

Univerzita Karlova v Praze, Filozofická fakulta

Ústav Filosofie a Religionistiky

Obor Filosofie

Disertační práce

Liberalism and its Justification in Contemporary Political

Philosophy

Liberalizmus a jeho ospravedlnenie v súčasnej politickej filozofii

Školitel: Jakub Jirsa, PhD.

2015

Matej Cíbik, M.A.

Prohlášení:

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne

.....

Matej Cívik

Abstract

This thesis attempts to answer one basic question: what we can philosophically say to justify liberalism as a mode of political existence of society. It is divided into three parts. In the first one, I critically survey two popular answers to this question, employing the concepts of self-ownership and value pluralism respectively. I argue that both of them are inadequate and unsatisfactory, mostly because they operate with a conception of person that is too thin for the justificatory task. In the second part, I develop an interpretation of John Rawls and the conception of person he uses. I argue that this conception is crucial with regards to his answer to my question, yet that he provides only a limited and in the final analysis unpersuasive justification for it. The third part tries to remedy the deficiencies of Rawlsian liberalism by providing a better argumentative support for his conception of person and developing from it two arguments aiming to justify liberalism as a mode of political existence of society

Key words

Liberalism – John Rawls – Pluralism – Conception of person

Abstrakt

Táto dizertačná práca sa snaží odpovedať na jednu základnú otázku: čo môžeme filozoficky povedať na obhajobu liberalizmu ako spôsobu politickej existencie spoločnosti. Práca pozostáva z troch častí. V prvej kriticky prezentujem dve populárne odpovede na túto otázku vychádzajúce z konceptov vlastníctva seba samého respektíve hodnotového pluralizmu. Argumentujem, že ani jedna z nich nie je adekvátne a uspokojujúca, hlavne z toho dôvodu, že operujú s koncepciou osoby, ktorá je príliš nerozvinutá na to, aby mohla zaštitovať obhajobu liberalizmu. V druhej časti prezentujem interpretáciu Johna Rawlsa, špeciálne jeho koncepcie osoby. Tvrdím, že táto koncepcia je pre jeho myslenie kľúčová, najmä vzhľadom jeho k odpovedi na moju otázku. V tomto kontexte ukazujem, že Rawls nepredkladá presvedčivú obhajobu tejto koncepcie. V tretej časti predkladám argumenty na podporu Rawlsovho liberalizmu, ako nápravu pochybení, ktoré som identifikoval. Ukazujem, ako môže byť jeho koncepcia osoby lepšie argumentačne podporená. Z tejto koncepcie následne odvodzujem dva argumenty podporujúce liberalizmus ako spôsob politickej existencie spoločnosti.

Kľúčové slová

Liberalizmus – John Rawls – Pluralizmus – Koncepcia osoby

High standards need strong sources

Charles Taylor, *Sources of the Self*

Contents

- Introduction - 5 -
- CHAPTER I - 13 -
 - 1. The paucity of self-ownership - 14 -
 - 2. The role of pluralism - 35 -
- CHAPTER II - 60 -
 - 1. Reading Rawls - 60 -
 - 2. Two justificatory strategies - 64 -
 - 3. The conception of person and its role - 75 -
 - 4. The two interpretations - 85 -
- CHAPTER III - 101 -
 - 1. The argument so far - 101 -
 - 2. Interpreting the question of justice - 103 -
 - 3. Seeing ourselves as normatively relevant political beings - 122 -
 - 4. The uniqueness of liberalism - 139 -
- Conclusion - 148 -
- Bibliography - 153 -

Introduction

This thesis is motivated by what I take to be an important and worrying gap between the beliefs of contemporary liberal-democratic societies and their possible theoretical justification. On the one hand, there is a wider than ever consensus on basic acceptable modes of interaction between the state and its citizens. The orthodoxy of extensive human rights and limited powers of the government is taken for granted and the actions of our societies, both internal and international, indicate a strong and universal belief in basic liberal principles. We are clearly upset when societies all over the world blatantly violate our values – even the ones that we ourselves have come to endorse only a few short decades ago (consider the international reactions to anti-gay laws in Uganda and the fact that gender equality is one of the priorities of Western aid for Afghanistan). There might be a certain level of hypocrisy involved when choosing to mute the criticism of certain countries while amplifying it towards some others, but the overall picture is nonetheless one of genuine confidence in the rightness of our principles concerning the equal political and individual rights of citizens and the limited role of government in the life of political society.

Given our strong political convictions, one would expect them to have a robust theoretical backing. It seems natural to think, as does Charles Taylor, that high standards need strong sources.¹ After all, the liberal standards do put quite demanding and non-trivial burdens upon governments and they are, importantly, built very deeply into the global international architecture as it was developed after the end of the Second World War (for example in Universal Declaration of Human Rights or Helsinki Accords). Nonetheless, the strong beliefs manifested in these international documents do not seem to display an equally strong theoretical support. The rough agreement we observe at the level of political values and principles in liberal democratic societies completely evaporates when confronted with the need to justify them. When we are asked about the basis of human freedom and equality, about the reasons behind our claim to human rights, protection, and a share of political power, the answers we come up with are much less confident and certainly less generally shared than

¹ Charles Taylor, *Sources of the Self: The Making of the Modern Identity* (Cambridge, Mass: Harvard University Press, 1992), 516.

what might be expected or even required given the strength of the practical principles we endorse.

This thesis offers a long answer to one relatively simple question. I enquire *what we can philosophically say to justify liberalism as a mode of political existence of society*. Even though I am going to discuss a fair share of authors along the way, I am not interested in giving a complete historical overview of answering this question. I only present some of the arguments within this field, ones that are or were recently popular and influential. Moreover, this thesis aims to have a certain kind of narrative structure, presenting a unified body of arguments rather than a series of disjointed sections. Therefore, any possible ambition for completeness is overridden by the aim to work organically towards the answer I consider to be the strongest and most persuasive, which is (unsurprisingly) laid out in the last chapter.

In Chapter I, I analyze the attempts to justify liberalism that seem to be both very simple and highly compelling, coming from the proponents of value pluralism and self-ownership respectively. Generally speaking, these philosophers claim that a justification of liberalism is a feature of one relatively simple argument, taking the facts of self-ownership or value pluralism as the starting points. If they are successful, the road towards liberalism is very much straight-forward, as it basically follows from one more or less transparent assertion. However, I examine both of these argumentative strategies and find them wanting. In a gross oversimplification, the main problem for both strategies lies in the radical discrepancy between the stripped-down sources of value they adopt and the complex nature of liberal principles and decrees. At the end of the day, no strategy justifying liberalism can take a form of a simple argument. What is needed is a relatively complex conception of person that can cover liberal principles while avoiding a radical normative *fiat* (an underexplained conjecture of overwhelming normative importance of certain factor/value/principle) that both theories in question seem to postulate.

Chapter II assesses a contemporary political theory that does have the requisite complexity – the theory of justice as fairness by John Rawls. In it, I focus on the factor that the justifications of liberalism mentioned above seem to be lacking – a worked-out conception of person that can explain and anchor liberal principles once the thin layer of value assertions is stripped down. Rawls' conception of person is compelling, although it is openly and unambiguously liberal. Thus, his theory might work relatively well while it is in place, yet it only moves the burden of justification one step further. Now, it is the conception of person that needs to be justified in order for liberalism to be persuasively supported. Unfortunately, I

find that the arguments that Rawls offers in support of his conception of person are either unpersuasive or incomplete. Therefore, the conclusion to the Chapter II is also critical. Rawls does not offer a compelling answer to my question – at least not all of it.

In Chapter III, I confront the problems I identified in Rawls directly and I develop a relatively complex argument in multiple stages. Firstly, I seek to justify the Rawlsian notion of moral personality and its key role in the enquiry about an acceptable political regime. Secondly, I purport to show that a certain form of political recognition is necessitated by our possession of moral personality. In the last stage, I claim that only the *liberal* forms of political recognition are possible in societies of certain type – and that this type is predominant in contemporary world. The arguments I propose should, hopefully, constitute a philosophical backing of liberalism that is both sufficiently robust and argumentatively sound, helping to at least partially fill the worrying gap between the domination of liberalism and its philosophical backing.

Before I start the argumentation proper, I need to elaborate more on the shape of the objective I undertake. The basic question (what we can philosophically say to justify liberalism as a mode of political existence of society?) is unclear in two important respects as the concepts of both ‘liberalism’ and ‘justification’ are too fluid to really fix ideas. Defining justification is necessary in order to make clear what sort of philosophical exercise I have in mind. And liberalism, as every concept used heavily both in philosophy and in everyday life, is very vague, often bordering on incoherence. Therefore, I naturally need to specify my approach towards it. Once both concepts are laid out, I can continue with answering the question itself.

I understand justification very broadly. To justify liberalism is to present it as the right alternative, to defend a claim that liberalism is to be preferred amongst other options how to politically organize a society. However, such preference can come in many varieties. A strong, fact-insensitive, universalistic, and metaphysical justification would present liberalism as the only legitimate mode of organizing every human society throughout the history. On the other hand, a weak, fact-sensitive, relativistic, and political justification would just claim that given the specific context, culture, economy, the state of scientific progress, and geopolitical situation, a liberal political regime is now better for the given society than the alternatives (but this situation might change with any of the circumstances). These two are of course only the extremes and any plausible justification of liberalism would lie somewhere in between. Moreover, the characteristics of the two poles (philosophical and social scientific) can come

apart, so there is no one unified axis that can be used to characterize every possible justification.

Nevertheless, given my ambition to uncover what we can *philosophically* say to justify liberalism, I do not want to come too close to the second option. If a justification of liberalism becomes very fact-sensitive, embedded more in political attitudes and highly contingent situation of a given society rather than in theoretical argument, then there might not be much philosophical to say. Sociology, political psychology, or even game theory might be better equipped to analyse the actual preferences and behaviour of political actors, providing perhaps strong instrumental reasons to adopt liberalism under certain conditions.

Still, I believe there is more to be said about justification of liberalism than that. Liberalism as a mode of existence of political society is not only about the everyday wheeling and dealing, but also about certain values and attitudes that political society stands up for and aims to embody. If this is the case, these values and attitudes do require a philosophical justification. Some argument is needed why they are to be preferred to other, incompatible values and attitudes. Throughout the text, I will be searching for this sort of philosophical justification.

This project understandably faces severe criticism right from the start. The basic worry is best expressed by Richard Rorty, who claims that a sound philosophical justification of liberal (or any other type of) institutions is neither possible nor needed.² In his interpretation of Rawls, the only relevant consideration when thinking about political institutions is the sum of beliefs held by the citizens of given society. In the end, a stable overlapping consensus is the only goal we aim at. Philosophical justifications of values within this consensus are superfluous at best and nonsensical at worst.

However, I believe that Rorty heavily underestimates the power and role of ideas in politics and human thinking generally. As human beings we not only have beliefs, we also want them to make sense. We ask whether the regime we have adopted is just, whether the values behind it are sound, whether we are indeed right in adopting this set of institutions and not some other. The openness towards philosophical inquiry of human agents may, of course, stop short of searching for an ultimate transcendental foundation of their beliefs, but it definitely goes further than simple overlapping consensus. At the end of the day, we cannot do with a justification reducible to Rortian ‘because we believe thusly’. Giving an unmitigated priority to a simple belief and not asking for its foundation or at least plausibility

² See Richard Rorty: “Priority of Democracy to Philosophy”, in: Richard Rorty, *Objectivity, Relativism, and Truth (Philosophical Papers, Vol. 1)* (Cambridge ; New York: Cambridge University Press, 1991), 257–282.

is an insult to the capacity of critical reflection we possess, even though our capacity might be severely limited and conditioned by the environment. To put it differently, even though there are valid reservations towards the enlightenment conception of reason as a universal entity upon which all rational beings can participate regardless of their history and background, this does not mean that every possible philosophical justification of political institutions is out of question. There is a space for practical (political and moral) reason in between hard scientific facts and bare culturally-determined beliefs, a space where we can reason about the best institutions, about values that deserve or not deserve to be pursued, about the limits of state power – or so I believe. If that is the case, then producing a philosophical justification of liberalism may just be possible. To show it will be my ambition in the rest of this text.

When it comes to liberalism, I adopt a view by Michael Freeden, who tried to make sense of vast differences among the different instances of liberalism while still preserving their fundamental relatedness..³ In his ‘morphological’ analysis, he understands ideologies as organized series of political concepts. Defining an ideology with a simple list of values it adopts and promotes is clearly insufficient. As remarked by Bernard Williams, “the schema or matrix attached to a given value concept is not going to achieve much by itself: it is too bare, indeed to schematic.” “For instance, in the case of liberty, it might be that the concern was not to be in someone else’s power. In the case of justice, it might be that of giving each person what he or she should have, or some similar idea.”⁴ In any case, the isolated political concepts are not going to be very effective at specifying the content of any given ideology.

The substance of an ideology is therefore provided by the morphology of concepts it adopts. Freeden calls the process ‘decontestation’. The basic idea is that concepts like freedom are “essentially contested concepts”.⁵ Defining them invites continuous disagreements about boundaries, definitions, and proper use which cannot be settled by empirical evidence, linguistic usage, or logic alone. There is a gap, to go back to Williams, between their very bare and schematic nucleus and their often highly developed use. Ideologies therefore aim to bridge this gap, ‘decontesting’ these concepts by connecting them with further concepts, values or goals, so as to elaborate on their basic concern. For example, liberty can be presented as being strongly connected to property rights, equality, autonomy,

³ Michael Freeden, *Ideologies and Political Theory: A Conceptual Approach* (Oxford: Oxford University Press, 1998).

⁴ Bernard Williams, “Liberalism and Loss,” in *The Legacy of Isaiah Berlin*, ed. Mark Lilla (New York: New York Review of Books, 2001), 93,92.

⁵ W. B. Gallie, “Essentially Contested Concepts,” *Proceedings of the Aristotelian Society* 56, no. 1 (1955): 167–98.

political power, value pluralism and so on, depending on the type of theory one is interested in. Moreover, it can be connected with some of these more strongly than with others. It is only by adopting a full morphological structure connecting liberty with a manifold of core and adjacent concepts that reveals liberty as a fully decontested concept playing a role in a given ideology.

Ideology as a whole thus consists of a web of meanings, a complex structure of many interrelated concepts. Some of them will be very close to the core and, in consequence, treated as very significant and connected to almost every purpose that the given ideology adopts, while others will be peripheral, invoked only occasionally and for specific purposes. However, the structure of any given ideology is not fixed and unyielding. Ideologies as morphological units are open to changes, some prompted by theoretical criticism and advancement, some by cultural and historical changes in the nations and groups cultivating them. Thus, equality, originally a peripheral concept for liberalism, moved ever closer to the core, so that now some philosophers see it as more important to liberalism than liberty itself.⁶ Different exemplifications of the same ideology might have very different morphological structures.

