice, appeal is made only to political values from a suitably general perspective. Consequently, *pro tanto* justification is all we need up to now.

However, as I have said, the political conception of justice, presented in theory as freestanding, nevertheless appears to particular individuals who affirm it in practice as part of their beliefs and commitments. Rawls refers sometimes to the political conception as a module, i.e. “an essential constituent part” (PL, 12; 145) of citizens’ moral outlook that fits in each perspective in different ways. That is one way of describing the overlapping consensus: the state where each citizen has incorporated the ‘module’ into their perspectives, or comprehensive doctrines. Take Rawls’ claim:

In a well-ordered society supported by an overlapping consensus, citizens’ (more general) political values and commitments, as part of their noninstitutional or moral

---

<sup>38</sup> Such a detachment is given the clearest expression in *Political Liberalism*, but Gališanka (2019) identifies a germ of that idea implicit in his earlier works, namely the assumption that “parts of reasonable persons’ conceptual frameworks are sufficiently independent from others” (8). Surveying Rawls’ private papers (opened to the public only recently), Gališanka tracks the various influences that shaped such an assumption, including Rawls’ early religious positions.

identity, are roughly the same” (PL, 31-32).

I believe that Rawls means something among the following lines. What distinguishes citizens’ particular perspectives is their ‘moral identity,’ their view of themselves as persons who acknowledge values from a variety of domains, including the political. Rawls assumes that the pluralism of moral identities in society are given by the different comprehensive doctrines that each affirms. Thus the questions “how can the values of the special domain of the political (...) normally outweigh whatever values may conflict with them” and “how can we affirm our comprehensive doctrine and yet hold that it would not be reasonable to use state power to gain everyone’s allegiance to it” (PL, 139) are treated as practically the same. And part of his answer is given by the following (for all intents and purposes) equivalent claims: in the well-ordered society of justice as fairness,

- (i) Each citizen acknowledges roughly the same political values from their own perspective.
- (ii) Each citizen acknowledges roughly the same political values in their moral identity.
- (iii) Each citizen affirms roughly the same political conception of justice in their particularly affirmed comprehensive doctrines.
- (iv) An overlapping consensus obtains.

Now, if Rawls wants to use an idea of justification in connection to such an incorporation, it seems clear that he needs something other than *pro tanto* justification. He needs a kind of justification that covers every possible domain of value, what he calls full justification. This is the justification “carried out by an individual citizen as a member of civil society,” by which “the citizen accepts a political conception and fills out its justification by embedding it in some way into the citizens’ comprehensive doctrine as either true or reasonable” (PL, 386). Rawls is quite laconic about how such a justification is to be carried out. But that is actually understandable, since the source of reasons for accepting the political conception are the existing reasonable comprehensive doctrines, and these are various and often radically different. Thus, he leaves to “each citizen, individually or in association with others,” (*ibid.*) to carry out full justification using the reasons particular to their comprehensive doctrines. A successful exercise of full justification yields the wholehearted acceptance of the political

conception: the citizen now endorses the political conception, with its principles, values and ideals, as part of their “noninstitutional or moral identity.”

Full justification of the political conception is characteristically an individual exercise<sup>39</sup>. But if “all reasonable citizens” have carried it out (PL, 387), then an overlapping consensus surely obtains. In order to secure stability, however, Rawls needs to impose yet another condition. Besides the achievement of an overlapping consensus, there has to be public knowledge of it. When that final step is made, it means that public justification has been carried out. To that effect, citizens must “take one another into account as having reasonable comprehensive doctrines that endorse [the] political conception, and this mutual accounting shapes the moral quality of the public culture of political society” (*ibid.*). I take him to mean that such a mutual acknowledgment provides each citizen with sufficient assurance that others will act on the principles and ideals of the conception of justice, because each knows that others’ acceptance of it is wholehearted<sup>40</sup>. And that mutual assurance “shapes the moral quality of the public culture” precisely because it is a condition of reciprocity without which the reasonable cannot flourish in the social world. As Rawls put it,

Insofar as we are reasonable, we are ready to work out the framework for the public social world, a framework it is reasonable to expect everyone to endorse and act on, provided others can be relied on to do the same. If we cannot rely on them, then it may be irrational or self-sacrificial to act from those principles (PL, 53-54).