This observation signals a possible problem. How can we distinguish any two ideologies amongst themselves while still preserving the inner relatedness of their (sometimes very different) instances? The problem can be formulated as follows: two instances of the same ideology can be quite far apart (some liberals did not care much about equality while others care about it a great deal), yet two different ideologies may share the same core concepts – both liberalism and socialism care about liberty and equality most of all. Here, the difference is not that the two ideologies believe in the same values, but that these values are differently understood. The values of liberty and equality are decontested in terms of individuality, contractual freedom or property in the hands of a liberal, while they are understood in terms of sociability, meaningful work and welfare for the socialist. The unity of one ideology is thus provided by the strings of connected political concepts similarly understood across its different instances. Even though no one single part of the structure is shared by all liberals, there are family resemblances among parts of their theories. This fits the basic aim of Freeden. Generally speaking, he wants to preserve different ideologies ('liberalism' or 'socialism') as meaningful and usable concepts while maintaining their radical inner diversity.

⁶ Ronald Dworkin, *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985), 183.

Liberalism as a meaningful and theoretically useful concept thus depends on certain basic ('decontested') understanding of its fundamental values. This is the closest we can get to liberal orthodoxy. In line with Freedén's vocabulary, I will call it 'the Millite core'. Henceforth, my use of the term 'liberalism' or 'liberal' should indicate any theory or a family of theories that can satisfy the 'Millite core', ignoring the inevitable subsequent intra-liberal struggles.

When it comes to the content, the Millite core is understood as giving high value to liberty and individuality understood as implying the absence of oppression, broad individual freedom of thought and action, equal rights, general distrust of state as a possible vehicle to actively achieve individual fulfilment and perfection, sufficient institutional stability and democracy.⁷ All of the above seem to stem from an understanding of an inner worth of human beings that deserve institutional protection and a space to develop. It is called 'Millite core' because John Stuart Mill was the first thinker to coherently advance the position and label it as liberal.⁸ In this sense, he is the founding father of liberalism. Exercising a sustained influence, Millian understanding of freedom, individuality, and equality can be found throughout writings of liberal philosophers up until today – and also retrospectively in writings of thinkers that have been labelled liberal *post mortem*, like Locke and Kant.

Given the values listed in the previous paragraph, the most basic contours of what I mean by liberal regime should be clear enough. In the spirit of "the Millite core", a liberal state is giving citizens sufficiently large leeway for self-expression, not harassing them in any way, giving them equal respect, minimizing coercion. It is exercising limited powers and trying to be as accommodating as possible to various forms of life while not adopting some overarching goal for the people to follow.

To put all of the above in one sentence, the quest for a philosophical justification of liberalism consists of trying to show that the above-mentioned values understood in a way characteristic for liberalism ought to guide the basic elements of a given political regime. In the next chapter I analyze the most direct, broad, and potentially the most powerful strategies seeking to do so.

⁷ Freedén, *Ideologies and Political Theory*, 144–155.

⁸ John Stuart Mill, *On Liberty* (Oxford: Oxford Paperbacks, 2008).

CHAPTER I

In this chapter, I introduce two liberal strategies of justification that are very simple and straight-forward. In virtue of this fact, they are the most obvious candidates to answer my basic question. The first strategy is directly deontological. It stipulates that human beings possess certain normative capacities that lead unswervingly towards a liberal state. The crucial term here is self-ownership, which is the political concept of choice for many liberals ever since the publication of John Locke's *Two Treatises on Government*. The basic move is to justify liberal state as uniquely respecting the duties correlative to the rights its citizens have towards their bodies and their property. Schematically, the argument runs as follows:

- (1) Individuals have certain rights based in their relation to their bodies and property;
- (2) These rights are manifested as duties that state ought to respect;
- (3) No illiberal state can do so;
- (4) Therefore, every state must be liberal in order to fulfil its duties.

The second strategy will focus on value pluralism and it will be similarly simple, at least in its basic form. It will try to show that since there are many incommensurable values, we have a reason to adopt a regime that respects this fact and is able to incorporate multiple ways of life and numerous normative commitments. Liberalism, in virtue of the structural tolerance and respect to diversity embedded into the Millite core, is the ideal candidate. To put it differently, the second strategy tries to establish a direct link from value pluralism to liberalism. If there is such a link – and if value pluralism obtains – then liberalism is justified.

Both strategies are similar in that they try to justify liberalism by presenting and elaborating on a single decisive feature. The self-ownership line stipulates that there is some deep-lying normative fact about us humans that favours liberalism as a mode of existence of a political society. The value pluralist line claims that there is a deep truth about the nature of values and their manifestations in human societies that, again, favours liberalism. Their starting points are vastly different – but the two-step arguments their present share the same structure.

However, the conclusion of both sections will be negative – none of the two strategies sketched above work. They are too simple and try to achieve too much with too little. Still,

their dismissal will serve as a platform to launch a more complex justification, perhaps with a better chance to succeed.

1. The paucity of self-ownership

It is an uncontested truth that the conceptual means used by a philosopher are of utmost importance for the resulting theory – they specify its limits and often pre-determine answers to important questions. In liberal tradition, one of the most influential concepts has been ‘self-ownership’. Originating from Locke’s *Second Treatise*⁹, it remained intertwined with the political thought from the 17th century up until today. In this section, I claim that using it as a justificatory tool for liberalism is very unfortunate – indeed, we should abstain from using it altogether. I present arguments undermining its internal consistency, so that talking about the ownership of the self comes close to being a category mistake. Moreover, I claim that the application of this concept completely distorts the landscape of political values, which means that self-ownership ultimately turns against the political insights that motivate it in the first place. With this result, we need to throw away the very last living kernel of the oldest justification of liberalism – Locke’s.

1.1 The appeal of self-ownership

Self-ownership is generally thought to gather under one conceptual roof several fundamental political and moral insights. Historically, liberal political thinking established itself in opposition to absolute monarchies, serfdom, slavery and other forms of unfreedom and political oppression. Therefore, the idea of being one’s own master without anyone having authority over one’s actions and life carries a lot of intuitive weight for this tradition. Secondly, there is an almost universal consensus amongst contemporary philosophers when it comes to a belief in some sort of private personal sphere where a person can do what she wants unless it impedes others. Respecting freedom of this kind is generally considered to be an unproblematic cornerstone of political legitimacy of any state. Thirdly, prohibition of unjustified harms caused by others is one of the fundamental moral decrees. Others are not allowed to injure me (unless special circumstances arise).

⁹ John Locke, *Two Treatises of Government* (Cambridge: Cambridge University Press, 1988).

These considerations are of course very vague – they represent broad intuitions shared almost universally across the Western world and beyond. What they have in common, however, is that they are seemingly very well accommodated by the notion of self-ownership. Why should a person be free to organize her life without anyone having authority to overturn her decisions and without anyone injuring her in the process? Why slavery and other forms of subjection and oppression are wrong? It is because every individual is a self-owner with all the corresponding rights.

Moreover, self-ownership does not get far from regular ownership, so it not only captures our intuitions – it is very clear and easy to grasp. If I own a car, I have a right to decide how and when it is to be used, where it goes, etc. The same goes, *ceteris paribus*, for my life. Also, as no one has a right to use my car without my permission, no one has a similar right towards me as a human being. Self-ownership is simple and does not require any elaborate theoretical structures for explanation – which is a clear and distinct advantage in a complicated philosophical world.

Thus defined, self-ownership can be viewed as a conceptual basis for individual freedom and autonomy. It provides, politically and morally, a border that ought not to be crossed by state power, or anything else. My self-ownership is to be respected regardless of the benefits that interference into my life might bring.

With these credentials, it can be clear why “the thesis of self-ownership has plenty of appeal”¹⁰, why “it is central part of morality in liberal democratic societies”¹¹. Indeed, “the specific motivation behind self-ownership involves the strong interest that I have in running my own life. If the state or other entity tells me when and where I must utilize my abilities—forcing me to produce this or preventing me from producing that—something deep and fundamental is sacrificed.”¹² Self-ownership seems to capture the significance of my relationship to myself, my body, my life, my work. Spelled out like this, its importance can hardly be overstated.

All of the above makes self-ownership a natural cornerstone of justification of liberalism. Stipulating that we are self-owners is, of course, a big normative commitment, but given how well it operates with our basic political and philosophical insights, such a commitment might just be warranted. Within this justificatory strategy, it is first argued that

¹⁰ G. A. Cohen, *Self-Ownership, Freedom, and Equality* (Cambridge: Cambridge University Press, 1995), 70.

¹¹ Carole Pateman, “Self-Ownership and Property in the Person: Democratization and a Tale of Two Concepts,” *Journal of Political Philosophy* 10, no. 1 (2002): 22.

¹² John Christman, “Self-Ownership, Equality, and the Structure of Property Rights,” *Political Theory* 19, no. 1 (1991): 39.

self-ownership is the right conceptualization of our relations to ourselves and to others. Therefore, as a consequence, it should govern our relations within political communities as well. These communities would therefore respect property and focus on protection of individual and his rights. All that would amount to endorsing a form of liberalism very much consistent with the ‘Millite core’.

To use the vocabulary of Michael Freedman, this justificatory strategy claims that self-ownership should become the primary decontestation of the liberal values of freedom and autonomy. This decontestation would give these values robust and intelligible content and would thus facilitate building further morphological relations – for example with regards to individual rights or duties of the state. Having self-ownership as a core liberal concept therefore seems to be a smart theoretical decision all-around, potentially producing a clear and robust justification of liberal political regime.

Nevertheless, admitting that self-ownership may capture some of our strongest political intuitions does not amount to its endorsement. Conceptually, its position is not so strong as to be irreplaceable – it is far from the only way to theoretically express the fundamental insights presented above. Other alternatives may include autonomy, integrity, or freedom, all of which can be – and have been – substantively elaborated without any reference to self-ownership”.¹³

But even though there are multiple concepts and values that can cover a set of intuitions broadly similar to self-ownership, a conceptual decision to use one and not others is very far from being just an innocuous play with words. Although all of them supposedly capture the strong interest I have in leading my own life, they are far from interchangeable. Once presented with contestable issues like redistribution, state perfectionism or the scope of justifiable coercion, self-ownership, unlike the others, tends to give answers in terms of property rights. When making a choice to adopt self-ownership as the starting point of a justification of liberal regime, it is thus not enough just to point towards its broad intuitive appeal – it is necessary to justify the general moral and political prominence it gives to property as the main decontestation of freedom. In the next section, I am going to do exactly the opposite, proposing a few arguments why self-ownership as an idea is dubious from the start exactly in virtue of its focus on property. Later, I am going to question the very intuitive appeal I sketched here. Altogether, the arguments I put forward should substantially erode the potential of self-ownership to produce a credible justification of liberalism.

¹³ In fact, most liberals do not use self-ownership as the basis of their theory, relying on other concepts instead. As Mill never used the concept, it is certainly not a part of “Millite core”.

1.2 The questionable concept of self-ownership

This section is divided into two parts. In the first one I am going to put forward an analysis of self-ownership, its characteristics and background. In the second, I will use this analysis to construct two arguments against self-ownership, aiming to show its incoherence and unsuitability for use in political philosophy.

The nature of self-ownership

What is human beings' relation to themselves? More specifically, is it a property-relation? These are the questions that stand behind the concept of self-ownership and need to be answered if it is to provide a persuasive justification of liberalism. It is quite striking that in the vast literature on self-ownership, they have often been neglected with the proponents rarely going beyond stating an interest in one's unhindered freedom or an opposition towards coercion and harm. However, the attractiveness of self-ownership hinges crucially on whether it can provide a plausible account of individual's connection to her 'self'. If it turns out that self-ownership for some reason cannot capture the nature of our self-relation, it would render the whole concept worthless.¹⁴

To start answering the questions above, it is important to ask in virtue of what features are human beings thought to be self-owners. It cannot be in virtue of having a body and basic preferences regarding one's life. Under this definition, most animals would become self-owners as well, with all the consequences that entails. Animals as self-owners would presumably have rights to the products of their labour, a right to sell their property to someone else, and so forth – which is absurd.¹⁵ Having a body and a preference for avoiding pain is therefore clearly *not* a sufficient condition to establish someone as a self-owner.

But having a body is not a necessary condition for self-ownership either. If we imagine a sort of crude Cartesian dualism where mind can exist independently from body to be true, then it once again seems absurd to claim that the mind in question would somehow

¹⁴ Of course, these questions might not be asked by people who use the concept very loosely, with self-ownership meaning in fact nothing more than vague personal autonomy. However, this usage is heavily misleading, as the term "ownership" does have clear property connotations. In the rest of this text, I will focus on self-ownership as meaning the property in the self – in the sense the concept is unambiguously used by the likes of John Locke and Peter Vallentyne.

¹⁵ I am not claiming that animals cannot have rights – only that they cannot have property rights in the usual sense of the word.

lose its property in itself as a consequence of not having a body. Psychological torture and coercion must breach self-ownership even if the entity in question is (by stipulation) disembodied. Whether any of this can happen in reality is beside the point. What is important for the argument at hand is that self-ownership does not conceptually require the self-owner to be a physical embodied being.¹⁶

So what *does* self-ownership require? It seems that ‘myself’ needs to be a certain kind of entity in order to be self-owned. I am going to argue that as a source of wide normative constraints on others, a self-owning entity must be a person, a *moral agent*. Being a moral agent is both a necessary and sufficient condition for an entity to be a self-owner. And this is so regardless of what specific conception of moral agency a friend of self-ownership adopts. No matter if moral agency lies in possession of practical reason, a character capable of virtue and vice, a personality able to form second-order desires, or a capability to pursue rational goals – all of these (if adopted) can establish a person as a moral agent and therefore a self-owner.

The line of thought goes as follows: all the conceptions of moral agency listed above establish a basis for a unique and normatively significant self-relation that is not present in animals or robots. Consequently, this creates a special sort of relationship I have towards myself, my future and everything I hold dear. It makes my character really *mine*; it makes me who I am. We need to be moral agents to secure this sort of self-relation, otherwise the reflective endorsement of one’s plans and actions would not be possible. Entities without moral agency (wantons, animals, or automatons) do not generate the requisite type of normative self-relation required for self-ownership to emerge. Animals cannot own themselves because their connection to their ‘selves’ is simply not as intimate and special so as to carry the normative significance that self-ownership supposedly creates. Therefore, some sort of moral agency along the presented lines is both necessary and sufficient for the emergence of self-ownership.¹⁷

¹⁶ That being said, it is obvious that our nature as vulnerable embodied beings does play a crucial role when it comes to specific rights and duties that presumably follow from self-ownership. But for now I am not interested in the scope of application of self-ownership – only in its fundamental characteristics.

¹⁷ A note on personal identity: The problems with personal identity are of course straightforwardly relevant to self-ownership – they question the nature of the ‘self’. However, there has not been much written on the connection of these two subjects (with the exception of Edward Feser, “Personal Identity and Self-Ownership,” *Social Philosophy and Policy* 22, no. 2 (2005): 100–125., who argues for a congruence between Thomistic hylemorphism and self-ownership). In my argument, I try to avoid any controversial claims regarding personal identity – with the exception of the claim that the ‘self’ in self-ownership needs to be a moral agent. The conception of the ‘self’ I am using here is thus (as with moral agency) compatible with a wide range of views – self as a moral agent can be a mind that is conceptually separable from body, a brain, a bundle of psychological properties, a full living organism, etc. The point I am making has no bearing on the truth in identity, no

Within the context of self-ownership, the normative significance of my self-relation manifests itself as a duty of others not to infringe upon my person. As a moral agent, I own myself – and therefore others should not interfere. Hence, it is who I am that creates this duty for others. Entities without the requisite self-relation do not oblige others in the same way as I do. Hence, for entities that have this self-relation, the rights stemming from self-ownership are innate – there is no point in time when moral agent acquires the right not to be interfered with. She has it insofar as she is a moral agent.

After the analysis of the features of normative self-relation of moral agents presupposed by self-ownership, it is also important to look more closely on the property part of it. Fortunately, the literature is much more welcoming here. In what follows I will adopt a widely accepted conception of property ownership by Vallentyne, Steiner and Otsuka.¹⁸ They claim that

“Full private ownership of an object consists of full set of the following ownership rights: (1) control rights over the use of the object; (2) rights of compensation if someone uses the object without one’s permission; (3) enforcement rights (to prevent the violation of these rights or to extract compensation owed for past violation); (4) rights to transfer these rights to others (by sale, rental, gift, or loan); and (5) immunity to nonconsensual loss of any of these rights of ownership.”¹⁹

There is one important feature that manifestly shines through all the items on this list – the interpersonal character of property rights. Ownership is established because someone may claim for himself something I consider to be mine. Any assertion ‘I own X’ necessarily carries with it a silent supplement ‘...so nobody else does and therefore they are prohibited from Y.’ This much is clear from the points (2) to (5), but it is necessarily presupposed in the control rights in (1) as well. Ownership rights are held against our fellow citizens and other human beings. It only makes sense to speak of theft or other property rights infringement if a

consequences for the trademark fission and fusion cases. It is only a conceptual point that when I say I own myself, I must have this relation insofar as I am a moral agent. Nothing else matters for self-ownership.

¹⁸Michael Otsuka, Peter Vallentyne, and Hillel Steiner, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant: A Reply to Fried,” *Philosophy and Public Affairs* 33, no. 2 (2005): 201–15. Their definition is nowadays widely cited as the benchmark conception of libertarian ownership. See for example Lea Ypi, “Self-Ownership and the State: A Democratic Critique,” *Ratio* 24, no. 1 (2011): 91–106; Kasper Lippert-Rasmussen, “Against Self-Ownership: There Are No Fact-Insensitive Ownership Rights Over One’s Body,” *Philosophy and Public Affairs* 36, no. 1 (2008): 86–118. Cf. Robert S. Taylor, “Self-Ownership and the Limits of Libertarianism,” *Social Theory and Practice* 31, no. 4 (2005): 468. “Control self-ownership (CSO) consists of the rights of use and exclusion, the power of transfer, and an immunity from expropriation with respect to one’s own body and labour power.”

¹⁹Otsuka, Vallentyne, and Steiner, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant,” 203–04.

person – a moral agent – commits the offence. A hurricane does not steal my car (i.e. infringe my property rights), only a person can do that.

Applying the conception quoted above, it follows that for the last survivor of human race the concept of ownership would be completely redundant. If the only person in the universe claimed an ownership of some piece of rock, this statement would be puzzling if not outright nonsensical. Ownership only makes sense if the object owned can be appropriated or abused by someone else – otherwise the rights listed above are simply inapplicable.

The critique

My first argument draws on the notions of moral agency and intersubjectivity of self-ownership as they were elaborated above. In order to avoid ascribing self-ownership to animals or machines, its proponents must acknowledge that moral agency (regardless of the specific conception adopted) is necessary and sufficient for establishing self-ownership. Only a moral agent can have the unique and intimate self-relation to justify the normative pull of self-ownership.

When it comes to the ownership of things, however, moral agency is no longer sufficient, though it is still necessary. In order to establish something external as mine, there needs to be some mechanism through which I can obligate others to respect my would-be property. This is the point where Kant is in agreement with Hobbes and Hume in claiming that the proper establishment of property (with the requisite rights and duties it creates for others) is only possible in civic condition.²⁰ Only in organized society can others be expected to accept limitations of their freedom created by my external possessions. As a consequence, property claims cease to make sense when I am alone in the universe. The concepts of ‘mine’ and ‘yours’ have significance only when I interact with others – only then can the rights to control, compensation or immunity to nonconsensual loss mean something. The conception of Vallentyne, Steiner and Otsuka quoted above²¹ seems to echo this insight.

However, my self-relation has no need for intersubjectivity. It does not change a bit if I am the last survivor of a nuclear apocalypse. I have a special relationship with myself insofar as I am a moral agent. Other people do not enter the frame. The shape of my argument, then, is quite simple:

(1) Property relations are necessarily interpersonal;

²⁰ Immanuel Kant, *Kant: The Metaphysics of Morals*, ed. Mary J. Gregor (Cambridge: Cambridge University Press, 1996), 6:256.

²¹ Otsuka, Vallentyne, and Steiner, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant,” 203–204.

- (2) My relation to myself is not interpersonal;
- (3) My relation to myself is not a property relation.

The fact that Locke and the whole libertarian tradition failed to distinguish between self-relation and property-relation is not that difficult to explain. We are born and always live in societies, so the perspective of a Rousseauian savage does not come naturally to us. ‘Owning’ a hand and owning a piece of land may seem to be two instances of the same thing. After all, they presumably generate rights and duties that are sufficiently similar. Only when we introduce a world with no other people in it does the difference reveal itself. In this lonely world, my hand is still ‘mine’, it is a part of me – but what is meant by ‘owning a piece land’ is no longer clear.

Of course, if someone cuts my hand off now, it is still a wrong thing to do. But the wrongness can no longer be explained in terms of property – the perpetrator was not trespassing or stealing. We would need some other type of normative claim to explain it.

To conclude, what I proposed here is a simple *reductio* argument. First, I claimed that self-ownership cannot depend on simple body-ownership. What is necessary and sufficient for it is some sort of moral agency. Only a moral agent can have the normatively-loaded self-relation required by self-ownership. In the second step, I presented what is a very commonplace conception of intersubjective ownership consisting in a set of rights. And in the last step, it turned out that the self-relation of moral agent is not intersubjective, so the concept of property is not applicable. Self-ownership thus becomes analogous to a claim that it is 5 o’clock on the sun – it transposes the concept of ownership, in itself intelligible, to a context where it ceases to make sense. Talking about ownership of the self therefore amounts to a category mistake.

This conclusion is concealed by the fact that my personal rights may be partly coextensive with the rights to my property. Nevertheless, there is no direct connection between the two. The mistake of the friends of self-ownership is that they develop a weak analogy between personal and property rights into a strong conceptual bond – which, as I showed, does not exist (at least not given the conception of property they tend to adopt). That is a reason to abandon self-ownership as a concept as well as the justification of liberalism that comes with it.

However, there is a relatively easy way to challenge the argument proposed above. Even though Hobbes, Hume, Kant and most contemporary philosophers (even libertarians,

otherwise very sympathetic to Locke) view property as “consisting of”²² a set of socially enforceable rights, which makes it necessarily intersubjective, that is not the only way to account for the property relations. In Locke’s theory of acquisition, property is naturally established by creating a special relation between me and the thing I (now) own by “mixing” my labour with it.²³ No intersubjectivity is needed at this point. The property rights are created only as an afterthought, as a social manifestation of the bond I already have with my possessions. Since Lockean ownership is not necessarily interpersonal, the first premise of my argument does not hold and self-ownership seemingly retains its coherence on this account.

Of course, there are good reasons why Locke’s account of property fell out of favour with contemporary philosophers, but I will not entertain them here. I only want to claim that self-ownership remains highly questionable as a concept even on this account. To start an exposition of this second argument, let me quote Kant’s critique of Locke from his *Lectures on Ethics*. There he writes:

„Man cannot dispose over himself, because he is not a thing. He is not his own property - that would be a contradiction; for so far as he is a person, he is a subject, who can have ownership of other things. But now were he something owned by himself, he would be a thing over which he can have ownership. He is, however, a person, who is not property, so he cannot be a thing such as he might own; for it is impossible, of course, to be at once a thing and a person, a proprietor and a property at the same time.“²⁴

The argument, reconstructed by G. A. Cohen, runs like this:

- (1) Man is a person;
- (2) Nothing can be both a person and a thing;
- Hence, (3) Man is not a thing.
- But:* (4) Only things can be owned;
- Therefore, (5) Man cannot be owned;
- Conclusion: (6) Man cannot own himself.²⁵

Put in this way, it depends crucially on (4) which is in this context entirely question-begging – it is “pulling a normative rabbit out of a conceptual hat”²⁶ – and that makes the

²² Ibid.

²³ Locke, *Two Treatises of Government*, 1988, §27.

²⁴ Immanuel Kant, *Lectures on Ethics* (New York: Cambridge University Press, 2001), 157 (27: 386).

²⁵ Cohen, *Self-Ownership, Freedom, and Equality*, 212.

²⁶ Ibid.

conclusion unpersuasive. The argument, however, can be reinforced. The crucial bit is to elaborate on why my relation to myself is so dissimilar to my relation to a thing I own that putting them under one label of ‘property’ would become deeply problematic.

The ‘self’ is unlike any property in that there is no moment of acquisition when we supposedly gain ownership of our person – as opposed to our car, house or anything else.²⁷ As I claimed in my analysis of moral agency, self-ownership and the rights stemming from it are supposed to be innate. But there is a blatant contrast between a necessary innate possession and a possible external ownership. So much so that it gives rise to two distinct kinds of relations.

The ownership relation within the Lockean paradigm is importantly unidirectional – when I own something it means that there is a sort of unilateral bond stemming from me to the thing involved. My property is at my disposal, serving as an instrument for my comfort, pleasure, sustenance or any other purposes I see fit (within moral constraints). Yet, while my relation to my possessions is made clear by my status as an owner, the other direction (from the owned to the owner) does not exist at all, as far as ownership is concerned – the objects do not obligate me, they are only there to serve my purposes.

However, when it comes to my self-relation, the simple disponibility disappears and there is no unidirectional relation. I do not own myself to attain some purposes of mine. When I look at the mirror, someone stares back at me. And the fact that this someone *is* me makes this relation necessarily reflexive. If I want to change myself, then the original *I* at the beginning of this statement will be automatically affected as well. Therefore, there can be no instrumentality, no unidirectional use. I cannot use myself for my purposes – the ‘myself’ is always a co-author. To use a terminology from Kant’s ethics, it may be possible to use others as means for my goals but it can never be possible to use myself for my goals – if they are my goals, then I am treating myself necessarily an end in itself, not as means. Owning a car therefore cannot be a model for my self-relation. If it is so conceived, it importantly misconstrues the concepts involved. Ownership implies instrumentality of the one-way relation between me and the object I own. However, when this object *is* me, no such relation obtains.

The issue can be illustrated by a problem posed in Nagel’s *The View from Nowhere*.²⁸ There he analyses a complex interplay between ‘the objective self’ empirically present in the

²⁷ If our selves were treated as originally unowned and only later (possibly) acquired, grave problems would ensue for the friends of self-ownership. See Ypi, “Self-Ownership and the State.”

²⁸ Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1989), 54–66.

world and the circumstance that one of these objective selves is *him*. Taken ‘objectively’, the fact that there is a person in the world named XY is not problematic. But once *I* am the person XY, a complex duality of perspectives emerges: I am capable of perceiving an intersubjective world accessible to everyone – yet somehow I am irreducibly present at one spot, perceiving the world from one ‘subjective’ point of view.

The analogical problem of self-ownership lies in the fact that it perceives human self-relation from the quasi-objective perspective, postulating a one-way connection between me and the thing I own. However, it cannot account for the fact that those two are identical. This identity then creates a complex self-relation which can never be captured in the simple instrumental property structure. It is in this way that we can understand Kant’s assertion that “it is impossible, of course, to be ... a proprietor and a property at the same time.”

Self-ownership is fundamentally flawed as a concept because it postulates a highly reductive relation to my person. I cannot own the very thing I am because there is a complex relationship between me and myself – while self-ownership postulates something crude, instrumental and one-dimensional. Property thus does not and cannot model the human self-relation, which means that the most important desideratum of self-ownership does not obtain. With this argument, Cohen’s (4) ceases to be question-begging. Returning to Kant, a charitable reading would now look roughly as follows:

(1) Ownership of something implies a type of relation X to the object owned (a relation to a thing, instrumental, one-directional);

(2) Man’s self-relation – and indeed his relation to other human beings – is of a different type, incompatible with X;

(3) Therefore, man cannot own himself.²⁹

1.3 The stringency of self-ownership

Even if we altogether disregard the case presented above, self-ownership remains an unattractive starting point to adopt. My next argument for this conclusion will run in two larger steps. In the first one, I will look at some well-known instances of theories that use self-ownership. I will analyse their tendency to put self-ownership on pedestal, allowing it to

²⁹ A short, broadly similar reconstruction Kant’s argument can be found in George G. Brenkert, “Self-Ownership, Freedom, and Autonomy,” *Journal of Ethics* 2, no. 1 (1998): 49–52.

either completely dominate or dismiss other values. In the second step, I will show how this stringency undermines the appeal presented in the first section of this paper, leaving self-ownership as a very dubious concept to adopt.

The Lockean tradition

In Locke's *Second Treatise*,³⁰ which popularised the whole idea in the Western thought, self-ownership constitutes the starting premise of the argument which aims to justify the acquisition of external objects. The argument itself is extremely simple: (1) I own myself, therefore (2) I own my labour, and therefore (3) I own products of my labour (§27). What is interesting for my purposes here is what happens afterwards, as self-ownership quickly gains crucial importance for the main argument of the book. The natural rights to life and liberty become property rights with self-ownership being the “natural” reason why they ought to be protected. In §87 Locke writes: “man being born with a title to perfect freedom...hath by nature a power...to preserve his property, that is, his life, liberty, and estate, against the injuries and attempts of other men.”