That is, being reasonable in politics when everyone else is not will be both ineffective, since politics is about collective action, as well as “self-sacrificial,” because others will

---

<sup>39</sup> Strictly speaking it is not individual, since Rawls adds that it might be done in association with others. That is an important point about full justification, especially once one acknowledges that comprehensive doctrines usually belong to traditions of thought. Talk of individual exercise is a simplification, but it serves to distinguish it from a *public* exercise. For even if one admits that full justification might be carried out in association, it is meant to be association within comprehensive doctrines. Given the fact of reasonable pluralism, that association cannot be extended to the public in general, at least without the oppressive use of state power. Rawls’ uses a peculiar notion of ‘public,’ which is restricted to the public association of every citizen in society; thus, his dichotomy is not ‘public vs. private,’ rather it is ‘public vs. nonpublic’. That is worth reminding, since Rawls categorically rejects the notion of “private reason;” according to him, “there is no such thing” (PL, 220, n.7).

<sup>40</sup> A qualification is needed here. Public justification requires public knowledge of the existence of an overlapping consensus, but it cannot require that such a public knowledge extends to the content of the particular doctrines. As Rawls puts it, “the express contents of these doctrines have no normative role in public justification” (PL, 387). Citizens must know *that* general acceptance of the conception of justice is wholehearted, nothing more.

probably take advantage of one's disposition refrain from pressing one's interests. One may question why it is that only the mutual acknowledgment of the existence of an overlapping consensus can provide such a mutual assurance that makes reasonableness possible. Couldn't citizens of the well-ordered society reach a weaker form of consensus over a political conception of justice that were yet sufficient to secure mutual assurance? For instance, one might entertain, people might stably accept a fair political process and a set of liberal basic rights as a sort of compromise. The competing doctrines in society are not fully satisfied, since they would prefer to have universal adherence to their cause, but stability obtains via the public knowledge that (i) no current political view, nor any conceivable comprehensive view, can convert everyone else; and (ii) having a fair decision-making process for collective action and the protection of those liberal basic rights is the second-best alternative. Perhaps this is correct, yet there are at least two reasons for insisting on a deeper consensus.

First, since a comprehensive doctrine also determines the nature and content of an individual's deepest moral reasons, knowing that an overlapping consensus obtains amounts to knowing that the moral reasons for accepting the political conception run as deep as they can for each particular citizen. This, besides enhancing mutual assurance, is crucial for ensuring stability for the right reasons. Showing the moral character of such a consensus is essential to Rawls' case for the possibility not merely of a stable society, but of "the deepest and most reasonable basis of social unity available to us in a modern democracy" (PL, 391).

Second, consider that a comprehensive doctrine is said to specify, among other things, an individuals' conception of the good. Therefore, general knowledge that an overlapping consensus obtains involves the general knowledge that the shared conception of justice is supported (in some way) by each citizen's good. Reasonable citizens, then, not only accept it as justified; they also want to be the kind of person that acts according to the conception in the political realm. Consequently, each knows that the allegiance of others is not easily shaken by self-interest considerations, nor by other nonpolitical values. They do not tend to get suspicious of each other's allegiance to the conception of justice, even if it involves at times considerable burden to them. Additionally, they judge that others will continue to uphold the scheme even in a scenario where the balance of power in society changes in a way that favors their particular doctrines.

Summing up, then, a political conception of justice is publicly justified if and only if

- (i) All reasonable member of the well-ordered society have carried out full justification of

- (roughly) the same political conception of justice (PL, 387).
- (ii) The political conception of justice is *pro tanto* justified.
  - (iii) Knowledge of (i) is general.

I have described how Rawls thinks of the reconciliation of citizens' motives as it is formed in an overlapping consensus, and how this relates to stability. Public justification, in its turn, relates to stability because it requires the existence of the overlapping consensus and public knowledge of it.

Now, if Rawls wishes to claim, responding to Habermas, that the overlapping consensus adds to the justification of the political conception, he needs one more thing. I have argued in 1.3 that appeal to wide reflective equilibrium is at the heart of Rawls' idea of justification in political philosophy. From the relevant perspective, I claim, it is the only test for conceptions of justice. Therefore, there must be a relation between public justification, on the one hand, and the search for wide reflective equilibrium on the other. Rawls in effect states that, when public justification of a conception of justice is carried out,

All reasonable citizens taken collectively (...) are held in general and wide reflective equilibrium in affirming the political conception on the basis of their several reasonable comprehensive doctrines (PL, 388).

That is, the political conception is the one that strikes them as a more convincing account of justice for a society marked by reasonable pluralism after carefully reflecting on their considered judgements of political justice and comparing the resulting conception with as many alternative conceptions as possible. Since everyone achieves a state of wide reflective equilibrium, it is also general (I will also refer to that state as *full* reflective equilibrium). That affirmation depends on the claim that those three types of justification cover all the relevant criteria for the reasonableness of a conception of justice. That is, *pro tanto* justification provides them with good reasons that they can share with other citizens; full justification assures that their deeper allegiances and their fundamental view of life are not frustrated by political justice; and public justification ensures both that other reasonable citizens in fact share (political) reasons for accepting the conception of justice and that other reasonable citizens' deeper, non-shareable motives will be likewise compatible with political justice. Political liberalism

conjectures that if a conception of justice achieves such a result, then it will be affirmed by each reasonable citizen in wide reflective equilibrium. It follows that public justification yields the best possible justification of a conception of justice<sup>41</sup>.

## **1.6 The role of reflective equilibrium in theory and practice**

In the earlier sections, I provided an overview of what ‘acceptance’ of a conception of justice in the well-ordered society is thought to be like. Now I make a final remark about the idea of a full reflective equilibrium in the well-ordered society. Rawls’ distinction between the point of view of “you and me” (call it the actual point of view) and that of the members of the well-ordered society (call it the ideal point of view) is crucial at this point. The first is made of flesh and blood persons, “here and now,” in non-ideal circumstances. The second does not exist in the world – it arises from a theoretical construction portraying ideal citizens living in a version of a perfectly just society<sup>42</sup>. Distinguishing those points of view will help understanding the different roles appeal to reflective equilibrium plays in Rawls’ project.

From the actual point of view motivation never leaves the scene. We arrive at political argument with both political and comprehensive reasons of justice. In fact, from that point of view political reasons are not detached, but a part of our net of beliefs and attitudes concerning justice. We might separate political from comprehensive reasons, and restrict our reasoning to the first, because we acknowledge the burdens of judgement and its implications. That is, insofar as we have the motive of reciprocity and are willing to justify our conduct to others, we feel the need for a common perspective. Because of our deep disagreements, in part because of the different comprehensive doctrines we affirm, we might want to go to a higher level of abstraction to try and find some common ground. (PL, 43-46).

Consider the order of things in theory construction. Rawls thinks that presenting a conception as detached from anyone’s view in particular is necessary if one wants it to be accepted with shared reasons in a context of reasonable pluralism. As I put in the earlier section,

---

<sup>41</sup> For completeness, considering that public justification presupposes the existence of an overlapping consensus, it also follows that the overlapping consensus does add to the justification of a political conception of justice. That is the response to Habermas.

<sup>42</sup> “A version” because the perfectly just character of a well-ordered society is defined formally as effectively regulation by *a* conception of justice. Of course, when I speak of Rawls’ ideal of society, or his utopia, I’m talking about the well-ordered society of justice as fairness (effectively regulated by justice as fairness), which includes the formal character of effective regulation as well as the substantive character of his specific conception.

he saw in the way democratic political ideals have been continuously institutionalized a promising starting point. Thus, the content of a liberal conception of justice, as a theory, is not only restricted to political values, but also appeals to political values in a special way. That is, appeal is not made to political values as they appear in anyone's moral identity, but to political values as they are publicly expressed in liberal constitutions, the tradition of their interpretation, in the way the law in a democracy is bound to press its claims to individuals, and so on. In a word, political liberalism (the theory) initially cuts political values from the particular reasons why each reasonable citizen endorses them.

Now consider the appeal to wide reflective equilibrium within this theoretical apparatus. As I put in 1.3, search for wide reflective equilibrium is carried out from a first person perspective because that is the only way of assessing the relative strength of one's convictions when judgements and conception are in conflict. Besides, moral considered judgments are not just any kinds of beliefs; they must have practical force, so that working out someone's moral conception will "characterize their moral sensibility" (Rawls, 1999b; 288). One may conjecture about how another's reflection would go, and perhaps impose some constraints of rationality to their perspective, but doing so is an exercise of role-taking – it requires that the one who wears other's shoes assume a first person perspective as well (and not that of a disinterested observer). Furthermore, the ultimate test for any conjecture is people's actual, serious attempt to reach wide reflective equilibrium.

As a consequence, search for wide reflective equilibrium essentially involves someone's perspective – someone's considered judgements which are to form a coherent whole after due reflection –, and it is connected to that person's practical reasons. However, if this is so, appeal to reflective equilibrium cannot serve any role at all in working out the content of a political conception of justice, since its content is supposed to be drawn directly from the public political culture of a democratic society. It is crucial that no reference to particular perspectives be done at this stage<sup>43</sup>.