The natural drive to protect property in my person and in external possessions then motivates the whole argument aiming to establish the state and justify its coercive powers. It all culminates in §138, where Locke claims:

“The supreme power cannot take from any man a part of his property without his own consent. For the preservation of property being the end of government, and that for which men enter into society, it necessarily supposes and requires, that the people should have property, without which they must be supposed to lose that, by entering into society, which was the end for which they entered into it.”

The quote above, although true to Locke's tendency to write absurdly long and complicated sentences, nicely captures the all-round pre-eminence of property rights in his theory. Self-ownership starts as an innocuous first premise of one argument, but it swiftly becomes much more than that, as it drives the aspiration to protect people's property (their bodies, possessions, and freedom) against unwanted intrusions. Facing the inability to secure a protection of property in the state of nature, people create the government which is then charged with a single big task, a “chief end” – to protect property.³¹ What happens here is that the different things we want from state (personal security, freedom to travel, protection of

³⁰ Locke, *Two Treatises of Government*, 1988.

³¹ §124. Cf. §94 and *passim*.

possessions) are all transformed into a unified set of property claims, under one heading, undistinguished from one another.

It is important to stress that the central conceptual tool that enables the unification of state's objectives is self-ownership. Without it, it would not be possible for my freedom, the unhindered use of my property, and my security to fall into the same category, as enforceable rights (with corresponding duties of others) which follow from my ownership. If freedom and personal security are not property rights, if they hold an independent standing, then the goals that the state aims to achieve necessarily multiply. In sharp contrast, Locke can analyse the role of the state using property rights as basically his only concern.

However, Locke wasn't the only philosopher to maintain such an absolute dominance of property rights. This tendency is very much present in contemporary thinking as well, especially in libertarianism. Robert Nozick's introductory claim, "individuals have rights and there are things no person or group may do to them (without violating their rights),"³² is a prominent example of that. In Nozick, the main drive behind his arguments is generated by his theory of "entitlements" which consist in a set of rights to dispose with my person and my (legitimately obtained) property. As "side constraints", these entitlements provide a line that ought not to be crossed.³³ When it is crossed, as when the government adopts an end-state distributive scheme to take away parts of my income and property, the government in fact assumes a partial ownership of my person – which is clearly unacceptable.³⁴

For Nozick, rival non-property-based rights would be possible only if they operated outside the limits imposed by historically established ownership entitlements. However, *there is nothing beyond those limits*. In a sharp statement, he claims that "the particular rights over things *fill the space of rights*, leaving no room for general rights to be in a certain condition."³⁵ Thus, property rights saturate the normative space, so to say. They provide the proper scope of justice and their protection is the one and only role that the minimal state ought to assume. In other words, all legitimate political demands take the form of entitlements based in ownership rights. The ban on theft and property seizure, freedom to travel, or a protection against assault are all an expression of what I can do with the stuff that is mine – and what limits does it place on actions of others.

Moving away from Nozick, the one perceived insufficiency of self-ownership was its difficulty to deal with the demands of equality. In words of G.A. Cohen, "there is a tendency

³² Robert Nozick, *Anarchy, State and Utopia* (Basic Books, 1974), ix.

³³ *Ibid.*, 28–33.

³⁴ *Ibid.*, 172.

³⁵ *Ibid.*, 238. Emphasis added.

in self-ownership to produce inequality, and the only way to nullify that tendency (without expressly abridging self-ownership) is through a regime over external resources which is so rigid that it excludes exercise of independent rights over oneself.”³⁶ However, there is a huge ‘left-libertarian’ literature now espousing exactly the settlement that Cohen thought was impossible.³⁷ Left-libertarians argue that the distributive inadequacy of self-ownership taken in isolation is mitigated by the shared or common ownership of the world at large. For them, respecting the ownership status of the external resources has radical egalitarian consequences, forcing the redistribution from the better-off to the worse-off in form of compensations, improved opportunities, etc.

It is important to note that self-ownership and world-ownership are here presented as two separate and unconnected principles³⁸ Indeed, the chief difference amongst left-libertarians seems to be provided by radically different conceptions of world-ownership they propose,³⁹ as opposed to them mostly agreeing on the basic contours of self-ownership consisting in the “fullest right a person (logically) can have over herself provided that each other person also has just such a right.”⁴⁰ However, the adoption of two separate principles of self-ownership and world-ownership does not entail the plurality of normative concerns. Indeed, the opposite is the case. Equality is not viewed as some independent value to be pursued – it is a feature and consequence of the entitlements to external resources, firmly rooted in the larger scheme of property rights.

In the end, left-libertarians do an interesting work in modifying Locke’s theory of acquisition so that the results are more egalitarian. However, the absolute focus on property rights as the only relevant normative political consideration remains. This is a feature their theories share with both Nozick and Locke, in spite of ideological opposition and radically different consequences.

Looking at the theories of its proponents in general, self-ownership seems to be a very greedy concept indeed. It does not stop until it dominates the landscape of values and rights altogether – until every right is a property right and every value is based on an ownership

³⁶ Cohen, *Self-Ownership, Freedom, and Equality*, 105.

³⁷ See for example Otsuka, Vallentyne, and Steiner, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant”; Michael Otsuka, *Libertarianism Without Inequality* (Oxford University Press, 2005); Peter Vallentyne and Hillel Steiner, *Left Libertarianism and Its Critics: The Contemporary Debate* (Palgrave Publishers Ltd., 2000); Peter Vallentyne and Hillel Steiner, *The Origins of Left Libertarianism: An Anthology of Historical Writings* (Palgrave Publishing Ltd., 2000).

³⁸ Otsuka, Vallentyne, and Steiner, “Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant,” 209.

³⁹ *Ibid.*, 202.

⁴⁰ Cohen, *Self-Ownership, Freedom, and Equality*, 213.

entitlement of some sort. Even when self-ownership is clearly not pulling all the strings, the principle(s) brought to complement it are of the same kind. For the friends of self-ownership, politics and justice seem to be reduced to the question of who owns what and what that entails.

The gravitational pull

The case of the thinkers presented above is curious. Some of them are manifestly moved by several quite distinct normative considerations (freedom, equality, autonomy), yet they end up stipulating a single monolithic conception of property rights and entitlements. But why is this so? Why the friends of self-ownership tend to promote property rights to the role of master-value nothing can oppose, even if they draw dramatically different implications from it? It seems to me that this is no mere coincidence, for three main reasons.

First, at the heart of the dominance of self-ownership in the theories of its proponents lies a simple fact that self-ownership simply *can* fill the role of the master-value. A surprising amount of our considered judgments about politics and morality can be in some way explained by it. A limited power of government, a broad range of freedoms, a large set of rights, and also equality (as the historically latest addition to the list) can be justified by the reference to self-ownership or to the associated property rights. Self-ownership, once adopted, becomes a Swiss Army knife of political theory, with no other tools required. And even if some additional considerations are required (as with left-libertarian world-ownership), they can be of the same kind as self-ownership, creating the same sort of rights and obligations, answering the same questions (e.g. ‘What should state do to respect everyone’s entitlements?’).

At this point, the considerations of parsimony, the Ockham’s razor, become very relevant. Once a theory can do with only one fundamental normative consideration (and a seemingly simple one at that), it would be unreasonable to unnecessarily supplement it with other considerations, only complicating the outcome and exposing oneself to different sets of objections. Therefore, the only sensible option is to stick with the dominance of property rights ultimately motivated by self-ownership.

The second reason for the gravitational pull of self-ownership is the sheer strength of the principle that cannot be easily mitigated. It rests on my relation to myself. And when it comes to my self-relation, I obviously have a great interest and intense feelings towards the thing that *is* me. My murder, rape, or torture are events I have every reason to avoid and strongly object to. But if the moral background behind my resistance to these unpleasant

events is in fact self-ownership, then the ‘moral force’ of self-ownership and property in general needs to be very stringent and extensive, so that it cannot be easily overridden.⁴¹ A principle that is set to explain my resolute moral rejection of being a subject of torture cannot be a weak one.

However, the problem arises when this strong principle is applied more widely, to the questions of distributive justice, scope of legitimate institutions, etc. Since we presumably apply the same principle here as in the cases of torture or rape, it is difficult to tone it down and relax its application – exactly because of the strength of the initial intuitions driving it. At the end of the day, Robert Nozick’s often-rejected argument equating taxation with forced labour⁴² is exactly a manifestation of this tendency. Taxation equals slavery because they are fundamentally only different instances of the same thing: the injury to self-ownership.

What happens in the forced labour-type arguments can be described roughly as follows: self-ownership, a principle that is driven by very powerful intuitions, is applied to cases where these intuitions are no longer present. However, *the structure of its application seemingly does not change* – so it retains the initial strength over all types of cases, from brutal assault to the mildest instances of property redistribution.

With this framework in place, we can understand why the friends of self-ownership seem so reluctant to introduce non-property-related considerations. It is difficult to imagine any principle that would overpower property rights understood as guaranteeing the very integrity and continuation of human life. As a result, the only forces restraining my property rights are the property rights of other people. That is why left-libertarians needed to define equality as a matter of ownership of common resources in order for it to get a proper hearing in the wider theory. Equality as an independent value would not stand a chance. In Nozick’s words, there is simply no room for it.

The last reason for the domination of property rights in the works of the friends of self-ownership is their very adoption of the language of rights in the first place. The long-acknowledged difficulty with rights-based accounts of justice is the fact that every conception of rights bears problematic absolutist undertones, with limited space to manoeuvre.⁴³ After all, people either have certain rights and corresponding enforceable duties, or they do not. The questions posed thus demand binary yes/no answers. Rights cannot be conflicting and once a

⁴¹ Self-ownership is defined as stringent and extensive in Steven Wall, “Self-Ownership and Paternalism,” *Journal of Political Philosophy* 17, no. 4 (2009): 400.

⁴² Nozick, *Anarchy, State and Utopia*, 169.

⁴³ Jeremy Waldron, “Nozick and Locke: Filling the Space of Rights,” *Social Philosophy and Policy* 22, no. 1 (2005): 109.

right is acknowledged, it is very difficult to compromise on it. Perhaps the clearest example here is the abortion debate. When the controversy is presented as the right of the mother to choose versus the right of the child to live, the two positions enter the trenches and no mutual accommodation is possible, as every party acknowledges one right and denies the other. Similarly, self-ownership proposes a wide-ranging scheme of rights penetrating every aspect of life. Introducing a different and potentially conflicting source of rights would be a theoretically dangerous move, inviting incoherence. It would be equivalent to acknowledging a stringent and extensive right to life, while trying to smuggle in a right to choose as well.

Furthermore, introducing a category of softer, non-right-based claims presupposes that there is some sphere unregulated by property rights (which would assume automatic preference in case both property rights and these non-right claims were present). But given the pervasive account of property adopted, such sphere may not exist. Therefore, the friends of self-ownership are left with property rights as their only tool.

Overall, the stringency of self-ownership is a product of multiple factors. (1) Its universal applicability to almost all issues in moral and political philosophy (a breach of every right becomes an instance of trespassing); (2) Its ability to provide answers to these issues almost single-handedly; (3) Its sheer strength when it is conceived as a moral reason of our rejection of torture and slavery; (4) Its adoption of the unyielding language of rights. When all these combine, there just may not be much space left for other unrelated normative considerations.⁴⁴

The appeal of self ownership reconsidered

In the two sections above I analysed the tendency of self-ownership to pull other values in and transform them in its own image. Now, I aim to show the consequences of this domination. I argue that under a closer inspection the frequently assumed appeal of self-ownership quickly disappears. The reason for this disappearance is the very fact that the liberal values that self-ownership was thought to capture must in the second step transform so that they can be expressed in terms of property rights. If they do not or cannot do that, so that

⁴⁴ In this step of the argument, I have not shown a logical impossibility of combining self-ownership with other values. I have only explained why the friends of self-ownership are often unwilling to do it. And even if they try, the combination of self-ownership with other values is always awkward and unsettled – because there is no convincing way to mitigate the strength of self-ownership. A few good articles appeared recently that exploit this structural weakness, applying it to particular issues: Stephen Wall on soft paternalism (Wall, “Self-Ownership and Paternalism.”) and David Sobel on overall practicability (David Sobel, “Backing Away From Libertarian Self-Ownership,” *Ethics* 123, no. 1 (2012): 32–60.). However, none of these authors attempts to explain how come self-ownership tends to be so strong and uncompromising in the first place.

the priority of property rights conflicts with the values that gave self-ownership its original appeal, one of those two will have to yield.

One of the principal sources of the appeal of self-ownership was autonomy understood as having meaningful options to exercise control over my life. However, many options for meaningful living depend on some sort of collective action that may just prove to be impossible in a society that gives too much weight to inviolability of ownership rights and is otherwise blind to divergent normative considerations. A good example of these activities is building of sport or recreational facilities in residential areas. Even if the process does not require any financial contribution on the part of uninterested citizens, it still modifies the environment that is common, and the building process may be loud, dusty, or otherwise discomforting for a time. In these situations, strong property and self-ownership rights would give everyone a veto, which would often hamper any possible collective action, as it gives disproportionate power to grumpy individuals.⁴⁵ Therefore, self-ownership not only does not entail autonomy, it may also restrain it by blocking some valuable options just because they may in some way include everyone, not only willing participants. Hence, to echo G. A. Cohen, who in *Self-ownership, Freedom, and Equality* develops a broadly similar point, “autonomy no more implies self-ownership than is implied by it”⁴⁶ – and my strong interest in leading a fulfilling life may be frustrated by giving an unrestricted priority to self-ownership.

Given that full ownership rights may prevent us from building facilities that are necessary for modern society and provide valuable options for us (it is not just public swimming pools but airports, factories, highways, or dams), many libertarians feel this practical pressure and they have been seeking for ways to soften their stance. A common strategy consists in allowing the others some unilateral non-consensual influence over my person and my property, while still trying to uphold the core ownership rights.⁴⁷ However, all this is immensely problematic. As I argued above, self-ownership with its built-in tendency to dominate cannot just yield to some other, unrelated considerations – it is too robust for this manoeuvre. I do not have space here to critically analyse the softening of the stance of the various friends of self-ownership in detail, but, luckily, I do not have to. There are several highly persuasive critiques claiming that the mitigating strategies employed by the friends of

⁴⁵ Cf. Nozick’s famous public address system argument Nozick, *Anarchy, State and Utopia*, 93.

⁴⁶ Cohen, *Self-Ownership, Freedom, and Equality*, 238.

⁴⁷ See for example Otsuka, *Libertarianism Without Inequality*, 30.

self-ownership so far have “an air of *ad hocness*.”⁴⁸ I only argue that we should not be surprised – self-ownership does not work in this way and cannot be easily restricted.

Still, since it was established above that the only clear and transparent way to limit property rights stemming from self-ownership is by introducing other property rights, I hold that there is one softening strategy readily available. Turning Nozick upside down, it is possible to claim that the community and state indeed are partial owners of individuals – and that is why they can limit and influence their property. After all, if individuals are not full self-owners, it seems to follow that someone else must own the remaining bit.

I take it that few if any libertarians (or indeed liberals at large) would be willing to accept this conclusion, as it goes against the most fundamental rationale of their position. As a result, they have to live with the fact that self-ownership may importantly limit collective action and valuable life choices. But that means that self-ownership limits autonomy – an important value that was supposed to motivate it in the first place.

The same point holds with regard to freedom. Self-ownership, as presented in the first section, was supposed to provide high-enough barriers against institutions and individuals dominating and controlling my life. However, the pre-eminence of self-ownership ensures that the only type of freedom libertarians focus on is the freedom to enjoy property. All its other occurrences are checked against this core one, and if they conflict with it, they are abolished.

The main philosophical question here concerns a possibility of translation. Libertarianism faces a challenge – it needs to translate freedom-related concerns as property-related concerns. In many core cases, this does indeed work. Assault or rape can be redefined as an especially serious form of property violation and it might even be argued that this redefinition perfectly captures the nature of the act in question. However, there are still certain aspects of freedom that are lost in translation. Especially vulnerable are some of its requirements that are a bit more immaterial – like dignity – when confronted with straightforward property rights. Typically, much of the unfreedom of black population in the American South before the civil rights movement was carried out by free exercise of property rights. The owners of restaurants and transportation companies, the boards of private schools, the large employers all set up rules (written and unwritten) for their property that in aggregate

⁴⁸Wall, “Self-Ownership and Paternalism,” 415–416. Cf. Sobel, “Backing Away From Libertarian Self-Ownership.”

made the African-Americans radically unfree.⁴⁹ The exercise of property rights therefore severely limited what we ordinarily understand as the core aspects of individual freedom.

The same point can be made when analyzing the archetypal example of this debate – slavery. Many libertarians famously believe that a voluntary enslavement is acceptable, since people own themselves and a right to transfer is one of the most fundamental property rights. Here, translating freedom as a property right seems to entail a possibility of enslavement, while this move ignores the more subtle questions of moral status of human being, which fall in between the cracks of property framework. Something is lost in translation, again. Freedom is upheld in a very questionable way, allowing the domination and control of others, opposition to which motivated self-ownership in the first place.⁵⁰

In a sense, my analysis of self-ownership just made a full circle. Strong and persuasive moral and political insights about controlling my own life, my freedom and autonomy, gave birth to a very strong principle of self-ownership. But it turns out that this principle is too strong, so that it cannot make room for values that stand as its core motivation. It has practically subverted its own roots.

Yet this is not a conclusive argument against self-ownership. It is only a simple dilemma. Since it turns out that adopting a strong conception of self-ownership not only does not entail protection of freedom and autonomy but may also go against it, there are two simple options. One can either (1) abandon self-ownership and find conceptual grounds that would better represent the initially attractive values; or (2) play hard ball and claim that if self-ownership can clash with freedom, so much worse for freedom. Put succinctly by Jan Narveson while stating an argument against Cohen: “since you are yourself, and self-ownership consists in having the right to do whatever you want with any part of yourself, up to and including your entire self, *liberty of the person ... (is) nothing more or less than self-ownership.*”⁵¹ And the same claim can be made for autonomy.

Surely, with this manoeuvre self-ownership lost a big part of its attractiveness. ‘Strong interest in running my own life’ becomes ‘strong interest in exercising property rights in my person’ – and those two are very far from each other, as the last paragraphs showed. All the

⁴⁹ In a profoundly racist society, this would be a problem even if the minority got their left-libertarian share of worldly resources.

⁵⁰ For a far more elaborated account of an opposition of the commonplace idea of freedom and freedom understood as a property right, see Samuel Freeman, “Illiberal Libertarians: Why Libertarianism Is Not a Liberal View,” *Philosophy and Public Affairs* 30, no. 2 (2001): 105–51.

⁵¹ Jan Narveson, “Libertarianism Vs. Marxism: Reflections on G. A. Cohen’s Self-Ownership, Freedom and Equality,” *Journal of Ethics* 2, no. 1 (1998): 9. Emphasis added.

different varieties of normative considerations have just been engulfed by property rights and denied proper hearing – even the ones that self-ownership was supposed to cover in the first place. The conclusion of my argument can thus be expressed in form of a conditional: If we believe in any sort of plurality of normative considerations in politics not covered by property rights, then self-ownership, with its built-in tendency to dominate, is not a good concept to adopt and use.

1.4 Justifying liberalism via self-ownership

Had it worked, self-ownership would serve as a great justificatory tool for liberalism. It is widely applicable, and it appears to rule out all the illiberal regimes by its strong protection of negative liberty. However, it cannot fulfil this role. At the end of the day, it wants to achieve too much with one – albeit robust and morally heavyweight – anthropological normative stipulation, that humans are self-owners. But even if self-ownership was fully coherent and if the pre-eminence of property rights was fully accepted, making it a cornerstone of liberal justification would still create massive problems. One of them is that self-ownership as a concept is highly partisan (as is libertarianism, the doctrine that so enthusiastically endorses it). For libertarians, self-ownership often serves as a creed to be adopted, not unlike the articles of faith for the organized religions. Once you start to believe in self-ownership, then the whole doctrine makes sense, the arguments seem forceful and the conception of justice sound. However, viewing it from the outside the doctrine may sound coherent enough – but there is no reason to adopt it in the first place. The justification of liberalism via self-ownership would be therefore conditioned on the acceptance of this highly contentious concept. In this case, the intuitive plausibility I examined at the beginning of this section can get the argument only so far. At the end of the day, what is required (even disregarding my arguments above) is a leap of faith, a questionable stipulation – and that is not a good strategy for a philosophical justification of liberalism.

Of course, many possible justifications of liberalism work with a normatively charged conception of human being – with humans understood as free, equal, autonomous, possessing rights, etc. and with political institutions having some sort of duty to respect this. However, a good justification should not end there; it should try to provide at least some reasons why it is indeed appropriate to adopt specifically this conception of human being and not some other.

Those reasons may exist for freedom or equality – I try to develop them later – but they do not exist for self-ownership.

Self-ownership presents us a picture of human beings that is made to fit into the world that supposedly has only one normative consideration – that of property rights. It is this simple move that creates all the problems I talked about in this section. The world of single normative consideration requires all the different strands of normative thinking to fall in line – which puts a high strain on coherence of the whole notion as it needs the ownership of myself and of my car to belong to the same category. In consequence, the resulting picture will necessarily be warped. As I showed, not all values are easily translatable into a single normative framework. There will be deformations and loose ends. As a result, even though a justification of liberalism with the notion of self-ownership may be argumentatively sound on the surface, it is not a justificatory strategy we can endorse. The concept itself is suffering under the pressure of different notions it tries to tie together – which may be an impossible task in itself. Moreover, the liberalism we get in the end is made to respect the unyielding primacy of property rights, which results in its being unresponsive to non-property related claims. This is a big problem, as many liberals think non-property claims are very much essential.⁵² The liberalism of self-ownership is of questionable coherence, disfigured, and partisan. Taking self-ownership as a part of what we can say to justify liberalism is thus ill-advised.

2. The role of pluralism

Justification of liberalism via pluralism takes a starting point that is importantly different from self-ownership. Instead of postulating a single normatively binding feature of human beings, it starts from social facts and nature of political values. This can be an advantage because such an analysis can aspire to have a strong and factually demonstrable empirical core which would therefore not succumb to the conceptual difficulties identified in self-ownership. Also, pluralism cannot, almost by definition, fall into the kind of partisan cage where so many libertarians seem to live these days – it is more inclusive and therefore *prima facie* better equipped for providing a broadly persuasive justification of liberalism.

The main argumentative move will not be “bottom-up” as with self-ownership (starting with the rights of individuals and moving to the nature of society), but reversed,

⁵² Freeman, “Illiberal Libertarians.”

“top-down”. It will start with society as a medium where values can develop, coexist or compete. This strategy then tries to show how pluralism in our political societies generates a presumption in favour of tolerance and peaceful coexistence, high barriers against oppression, and strong individual freedoms. This would satisfy the Millite core and thus greatly boost the case for liberalism.

In what follows, I firstly focus on Isaiah Berlin. He is the most important champion of pluralism in the 20th century and his contributions are rarely surpassed by more recent pluralist philosophers. However, his free-flowing essayist style leaves much to be demanded in philosophical rigor and clarity. Later in this chapter I will, therefore, use the work of George Crowder and John Gray to figure out the exact contours of the argument from pluralism to liberalism. This will be, as I already announced above, a critical enterprise, as I will strive to show that none of the multiple variations of this argument really work. Still, pluralism offers an undoubtedly interesting perspective on the nature of our political societies – one that cannot be overlooked by any potential justification of liberalism.

2.1 Isaiah Berlin

The main insight that Berlin wants to communicate to his readers throughout his texts is that despite the prevalence of ‘monist’ thinking in the history of human intellectual life, such thinking is not only profoundly wrong but also terribly dangerous. Philosophers from Plato to Kant and theologians from all over the globe believed and believe that there is a single best way to live a human life; that, when it comes to human flourishing, some ways of life are objectively and demonstrably better than all the others. They are the ‘hedgehogs’. Against them, Berlin (as a ‘fox’) is an advocate of far-reaching diversity of our experience that is impossible to constrain by a single philosophical conception arranging all values into one organized whole.⁵³

For Berlin, the most fascinating and important feature of human beings is their capacity for choice, their ability to pick goals, pursue them, and consequently transform their

⁵³ For contrast between the hedgehog and the fox, see Berlin’s powerful analysis of Tolstoy’s *War and Peace* in his essay “The Hedgehog and the Fox” (Isaiah Berlin, *Russian Thinkers* (London; New York, N.Y.: Penguin Classics, 2008), 48–58 and passim). His thesis is that while Tolstoy possesses an unparalleled talent in being able to observe and represent life in its irreducible variety, he is always driven by an ambition to find the one complete overarching explanation. This means that Tolstoy is captured by an inner struggle to find something he knows, deep inside, does not and cannot exist. From this inner struggle, masterpieces like *War and Peace* emerge – but, ultimately, we are better off accepting the world as it is, in its diversity unbridled by vain theoretical ambitions.

conceptions of value and forms of well-being into infinite varieties. No less important is the fact that despite this infinite variety these values still retain their distinctly human quality, so that we are able to perceive and recognize them as such across the differences in historical situation and culture. For Berlin, there exist so many genuine human values that no one life and no one culture can encapsulate them all. To bring up his favourite example taken from Machiavelli, no society and no single individual can embody both the basic tenets of Christian morality and Roman *Virtu*⁵⁴ At some point we therefore have to choose one or the other regardless of the fact that both of these conceptions have their undeniable appeal. We can never have all that is good and our choices involve inevitable loss that can never be compensated by the other good(s) we gain. To use the established philosophical vocabulary, the multiple goods to choose from will often be irreducible to any one common denominator – they will be incommensurable.⁵⁵

Berlin's view does not imply relativism. As he characterises it, "I prefer coffee, you prefer champagne. We have different tastes. There is no more to be said. That is relativism."⁵⁶ Relativism for Berlin does not take seriously the world of human values and ignores the possibility of mutual understanding and appreciation of different cultures. He claims that "there is a world of objective values"⁵⁷ – they are real and exist within human horizon which we all share. However, as I noted above, in spite of our possible broad understanding and appreciation of different ways of life, the world of values is too big for anyone to live them all – and even if it were smaller, some values tend to not only exclude but oppose each other (e.g. justice and compassion). Therefore, combining them all is conceptually impossible.⁵⁸

So, how could monism cope with what is for Berlin an undeniable and irreducible plurality of human goods? Given that Machiavelli's observation of incompatibility seems unbeatable, what must be done for all values to be successfully combined into one comprehensive conception? Such a task would first of all involve a sharp distinction between genuine and false values. It would claim that either some parts or all of the Roman *Virtu* (or Christian morality) are misguided or even plainly wrong – they are not genuine human goods

⁵⁴ Isaiah Berlin, "The Pursuit of the Ideal," in *The Crooked Timber of Humanity* (Princeton: Princeton University Press, 2013), 8.

⁵⁵ There may be some problems with incommensurability as a technical term used by certain pluralists, but there is no space to get to them here. For a comprehensive guide, see Ruth Chang, *Incommensurability, Incomparability, and Practical Reason* (Cambridge, Mass.: Harvard University Press, 1998).

⁵⁶ Isaiah Berlin, *The Crooked Timber of Humanity: Chapters in the History of Ideas*, 2nd edition (Princeton: Princeton University Press, 2013), 11.

⁵⁷ Ibid.

⁵⁸ See also Isaiah Berlin, "The Decline of Utopian Ideas in the West," in *The Crooked Timber of Humanity* (Princeton: Princeton University Press, 2013), 20–48.

they claim to be. Secondly, any possible conflict of the remaining values would be solved by strict priority rules and by assigning each value its proper space, so that all possible incompatibility would turn out to be merely apparent and therefore no cause for concern. What would remain is a complete edifice of goodness/rightness for human beings and their societies, eternal and unmovable.

This enterprise is, for Berlin, an exercise in “throwing a metaphysical blanket over either self-deceit or deliberate hypocrisy.”⁵⁹ It gives priority to vain intellectual self-indulgence over honest and perceptive insights about the nature of human good. The perfect monist society would be lifeless and static. Indeed, “Plato saw it correctly that if a frictionless society is to emerge than poets must be driven out”⁶⁰ – and in this process, almost all that is admirable and valuable in the life of human beings would be thrown away.

It may seem to us now that Berlin overstates his case a bit – but as a Jew born in Riga in 1909, he does not lose sight of the real issue that motivates his thinking. If there is only one right conception of ethics and politics, the one correct way to regulate human affairs, if the ‘final solution’ is possible, then “surely no cost would be too high to obtain it: to make mankind just and happy and creative and harmonious forever.”⁶¹ Indeed, “the search for perfection does seem to me a recipe for bloodshed.”⁶² Therefore, monism is not only incorrect – it is very dangerous as well. For Berlin, the distance between trying to fit radically different values and ideas into one frame is painfully close to fitting different people into one ideal. In both processes, the items in question will necessarily be distorted, deformed, disfigured.

In the end, monism is ‘conceptually incoherent’ because of the deepest character of human beings. In his essay on Mill, which is probably as useful an exposition of his own views as Mill’s, Berlin claims approvingly that Mill “saw that men differed and evolved not merely as a result of natural causes, but also because of what they themselves did to alter their own characters, at times in unintended ways. This alone makes their conduct unpredictable, and renders laws or theories, whether inspired by analogies with mechanics or with biology, nevertheless incapable of embracing the complexity and qualitative properties of even an individual character, let alone of a group of men. Hence the imposition of such construction

⁵⁹ Isaiah Berlin, *Four Essays on Liberty* (London: Oxford Paperbacks, 1969), 171.