But appeal to reflective equilibrium is allowed at a second stage of argument, in which political liberalism tackles the problem of stability by conjecturing how, in well-ordered

---

<sup>43</sup> See Daniels (1996b). However, as the author puts, in public reason there might be *prima facie* conflicts between political values the relative weights of which are not always easy to establish. In those cases one is required to strike a balance between political values using the material available in the public culture and its possible interpretations. That Daniels calls "political reflective equilibrium" (164). It should be clear by now the difference between political reflective equilibrium and the thorough self-assessment entailed by wide reflective equilibrium.



society, the conception would figure in citizens' particular deliberations about justice<sup>44</sup>. Here, it drops the freestanding position and faces the full blown ideal, unlocking the ideal point of view. It is from that point of view, Rawls says, that members of the well-ordered society are held together in general and wide reflective equilibrium. It follows that, in the well-ordered society, full reflective equilibrium is an achieved state. Now, achieving that state would amount to reaching an endpoint of the justification programme, which just is the search for shared reasons in wide reflective equilibrium. Its role within the theory would be connected to the problem of stability: showing that the well-ordered society of justice as fairness achieves the deepest and most reasonable form of acceptance.

To wrap it up, at the actual level, search for wide reflective equilibrium is our way of carrying out the justification programme. At the ideal level, wide reflective equilibrium seems to be a generally achieved state – the imagined endpoint of the justification programme. However, as it becomes clear in the second overintellectualization objection (section 2.2), such an imagined endpoint seems to be at odds with the practical condition.

---

<sup>44</sup> Such a theoretical division of labor is perspicuous because it clearly distinguishes two challenges a conception of justice should meet – that of working out a conception that can be affirmed with shared reasons, but which can also gain support of the radically different views that compose reasonable pluralism. And if justice as fairness fares well in both accounts, then Rawls has shown everything that he wanted.

## REFERENCES

- APPIAH, A. **As If: Idealization and Ideals**. [s.l.] Harvard University Press, 2017.
- BENTHAM, J. **An Introduction to the Principles of Morals and Legislation**. Kitchener: Batoche Books, 2000.
- BERLIN, I. The Pursuit of The Ideal. Em: **Moral Disagreements**. [s.l.] Routledge, 2000.
- BUTLER, J.; MCNAUGHTON, D. **Fifteen sermons preached at the Rolls Chapel: and other writings on ethics**. First edition ed. Oxford, United Kingdom: Oxford University Press, 2017.
- COHEN, J. A More Democratic Liberalism. **Michigan Law Review**, v. 92, p. 45, 1994.
- DANIELS, N. Wide Reflective Equilibrium and Theory Acceptance in Ethics. Em: DANIELS, N. (Ed.). **Justice and Justification: Reflective Equilibrium in Theory and Practice**. Cambridge Studies in Philosophy and Public Policy. Cambridge, Massachusetts: Cambridge University Press, 1996a.
- DANIELS, N. Reflective Equilibrium and Justice as Political. Em: DANIELS, N. (Ed.). **Justice and Justification: Reflective Equilibrium in Theory and Practice**. Cambridge Studies in Philosophy and Public Policy. Cambridge, Massachusetts: Cambridge University Press, 1996b.
- DANIELS, N. On some methods of ethics and linguistics. Em: DANIELS, N. (Ed.). **Justice and Justification: Reflective Equilibrium in Theory and Practice**. Cambridge Studies in Philosophy and Public Policy. Cambridge, Massachusetts: Cambridge University Press, 1996c.
- DWORKIN, R. **Justice for Hedgehogs**. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2011.
- GALIŠANKA, A. **John Rawls: The Patch to a Theory of Justice**. Cambridge, Massachusetts: Harvard University Press, 2019.
- GAUS, G. The Turn to Political Liberalism. Em: MANDLE, J.; REIDY, D. A. (Eds.). **A Companion to Rawls**. [s.l.] John Wiley & Sons, 2013.
- KNIGHT, C. Reflective Equilibrium. Em: BLAU, A. (Ed.). **Methods in Analytical Political Theory**. Cambridge, United Kingdom: Cambridge University Press, 2017.