⁶⁰ *Ibid.*, 196.

⁶¹ Berlin, *The Crooked Timber of Humanity*, 15–16.

⁶² *Ibid.*, 19.

upon a living society is bound, in his favourite words of warning, to dwarf, maim, cramp, wither the human faculties.”⁶³

Assuming for now that the analysis presented above is roughly correct and we indeed do face the world of plural and irreconcilable choices, Berlin presents us with further political consequences of this position. In the last section of his most famous essay, *Two Concepts of Liberty*, he postulates a deep connection between pluralism and freedom. It seems that for him they mutually imply and entail each other, although this must not be understood in the strict logical sense – because Berlin simply is not that kind of thinker. Maybe the best way to express it would be to claim that Berlin views pluralism and freedom as intimately connected.⁶⁴

The way from pluralism to freedom leads through the value of choice. In his own words, “the world that we encounter in ordinary experience is one in which we are faced with choices between ends equally ultimate, and claims equally absolute, the realization of some of which must inevitably involve the sacrifice of others. Indeed, it is because this is their situation that men place such immense value upon the freedom to choose.”⁶⁵ In an alternative formulation, “the necessity of choosing between absolute claims is ... an inescapable characteristic of human condition. This gives value to freedom...”⁶⁶ Simply put, Berlin wants to claim that if pluralism of objective values is true, our choices really matter. In contrast, this is not the case if values are only subjective or relative – because then nothing deep is lost if we happen to have one set of preferences rather than another. Neither is this the case if some set of values is objectively preferable over the others – that might mean that we are actually better off following some authority that tells us which one is which. The agony of choice only arises if we come to understand that there are no experts, no right answers, no one to take the choice for us – but that our choices are immensely important, as they sometimes radically transform our lives and ourselves. Freedom thus seems to follow from the importance of choices combined with the non-existence of experts and objective rules. This makes the

⁶³ Berlin, *Four Essays on Liberty*, 193.

⁶⁴ In the last 20 years, much has been made about the entailment claim in the literature. Some have tried to rescue it by “completing” Berlin’s argument (see George Crowder: “Value Pluralism and Liberalism”, in: Henry Hardy and George Crowder, *One and the Many: Reading Isaiah Berlin* (Amherst, N.Y: Prometheus Books, 2007), 207–230. Some dismissed it out of hand as incomprehensible and wrong - see Alex Zakaras, “A Liberal Pluralism: Isaiah Berlin and John Stuart Mill,” *The Review of Politics* 75, no. 01 (December 2013): 74–82; Robert Talisse, *Pluralism and Liberal Politics* (Routledge, 2013), 22–23. In the end, taking the entailment claim too seriously is a sure recipe for confusion, as it is too easy a target for criticism, consequently blinding us in searching for a more subtle connection between pluralism and freedom – as analysed below.

⁶⁵ Berlin, *Four Essays on Liberty*, 168.

⁶⁶ *Ibid.*, 169.

individual the best and the right entity to make these choices – and that results in a preference for individual freedom.

The connection in the opposite direction, from freedom to pluralism, is made clear as well. In the course of his argument against monism, Berlin claims: “To assume that all values can be graded on one scale, so that it is a mere matter of inspection to determine the highest, seems to be to falsify our knowledge that men are free agents ...”⁶⁷ Few lines further on he asserts that “(pluralism) is more humane because it does not deprive men, in the name of some remote, or incoherent ideal, of much that they have found to be indispensable to their life as unpredictably self-transforming human beings.”⁶⁸ All this shows how big a role a sort of natural human freedom played in Berlin’s understanding of pluralism. Value pluralism for him obtains exactly because humans are free agents whose actions (and their consequences) cannot be contained into some rigid structure. To paraphrase the quotes above, if philosophers claim that men indeed can be fitted into some overarching moral or political conception, then they clearly misunderstand how humans act and function – they underestimate human freedom and spontaneity, their “self-transforming” nature. In one short sentence, due to the way humans function, pluralism is necessary while monism is deeply misguided.

Getting to the last stage of the argument, the marriage of pluralism and liberty has important political consequences. “Pluralism, with the measure of ‘negative’ liberty it entails, seems to be a truer and more humane ideal than the goals of those who seeking in the great, disciplined, authoritarian structures the ideal of ‘positive’ self-mastery by classes, or peoples, or the whole mankind.”⁶⁹ The connection of pluralism and freedom makes us prefer a regime that tolerates difference and gives people adequate space to express their liberty – a liberal political order. Only this kind of regime respects the fact that there are many ways for humans to flourish and gives sufficient leeway for its citizens to make use of them without any scripted paths and pre-determined ends. Thus, if we want to politically articulate pluralism and freedom, liberalism is the best way to do it.

Even though there is some dispute about how to understand the position presented above,⁷⁰ the prevailing reading is that, if we wanted to take Berlin seriously, the argument would go roughly like this:

⁶⁷ Ibid., 171.

⁶⁸ Ibid.

⁶⁹ Ibid.

⁷⁰ Alex Zacharas, for example, argues that we should completely disregard the last section of *Two Concepts of Liberty* (which provides us with the clearest and most radical statement of the connection between pluralism

- (1) Pluralism is true;
- (2) Pluralism entails the value of choice;
- (3) The value of choice entails a presumption in favour of (negative) freedom;
- (4) Negative freedom entails a (minimally) liberal state.

As a piece of reasoning, this is laden with problems. Indeed, there is more or less a consensus amongst contemporary scholars (even amongst those otherwise sympathetic to Berlin) that this argument does not work.⁷¹ The role of (1) in the overall structure of the argument is complex, but it will be examined only later on – the problems with it resurface better in Crowder and other neo-Berlinians. Otherwise, none of the subsequent steps follow.

The second step makes a mistake, familiar from Mill, in seamlessly jumping from ‘valued’ to ‘valuable’ almost without noticing the difference.⁷² Berlin asserts that pluralism brings about “the necessity of choosing” and he even delves into the psychology/phenomenology of why “men place such immense value upon the freedom to choose”⁷³ – but the argument telling us why choice is valuable as such is nowhere to be found. Given that there are many cultures and worldviews that in fact place no value at all on freedom to choose, and given that under Berlin’s pluralism there exists an infinite variety of different and incompatible things to value with no set of universal common elements, it is unclear how it even may be possible for Berlin to arrive at such conclusion. For this step to work, it would have to proceed from truth about values to universal ‘valuability’ of choice across value systems – subverting pluralism in the process. A value of choice therefore does not follow from pluralism – only a fact of choice does.⁷⁴

As to the third step, even if we valued choice as such, a political conception of negative freedom does not follow. The simple reason is that choice does not occur only in the political framework of negative freedom. Indeed, it would take an implausibly dystopian society, Orwell’s 1984 or North Korea 2.0, to eliminate choice from human life altogether. All other societies give their members options to choose from in a variety of personal or

and liberalism). For him, we would be much better off reading Berlin’s subsequent pieces, especially his essay on Mill. See Zakaras, “A Liberal Pluralism.”

⁷¹ See for example John Gray, *Berlin* (London: Fontana Press, 1995), chap. 6; Gerald F. Gaus, *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project* (London: Sage Publications, 2003), chap. 2.

⁷² For the original Mill’s assertion, see J. S. Mill, *Utilitarianism*, ed. Roger Crisp (Oxford: OUP, 1998), chap. 4.

⁷³ Berlin, *Four Essays on Liberty*, 168, 169.

⁷⁴ This critique is summarized in its general form by Robert Talisse. He claims that “the inference from pluralism to these prescriptions requires some additional premise. The worry, however, is that any such premise will introduce commitments inimical to pluralism. Pluralism may be prescriptively barren.” Talisse, *Pluralism and Liberal Politics*, 23.

political, important or trivial issues. Some freedom exists in every society. Moreover, because Berlin takes the difference between freedom and other values seriously – he does not have a moralized conception of freedom – he holds that laws necessarily restrict freedom even in the best liberal society – their role is to disallow and regulate certain forms of conduct, not to liberate.⁷⁵ Therefore freedom is (and must be) curtailed in every single society – and advanced liberal societies often contain so much rules and regulations restricting the most mundane activities (building a house, selling food) that many actions are simply banned or heavily restricted. As a result, without an additional argument showing why Berlinian negative liberty is somehow special, he must concede that it is on par with other kinds of liberty that can be allowably abolished in favour of environmental concerns, protection of others, justice, equality, or indeed other forms of freedom. No specific political conception of freedom thus follows from a value of choice.

Furthermore, a similar problem can occur here as in the second step. Even if Berlin had come with an explanation showing why negative freedom is somehow special and needs to be respected universally, this would have immediately undermined value pluralism he so vigorously upholds. Simply said, if one value is more important than others, they are not incommensurable anymore.⁷⁶

As to the last step, multiple authors asserted that if pluralism entails liberalism, this must mean that no other type of regime is compatible with pluralism. In this picture, if a state vigorously promotes a specific set of values within its territory (not allowing the openness and a measure of negative liberty typical for liberalism), then it disrespects a plural nature of values. This piece of reasoning, however, is prone to counterexamples. If we can find a particularistic illiberal regime compatible with pluralism, then the whole analysis fails, liberalism does not follow from pluralism, and its justification is nowhere to be found.

As it happens, literature is full of examples of this type.⁷⁷ The main thought is that a particularistic preference for a set of values in a given political context does not entail supreme validity of these values universally – so it does not challenge pluralism in any shape or form. In other words, the fact that values are plural does not exclude a possibility that one specific set of values might be uniquely suitable to one specific context. Israel, for example, clearly promotes Jewish values and Jewish way of life as the best way to live for its citizens (i.e. Jews – putting the acute difficulties with resident Palestinians aside), but it will never

⁷⁵ Berlin, *Four Essays on Liberty*, 148.

⁷⁶ See the quote from Talisse in the footnote above.

⁷⁷ See for example John Gray, *Post-Liberalism: Studies in Political Thought* (New York: Routledge, 1996), chap. 20; Zakaras, “A Liberal Pluralism.”

claim that these values are somehow valid *simpliciter*. In other words, Israel may well endorse value pluralism while still maintaining and upholding its own traditional values – there is nothing conceptually incoherent in that. The same is true for North Korea. Its officials claim that their “juche” ideology is the most suitable form of political organization for their people, without holding any more general thesis. We may (and do) of course disagree with that – but not because it is somehow *apriori* incompatible with value pluralism. And if illiberal regimes can in principle be compatible with value pluralism, then it clearly does not entail liberalism.

Since none of the steps of the argument presented above holds (pluralism does not entail a value of choice and value of choice does not entail freedom) is there a more charitable reading of what Berlin tries to say? Alex Zakaras offers one in his recent article.⁷⁸ He claims that a connection between pluralism and liberalism in Berlin is not so much logical or even philosophical, but *psychological* and comes as a by-product of critique of monism. For Zakaras, Berlin “proceeds in two steps. First, Berlin thought that those who understood other values and ways of life with empathy and imagination would be more disposed to tolerate them. Second, Berlin saw monism as an impediment to such understanding; he thought that value pluralism, on the other hand, tended to encourage it.”⁷⁹ The upshot of the argument is that value pluralists are more open-minded and self-critical and therefore much more inclined to provide inclusion and acceptance. They are thus immune to the dangers of monism and are more in favour of (negative) liberty and liberalism in general.

The argument of Zakaras is decidedly Berlinian and there is no question whether Berlin would accept it – he clearly did. Our capacity of empathy and our insight into the value and relevance of other cultures is Berlin’s main point against relativism, while a refusal to see the worth of different worldviews is an exercise in “either self-deceit or deliberate hypocrisy.”⁸⁰ However, the argument is clearly quite limited, so the question remains whether this is all that Berlin wanted to say. After all, it does not even claim that if a person embraces pluralism, she has a (however weak and first-personal) reason to prefer liberalism – all the weight is pulled by a psychological propensity where empathy and understanding might channel themselves into the preference for inclusive political regime. And even this connection may not work for everyone – there are quite a few pluralists amongst contemporary philosophers who describe themselves as critics of liberalism, even arguing for

⁷⁸ Zakaras, “A Liberal Pluralism.”

⁷⁹ *Ibid.*, 85.

⁸⁰ Berlin, *Four Essays on Liberty*, 171.

a conflict between pluralism and liberalism.⁸¹ In any case, the psychological interpretation of Berlin is very weak and also completely inept as a justification of liberalism – it will therefore not help with the main concern of this thesis.

My understanding of Berlin differs from the positions described above. I believe that the deep connection between pluralism and liberty he postulates is not reducible to either logic or psychology – and (as I already claimed in my sketch of Berlin above) it is definitely not only one-sided. Liberty is not a mere implication of pluralism – it is also the reason why we have pluralism in the first place. At the end of the day, what underlies *both* freedom and pluralism is a picture of human beings as spontaneous and self-transforming. In this sense, humans are seen as naturally protesting against chains of all kinds – and naturally creating a world that we cannot express by a single conception of value. This is the most fundamental conviction of Berlin, one deeply echoing and radicalizing John Stuart Mill and his preoccupation with “experiments in living”.

Therefore, if humans are denied negative liberty or are forced into following a monist worldview, they are de-humanized. Such coercion does not harm them in a mere utilitarian sense – it harms them “intrinsically”, in their very humanity.⁸² Therefore, claims Berlin in a very personal passage, “those who ever valued liberty for its own sake believed that to be free to choose, and not to be chosen for, is an inalienable ingredient in *what makes human beings human*.”⁸³

In the end, the important role that is played by a conception of man in politics and its problems is not wholly surprising. For Berlin, our values are

“...bound up with our conception of man, and of the basic demands of his nature. In other words, our solution ... is based on our vision, by which we are consciously or unconsciously guided, of what constitutes a fulfilled human life, as contrasted with Mill’s ‘cramped and warped’, ‘pinched and hidebound’ natures. To protest against the laws governing censorship or personal morals as intolerable infringements of personal liberty presupposes a belief that the activities which such laws forbid are fundamental needs of man, in a good (or indeed any) society.”⁸⁴

In my reading of Berlin, it is this specific conception of man (as “self-transforming” in an ever-creative and unpredictable manner) that explains pluralism and justifies liberty – we

⁸¹ John Kekes, *The Morality of Pluralism* (Princeton, N.J.: Princeton University Press, 1996); Gray, *Post-Liberalism*.

⁸² Berlin, *Four Essays on Liberty*, lii.

⁸³ *Ibid.*, lx. Emphasis added.

⁸⁴ *Ibid.*, 169.

cannot understand Berlin without it. Yet he still thinks that his conception of man is just one of the many possible. This is a philosophically difficult position to stomach.

Berlin is fully aware that his conception of freedom and its importance does not have any objective standing within the realm of human values – but he still deeply believes in it, he cannot do otherwise. As a result, multiple interpreters suggested that we may in fact be dealing with two distinct Berlins – a devoted radical pluralist and an old-fashioned Enlightenment liberal – with deep, irreconcilable tension running in between them.⁸⁵ However, I do not believe this is a good model of Berlin’s thinking, not least because, as I explained above, the two views seem to be deeply interconnected. I believe that Berlin saw that even given the endless human plurality we cannot but pursue liberty as a political value – because it is the value that resonates the deepest within our political societies and ourselves. Human beings are seen as possessing an inner worth that gives value to their choices and decisions – and that, for Berlin, has political consequences in giving an important place to negative liberty as the most important pillar of the liberal Millite core.

This of course isn’t a satisfactory argument, it is only a personal confession of a man who lived through the 20th century and understood it better than most. At the end of the day, it is not too unlike the view presented in the previous section on self-ownership: it postulates a conception of man as normatively relevant, whose ability to self-transform deserves recognition in form of giving him a measure of negative liberty. No reason to adopt this conception of man was given yet – so a justification of liberalism was not found, in spite of Berlin offering us some truly interesting philosophical insights. However, multiple philosophers took up the challenge and tried to build a better philosophical edifice on these Berlinian insights. We now turn into them in our quest for justification of liberalism.

2.2 Pluralism after Berlin

I am not the only one who reads Berlin’s remarks on a connection between pluralism and liberalism as truly interesting but at the end of the day somewhat unsatisfactory. In this section, I will analyse the arguments of philosophers who, while clearly and explicitly inspired by Berlin, thought and think that his insights and arguments were right in their direction, but often wrong in execution. Looking at Berlin’s work, such an observation can be

⁸⁵ Gaus, *Contemporary Theories of Liberalism*, 49. See also George Crowder: “Value Pluralism and Liberalism”. In: Hardy and Crowder, *One and the Many*, 207.

made relatively easily – he manifestly shunned “philosophy” in favour of doing “history of ideas”. What that actually meant is that he was going great philosophy – but doing it very differently from the industrial standards typical of political theory. Thinkers like John Gray and George Crowder thus think that some aspects of Berlin’s thought can be greatly boosted by removing the inconsistencies and adding more rigour and better interaction with contemporary scholarship, thus creating on the whole a clearer, more defensible theoretical position. Crowder and Gray both want to provide an explanation and justification of a certain type of political regime starting from Berlinian positions, answering my question in the process. Therefore, my analysis of value pluralism is not complete until I examine what they have to say.

The central issue in their thinking is, unsurprisingly, occupied by the interaction between pluralism and freedom, although their approaches differ sharply. Crowder shares Berlin’s preoccupation with freedom understood negatively and individualistically. He therefore wants more or less to “complete” Berlin, rather than transform his thinking. In contrast, Gray is an explicit critic of expansive universalist liberalism and wants to argue for a much more open conception of legitimate political regime. Still, in his argument he acknowledges the need to uphold certain minimal criteria for political societies that correspond to some sort of liberal political minimum. I analyse Gray first.

Gray’s second face

John Gray, perhaps the most influential living philosophical pluralist, is much more radical in his critique of liberal orthodoxy than Berlin, steering away from the tradition of Locke and Mill (which so often focused on ‘negative liberty’), back to Hobbes.⁸⁶ He claims that given the irreconcilable pluralism of contemporary societies, no ambition trying to prove a universal validity of one value (or a set of values) can ever succeed – even if that value is negative liberty. The best we can hope for is a peaceful coexistence between different elements of society – a social peace, a *modus vivendi*.

Since values necessarily conflict, the task of politics is to ensure that this conflict does not result in bloodshed, civil war and, at the extreme, a total disintegration of society (with a life of an individual being solitary, poor, nasty, brutish, and short). With this task in mind, no universal set of necessary and sufficient conditions for political legitimacy can exist. Every potential solution to a given conflict of values and interests will be local. The right solution,

⁸⁶ My analysis will focus mostly on two books: John Gray, *The Two Faces of Liberalism* (Cambridge, UK: Polity Press, 2000); Gray, *Post-Liberalism*.

the right form of government, is the one that works best in a given context, with all its quirks and particularities.

However, Gray does not want to assume that *modus vivendi* is an ultimate value that we universally ought to pursue. He is unequivocal in claiming that values are plural with no normative conclusion (not even social peace) entailed. Nevertheless, he claims that some level of peaceful coexistence is necessary for pursuit of a great majority of all the different conceptions of good, whether they in fact value *modus vivendi* as such or not. Echoing Hobbes, he holds that social peace is a necessary condition of art, industry, and almost any form of complex human cooperation. Therefore, no matter what conception of good we adopt, *modus vivendi* gains great instrumental value. The fact that it may be temporarily trumped under extreme conditions by other considerations does not change its importance.⁸⁷

Gray therefore discards (1) the search for necessary conditions of political legitimacy in favour of finding (2) a *modus vivendi* as a realistic precondition for a pursuit of almost any good in the context of human society. For him, these are the two faces of liberalism. While the former is, in his view, normative and inevitably inimical to deep pluralism of human societies, the latter is minimal in its assumptions, empirical and highly plausible. The first face of liberalism is represented by Locke, Kant, or Rawls and has been dominant in the West almost ever since the inception of liberalism. However, its main ambition, to find a single best way for human societies to live, is doomed from the start by pluralism deeply ingrained within the values themselves. In the end, whatever set of values the first face of liberalism promotes is either vague or conflicting.⁸⁸ The second face, established by Hobbes but dormant for a long stretches of time, does not have these shortcomings and it is thus perfectly suitable for conceptualizing the political life in fragmented and multicultural societies of the late 20th and early 21st century.

Gray is heavily critical of Berlin in the moments where he seems to be endorsing the first face of liberalism, deriving freedom from pluralism.⁸⁹ His pluralism stays closer to the empirical surface of factual disagreement of values held in human societies. Thus, his project can be viewed as a more modest version of Berlin's. He abandons the attempts to derive some normative content from pluralism and is therefore happy to endorse even illiberal societies, as long as they can maintain a decent-enough *modus vivendi* incorporating a presumption against human suffering and a plea for tolerance and decency in social and interpersonal relations. He

⁸⁷ For the full argument see Gray, *The Two Faces of Liberalism*, chap. 4.

⁸⁸ For his analysis of Rawls, see *Ibid.*, chap. 3.

⁸⁹ Gray, *Berlin*, chap. 6.

views these features of *modus vivendi* as general maxims for political conduct. Value pluralism may not justify liberalism for Gray – but it does, through our endorsement of *modus vivendi*, justify basic humanity. And that may be good enough, even as an answer to my question.

However, Gray's *modus vivendi* is greatly unstable, bordering on incoherence. He claims that, unlike Berlin's endorsement of negative liberty, his reliance on social peace is not a value in itself as it merely facilitates a pursuit of almost any human good. But a harder look reveals that Gray's argument is more complicated than that. He needs *modus vivendi* both (1) to entail a level of toleration and decency in political conduct, and (2) be an (almost) necessary condition for a pursuit of human good (and therefore instrumentally valuable). In isolation, both of these claims can be more or less defensible, but they quickly fall apart when put together.

Analysing (1), it quickly becomes apparent that Gray's concept of *modus vivendi* is quite expansive. In the most radical statement he claims:

"In contemporary circumstances, all reasonably legitimate regimes require a rule of law and the capacity to maintain peace, effective representative institutions, and a government that is removable by its citizens without recourse to violence. In addition, they require the capacity to assure the satisfaction of basic needs to all and to protect minorities from disadvantage. Last, though by no means least, they need to reflect the ways of life and common identities of their citizens."⁹⁰

The thought behind this assertion is that any regime disrespecting some of the above will alienate some segments of population and an internal conflict will quickly ensue. We want to eliminate mutual grudges so that citizens can effectively cooperate and be empowered by a truly representative government, thus preventing possible revolution or civic strife. Only in this way a true and long-term *modus vivendi* can be achieved.

Multiple authors have noticed that most if not all crucial liberal positions have been smuggled into Gray's *modus vivendi* – tolerance, representative democracy, or rule of law most prominently.⁹¹ The easy argument here is that Gray is doing exactly the same mistake he so expertly uncovered in Berlin – promoting, under the value pluralist pretext, some values as more important than the rest. However, we must not forget that *modus vivendi* is supposed to have an independent grounding as an instrumental condition of possibility for a pursuit of

⁹⁰ Gray, *The Two Faces of Liberalism*, 107. A broadly similar analysis of "a living kernel of liberalism" can be found in Gray, *Post-Liberalism*, 314–320.

⁹¹ Matthew J. Moore, "Pluralism, Relativism, and Liberalism," *Political Research Quarterly* 62, no. 2 (n.d.): 244–56.

human good – and Gray can escape the charge that he arbitrarily promotes some values over others if he can vindicate this claim.

However, the problem with *modus vivendi* as a condition of possibility for a pursuit of human good (2) when contrasted with a *modus vivendi* as entailing representative government and rule of law (1) is that the former is considerably more modest. The second claim encapsulates a Hobbesian thought that for human society to function at all, some level of cooperation is necessary. When this is achieved, the members of society are able to pursue many conceptions of good (the notorious Hobbesian list of industry, culture, navigation, use imported commodities, commodious building, knowledge of the face of the earth, arts, letters, society⁹²). However, the ideal cooperative *modus vivendi* envisaged by Gray is very far from a Hobbesian minimum. To illustrate, if we imagine a relatively cruel slave-owning society that is stable and peaceful, we can plausibly claim that some sort of *modus vivendi* has been achieved – albeit one that allows one group to be almost completely dominated by others. But notwithstanding the slavery and cruelty in this society, it is surely possible to pursue valuable goods for some portion of its population. Indeed, many miracles of human civilization, both artistic and scientific, have been achieved in this sort of society by people who could devote their lives entirely to one task because all the other duties were performed by their servants and property.

A slave-owning society is cooperative but not humane. It allows for pursuit of complex values but does not require its members not to be cruel and tyrannical. Therefore it fulfils (2) but falls very much short of (1) – revealing a deep inconsistency within Gray's concept of *modus vivendi*. Simply said, Gray puts forward an argument for *modus vivendi* as (2) – a necessary condition of cooperation. However, in the rest of his argumentation he uses much more demanding (1) – *modus vivendi* as requiring tolerant and inclusive society. All this creates an enormous justificatory gap in Gray's argument, completely undermining his positive argument for humanity and toleration.

However, Gray could still say that the ruling class in my example still does have a value-independent, instrumental reason to empower the slaves. If it refuses to do so, it runs a high risk of slave revolt, war, and subsequent turmoil. But, conceived as a purely instrumental reason, this is very weak. The cynical truth is that all human societies end by some sort of catastrophe, often helped and aggravated by their own citizens. And some cruel slave-owning societies proved to be remarkably stable, spanning centuries. At the end of the day, an

⁹² Thomas Hobbes, *Leviathan* (Oxford: Oxford Paperbacks, 2008), chap. 13.

inclusive society is no guarantee for stability, just as a cruel slave-owning society is no automatic guarantee for turmoil. We do not have an instrumental reason to uphold Gray's *modus vivendi*.

To save his position, Gray may (and on certain reading actually does) add some measure of welfarism into his argument. This step would enable him to argue against cruel societies much better than a merely instrumental argument ever could. He may claim that these societies produce suffering and systematically upset the conceptions of good of its members – and therefore are inhuman and wrong. However, this sort of vague gesture towards welfarism does not and cannot count as a serious philosophical argument without an actual theory specifying what does and what does not constitute human welfare, which Gray does not offer. Moreover, any such theory would (again) run a risk of conflicting with his point of departure, value pluralism. Therefore, his defence of humanity in politics via *modus vivendi* is ultimately unsuccessful.

As a conclusion, it turns out that nothing of importance seems to follow from Gray's pluralism. Even his considerably lowered ambition, targeting to justify not liberalism but basic decency, has failed. In a sense, this is much less satisfactory than the justification of liberalism that Berlin aimed to produce. Next, I examine George Crowder, who is the most direct heir to Berlin's position.

Crowder and the role of pluralism

It is not too big an overstatement to say that the justification of liberalism via value pluralism constitutes a life-project of George Crowder – since 1994 up until today he has published roughly dozen papers, book chapters and a monograph on the topic.⁹³ Initially sceptical, he later developed his own argument from value pluralism to liberalism via the notion of diversity. The argument itself tries to stay relatively close to Berlin, focusing on individual pursuit of good. However, in contrast to Berlin, its ambition to justify liberalism from value pluralism is very much explicit and taken as a quest to produce a clear-cut piece of reasoning. In what follows, I analyse only the latest (and, in my opinion, the best) iteration of this argument,⁹⁴ ignoring some interesting developments and omissions with regards to the previous versions.

⁹³ The latest contributions include: George Crowder, *Liberalism and Value Pluralism* (New York: Continuum, 2002); Hardy and Crowder, *One and the Many*; George Crowder, "Two Concepts of Liberal Pluralism," *Political Theory* 35, no. 2 (2007): 121–46; George Crowder, "Value Pluralism, Diversity and Liberalism," *Ethical Theory and Moral Practice*, forthcoming, 1–16.

⁹⁴ Crowder, "Value Pluralism, Diversity and Liberalism."

Crowder's argument justifying liberalism runs in two steps: from pluralism to diversity and from diversity to liberalism. Pluralism is defined as a belief that there is a great variety of possible human goods, manifesting themselves in different and incommensurable values and ways of life. Individuals and societies thus necessarily and inescapably adopt and encourage certain genuine human goods, while rejecting or ignoring others.

The step from pluralism to diversity starts with a claim that we indeed can compare different societies and lifestyles, in spite of incommensurability. Even though qualitative contrast may not be possible, a quantitative measure is still attainable. Here, Crowder quotes an argument by Bernard Williams: "If there are many and competing genuine values, then the greater the extent to which a society tends to be single-valued, the more genuine values it neglects or suppresses. More, to this extent, must mean better."⁹⁵ Crowder understands this quote as claiming that we have a reason to adopt all genuine human goods as they do possess real value. In practice, however, we cannot pursue them all even though they have, in abstract, an equal claim on us (in being genuine human goods). This introduces "a presumption in favour of attention to a greater multiplicity of values."⁹⁶ Thus, a preference for diversity is a result of us trying to accommodate as many competing and equally compelling claims as we can.

This picture is for Crowder less acute when it comes to individuals, as they can be excused (as conditioned by specific circumstances such as talent and opportunity) to pursue "lopsided", single-valued lives. His examples are that of an ingenious scientist, a devoted artist, and a captain of industry. The presumption in favour of diversity therefore gains its full force only at the social level, where every society should aspire to accommodate as many goods and ways of life as possible.