- LADEN, A. S. Constructivism as Rhetoric. Em: MANDLE, J.; REIDY, D. A. (Eds.). **A Companion to Rawls**. [s.l.] Wiley Blackwell, 2014.
- LORD, E. On The Intellectual Conditions for Responsibility: Acting for the Right Reasons, Conceptualization, and Credit. **Philosophy and Phenomenological Research**, v. 95, n. 2, p. 436–464, 2017.
- LORD, E. **The importance of being rational**. First edition ed. Oxford, United Kingdom: Oxford University Press, 2018.
- MILLS, C. W. “Ideal Theory” as Ideology. **Hypatia**, v. 20, n. 3, p. 165–183, ed 2005.
- NAGEL, T. Moral Conflict and Political Legitimacy. Em: HERTZBERG; PIETARINEN (Eds.). **Perspectives on Human Conduct**. [s.l.] Brill, 1988. p. 110–127.
- NEUFELD, B. **Public Reason and Political Autonomy: Realizing the Ideal of a Civic People**. New York: Routledge, 2022.
- NUSSBAUM, M. C. Introduction. Em: BROOKS, T.; NUSSBAUM, M. C. (Eds.). **Rawls’ Political Liberalism**. New York: Columbia University Press, 2015.
- O’NEILL, O. Constructivism in Rawls and Kant. Em: FREEMAN, S. (Ed.). **The Cambridge Companion to Rawls**. Cambridge: Cambridge University Press, 2003.
- RAWLS, J. Reply to Alexander and Musgrave. Em: FREEMAN, S. (Ed.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1974.
- RAWLS, J. Kantian Constructivism in Moral Theory. Em: FREEMAN, S. (Ed.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1980.
- RAWLS, J. The Idea of an Overlapping Consensus. Em: FREEMAN, S. (Ed.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1987.
- RAWLS, J. The Idea of Public Reason Revisited. **The University of Chicago Law Review**, v. 64, n. 3, p. 765, 1997.
- RAWLS, J. **The Law of Peoples**. Cambridge, Massachusetts: Harvard University Press, 1999a.
- RAWLS, J. **A Theory of Justice: Revised Edition**. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1999b.
- RAWLS, J. The Independence of Moral Theory. Em: FREEMAN, S. (Ed.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press,

1999c.

RAWLS, J. **Justice as Fairness: A Restatement**. Cambridge, Massachusetts/Londres: The Belknap Press of Harvard University Press, 2001.

RAWLS, J. **Political Liberalism: Expanded Edition**. Nova Iorque: Columbia University Press, 2005.

RAWLS, J. **Lectures on the History of Political Philosophy**. Cambridge, Massachusetts, and London, England: The Belknap Press of Harvard University Press, 2007.

RICKS, C. **T.S. Eliot and Prejudice**. [s.l.] University of California Press, 1988.

SCANLON, T. M. Rights, goals, and fairness. **Erkenntnis**, v. 11, n. 1, p. 81–95, 1 jan. 1977.

SCANLON, T. M. Rawls on Justification. Em: FREEMAN, S. (Ed.). **The Cambridge Companion to Rawls**. Cambridge: Cambridge University Press, 2003.

SCANLON, T. M. **Being Realistic about Reasons**. Oxford, United Kingdom: Oxford University Press, 2014.

SCHROEDER, M. Believing Well. Em: MCHUGH, C.; WAY, J.; WHITING, D. (Eds.). **Metaepistemology**. United Kingdom: Oxford University Press, 2018.

SEN, A. The Idea of Justice. **Journal of Human Development**, v. 9, n. 3, p. 331–342, 1 nov. 2008.

SEN, A. Dworkin on Ethics and Freewill: Comments and Questions. **Boston University Law Review**, v. 90, p. 657, 2010.

SIDGWICK, H. **The Methods of Ethics**. 7<sup>a</sup> ed. Londres: Palgrave Macmillan, 1962.

SIMMONS, A. J. Ideal and Nonideal Theory. **Philosophy & Public Affairs**, v. 38, n. 1, p. 5–36, 2010.

SYLVAN, K. What apparent reasons appear to be. **Philosophical Studies**, v. 172, n. 3, p. 587–606, mar. 2015.

SYLVAN, K.; SOSA, E. **The Place of Reasons in Epistemology**. [s.l.] Oxford University Press, 2018. v. 1

WAY, J. **Creditworthiness and Matching Principles**. University of Southampton, [s.d.].

WEITHMAN, P. John Rawls and the Task of Political Philosophy. **The Review of Politics**, v. 71, n. 1, p. 113–125, ed 2009.

WEITHMAN, P. J. **Why political liberalism? on John Rawls's political turn.** Oxford ; New York: Oxford University Press, 2010.

WILLIAMS, B. Persons, character and morality. Em: **Moral Luck: Philosophical Papers 1973-1980.** Cambridge, United Kingdom: Cambridge University Press, 1981.

WILLIAMS, B. **Shame and Necessity.** Berkeley, Los Angeles, Oxford: University of California Press, 1993.

WILLIAMS, B. **In the Beginning Was the Deed: Realism and Moralism in Political Argument.** [s.l.] Princeton University Press, 2009.

WILLIAMS, B. **Ethics and the limits of philosophy.** First issued in hardback ed. London New York: Routledge, 2015.

WITTGENSTEIN, L. **Philosophical Investigations.** [s.l.] John Wiley & Sons, 2010.