One additional qualification needs to be taken into consideration, though: coherence. Crowder acknowledges that a society that tries to pursue great many disparate conceptions of good can be too scattered to be successful – disintegration becomes a real possibility. Every society must therefore limit its pursuit of diversity by a "tolerably coherent package of social values."⁹⁷ Pluralism therefore requires a pursuit of diversity qualified by some level of social coherence, so that it becomes feasible and not self-destructing.

The second step, from diversity to liberalism, consists in showing that liberal regimes are very good at balancing competing claims, especially when it comes to harmonizing

⁹⁵ See Williams' introduction to Isaiah Berlin, *Concepts and Categories: Philosophical Essays* (London: Pimlico, 1999), xvii.

⁹⁶ Crowder, "Value Pluralism, Diversity and Liberalism," sec. 2.

⁹⁷ *Ibid.*

different competing values (like freedom and equality). Liberal establishment thus provides powerful enough institutional tools to achieve significant diversity while maintaining the necessary social coherence. Crowder claims that, historically, no illiberal regime has provided this degree of flexibility and stability at the same time. Therefore, the presumption in favour of diversity leads directly to the adoption of liberalism as the best possible means to this end. And since diversity follows from value pluralism (as was showed in the previous step), then we can say that liberalism follows from value pluralism.

Here, I will not question the second step of the argument (from diversity to liberalism). It seems to me that once we have a strong presumption in favour of diversity (understood as a multiplicity of options for individual pursuit of good within a functioning society) then liberalism follows more or less unproblematically. It is establishing this presumption that constitutes the hard part – and in the end an undoing – of Crowder’s endeavour.

Prima facie, the most questionable move in Crowder’s argument goes from a multiplicity and incommensurability of values to the conclusion that these values then have an equal claim on us – and we therefore should value them equally. Here Crowder does not have in mind any sort of numerical equality, which would be clearly implausible. He claims that we need to acknowledge that human goods are, in being genuinely valuable, somehow “in the same league”.⁹⁸ Thus, we should value them roughly equally and therefore recognize that having more of them is better. And this recognition would establish a presumption in favour of diversity.

As I read it, Crowder adopts a version of a principle of insufficient reason. If I have five different roads ahead of me and I know that one might lead me to 100 dollars – and I do not know which one is it – then the most reasonable thing to do is to attribute a roughly equal value to every road ahead of me and then try as many as I can. Similarly, if a society has no way of comparing the possible instances of human good it can allow or promote, the best course of action is to attribute them roughly equal value and then accommodate as many of them as possible (as basic social coherence allows).

Despite some theoretical appeal, this scenario is utterly impracticable. It invites us to imagine an abstract society with no prior value commitments that contemplates how many instances of abstract and (roughly) equally valuable human good can it absorb. This is analogous to famous Sartre’s radically situated existentialist individual, by nature detached from all values, who completely invents herself by radical choice. I do not believe that such

⁹⁸ *Ibid.*, sec. 3.

an individual can ever exist – we all have prior commitments or at least inclinations of certain sort. On societal level, however, this is doubly implausible. Any real society consists in individual human beings who already live in a certain way and believe in certain things. Therefore, in practice, all different forms of human good are never in the same league. No society cares about a diversity of “human goods” *simpliciter*, they are always already somehow specified.

To emphasize my point on Crowder’s own example, he illustrates the power of liberal societies to balance competing goods (and thus preserve diversity) by analysing an ongoing struggle between freedom and equality. None of the two are endorsed without qualification yet both of them find their place in our societies as they facilitate different human goods. We might not think that they should command exactly the same respect, but we agree that they are “in the same league” with regards to social importance. However, this example falters with almost any other value. We can all agree, for example, that beauty is an important human good, one that should command some level of social importance. But is it in the same league as equality? No it is not, far from it. Practices that state may develop to bolster equality (like heavy taxation and large redistribution schemes) could never be adopted to promote beauty in whatever form. Even though we may agree that equality and beauty are incommensurable and none of them commands universal precedence, the notion that we should, because of this, value them equally does not make sense – especially as the social “we” that I am using here always reasons with the background of some already adopted values and ways of life.

The abstract scenario of all the goods having roughly the same appeal for some strange society is completely fictitious and therefore cannot have much use in promoting diversity in a way Crowder wants it to. There is no reason for any society to adopt the reasoning that he proposes, unless it is truly “unencumbered” – and no society ever is. This means that there is no reason for us to value different human goods equally.⁹⁹ ‘More’ does not mean ‘better’ and the presumption in favour of diversity does not follow from value pluralism. This attempt to justify liberalism thus does not work.

⁹⁹ To clear a possible misunderstanding: In my rebuttal of Crowder I do not claim that no abstract thought experiment can help us determine what values our society should endorse. Quite the contrary, I do believe that Rawls’ original position is an extremely helpful tool in this respect. However, Crowder’s argument is much shallower than Rawls’. Crowder goes directly from abstractly equal value to concrete value for any given society. Rawls, on the other hand, makes it very clear that his construction of the original position is no merely abstract metaethical construct but does embody certain values he finds important for justice and fairness. But all this will be discussed in detail in the next chapter.

2.3 The trouble with value pluralism

While the last subsections were mostly engaged in separated and seemingly unconnected disputes of single arguments of given authors, this one takes a more systematic view on the problems associated with using value pluralism as the main building block of liberal justification. The basic question, partially anticipated in my analysis above, concerns the role of value pluralism in the liberal justificatory exercise. Interestingly, despite a broad agreement on how value pluralism actually looks like, its status in the argumentation has rarely been discussed. This has led to confusions which I would like to put forward in the next paragraphs, listing the different roles value pluralism could possibly take in justifying liberalism.

Maybe the theoretically most straight-forward way to view value pluralism is as an *empirical thesis* about what conceptions of good do people pursue or have historically pursued. In certain moments, both Berlin and Gray seem to adopt this attitude. Its strength consists in adopting a social-scientific perspective with robust factual evidence that is all but indisputable – people did and do pursue many incompatible forms of good. It remains unclear, however, whether anything philosophically interesting follows from this thesis. Robert Talisse, who calls this variation of pluralism “manyism”, claims that it cannot, as it is strictly descriptive.¹⁰⁰ I disagree with such a sweeping statement. The acknowledgement of manyism gives an additional burden to any proponent of value monism – she must now come up with at least some form of error theory, showing us how all the different generations and cultures did not stumble upon the answers proposed by her now and, consequently, were simply wrong about the human good. It is this argumentative strategy that Gray uses to criticize (monistic) liberal universalism as unable to show how it can be the only right option.

Nevertheless, the added critical burden for monists is certainly not enough to establish anything positive for pluralists. If manyism is a social-scientific attitude, it serves for observation, not for normative assessment. In a sense, this was quite apparent in my analysis of Gray. To argue for his political conception of minimal human decency (which turned out to be not so minimal after all), he must in the end adopt some sort of welfarism on top of his value pluralism. However, with no deeper conception of good present, no developed conception of person, his welfarism remains underdeveloped and on the surface. If human

¹⁰⁰ Talisse, *Pluralism and Liberal Politics*, 12. Talisse has his own list of distinct forms of value pluralism (and non-pluralism). However, he asks a completely different questions than me – he is not interested in the roles value pluralism can play in a justificatory argument, but in forms that value pluralism can (and cannot) take when it is contrasted with monism. Therefore, his list is very unlike mine.

well-being is not the ultimate value (and Gray is no utilitarian), then Gray can use nothing but a very rough intuition to manage possible trade-offs, even for actions that harm what is for him the basic human decency. Therefore, pluralism as manyism may be an interesting critical tool to be used against some more ambitious theories, but it is very much insufficient to establish any positive case for minimal political constraints, not to mention full-fledged liberalism. Here, I must agree with Talisse in claiming that this kind of pluralism is prescriptively barren.

Still, manyism on the social-scientific level may lead to a distinct *first-personal attitude* of value pluralism. Here, the individual would acknowledge that there are multiple ways for her to prosper – and even more ways are open to others. The simple impersonal manyism is supplemented by empathy and understanding of different cultures and conceptions of good, all viewed as genuine and valuable. This strand of first-personal ethical insight is very much present in Berlin, so much so that Alex Zakaras, in the article analyzed above, contends that Berlin’s connection of pluralism and liberalism leads only through this inner appreciation of diversity.¹⁰¹

The problem here is that an empathetic understanding of different forms of human good may (still) not generate any prescriptions and may not be applicable to society at large. As to the first point, we might ask whether empathy towards diversity is universally valuable and whether it is to be demanded from everyone. In other words, we ask whether someone who is not empathetic and only follows her conception of good is doing something wrong. Of course, some wrong is certainly done in cases of coercion or intolerance, but that is not the point here. The point is whether we can demand from everyone a perceptive appreciation of different cultures and individuals. We certainly cannot, for multiple reasons. First of all, requiring not only liberal political regime but also liberal personal attitude from everyone is much too demanding. Generally speaking, emotions are a very problematic subject of obligations. It is unclear how I can demand someone to feel a certain way towards someone else. Secondly, if anything like ‘competitive pluralism’ proposed by Joseph Raz is true,¹⁰² then the pursuit of one type of good (like devotion) makes us naturally apprehensive towards certain other types of good (like libertine freedom) – and vice versa. In these cases, the open-minded empathetic insight is ruled out by the very nature of the good in question. Therefore, even though some people can exhibit understanding and appreciation to diversity in all its

¹⁰¹ Zakaras, “A Liberal Pluralism.”

¹⁰² Joseph Raz: “Autonomy, toleration, and the harm principle”. In: Susan Mendus, *Justifying Toleration: Conceptual and Historical Perspectives* (Cambridge: Cambridge University Press, 2009), 163–165.

forms, some people simply cannot – and it makes no sense to demand it from them. Berlin’s insights about the value in diversity thus have no universal prescriptive bearing.

When it comes to the application of Berlin’s first-personal insights to society at large, the ensuing problems were demonstrated by my analysis of Crowder’s argument. It seems to me that Crowder wanted to channel the first-personal understanding of the plurality of genuine human goods into the social decision-making process, which would then supposedly result in giving equal value to all conceptions of good and thus pursuing the greatest possible diversity. Society would thus ‘understand’ different human goods as genuinely valuable and therefore would look to endorse as many of them as is feasible. However, on the societal level, the unconstrained belief in abstract pluralism of values is not possible, unless the given society is ‘unencumbered’ – which it never is. Society at large cannot empathetically understand all diverse values equally which means that they will never be “in the same league” – and the rest of the argument therefore does not follow. At the end of the day, a personal attitude of appreciation of pluralism is nothing more than that – a personal attitude. Some of us might find it attractive for various non-universalizable reasons but that is all that can be said about it. No social or political consequences follow.

However, one can argue that value pluralism is not merely a personal attitude – it is a hard *metaethical fact*. People (and societies) who appreciate diversity are therefore not only more empathetic – they are closer to truth. In different forms, Berlin, Gray and Crowder all seem to endorse this thesis at times. The nature of human values is such that they are real, many and incommensurable – and value pluralists are the ones who understand this fact. Nevertheless, the role of this fact in the argument for liberalism is still quite unclear.

First of all, it is doubtful whether this ‘fact’ even obtains. As Robert Talisse noted recently,¹⁰³ the arguments for pluralism as metaethical fact are suspiciously thin. Mostly they take the form of a claim that value pluralism best reflects our moral experience, especially when it comes to choosing, where often no highest principle can be found and irreducible losses occur. However Talisse offers an alternative interpretation of the phenomenology of choice, claiming that the regret we feel when having to choose would be quite irrational if pluralism really obtained (Why regret something absolutely inevitable?). Therefore, our experience of choice pushes us towards the monist alternatives where hard moral dilemmas are not a necessary part of the universe but an undesirable consequence of a situation that should have been avoided. Moreover, the fact that we often do not know the right answer to

¹⁰³ Robert B. Talisse, “Value Pluralism and Liberal Politics,” *Ethical Theory and Moral Practice* 14, no. 1 (2011): 87–100.

the dilemma may just mean that finding it is extremely hard, maybe even borderline epistemically impossible. But that is very far from the metaphysical impossibility postulated by value pluralists. In the end, Talisse as a Rawlsian is happy to provide the epistemological ‘fact of pluralism’ and ‘burdens of judgement’ explanation of diversity (with pluralism caused by a free exercise of human reason in varied conditions), without postulating any overarching metaethical story.

The only way for value pluralism to become a real *institutional guiding idea* is to establish the duty of political institutions to follow the truth that values are plural. However it is really unclear how such a duty can be established, not to mention that for anything politically conclusive it still requires the additional step from accepting the metaethical truth to valuing diversity in concrete political circumstances – a step that Crowder and others failed to make persuasively.

All in all, value pluralism as introduced by Gray and Crowder is prescriptively barren, at least when it comes to it providing a positive doctrine. This conclusion was often masked either by introducing additional assumptions that take the leading role in an argument, or by adopting half-implicit welfarist considerations, or by proposing unsuccessful arguments unable to connect value pluralism with some positive prescriptions. It seems, therefore, that the neo-Berlinians found a dead end in the quest of liberal justification.

However, this grim conclusion only follows if we adopt the conception of value pluralism favoured by the two thinkers above. For them, value pluralism is a fact (empirical or metaethical) about our moral universe that can be analysed more or less independently of any other factual or normative considerations. In contrast, as I have shown, Berlin operates with a radically different suggestion. For him, pluralism and human liberty are not an assumption and a desired conclusion but two sides of the same coin. Pluralism is a result of adopting a certain conception of person and *vice versa*. The two of them together can then establish a preference for liberty and a protection of an individual. Sadly, the Berlinian justification of liberalism is radically incomplete. We do not learn much about the conception of person underlying pluralism and its consequences are spelled out only in the roughest sketch. Nevertheless, there might be more to say about such approach and in the third chapter I will actually come relatively close the general picture that Berlin seems to endorse.

This chapter presented two simple and straightforward strategies to justify liberalism. One stipulated normative significance of human self-relation with the notion of self-

ownership, the other wanted to use value pluralism as the driving force towards liberalism. Both failed.

Self-ownership proved to be too crude an instrument to even be sufficiently coherent. Its aspirations to be the bedrock of liberal justification therefore turned out to be hollow. The most important reason for this failure is the fact that self-ownership works with a really contestable and paper-thin conception of person. The prescriptions it wants to achieve are, as a result, one sided and extremely controversial, produced by the power of one monumental and underexplained normative *fiat*.

Still, the basic model presented by the self-ownership justification of liberalism is very appealing. It seems clear that the nature of human beings should play a role in asking what type of political regime is the best; that human beings *as human beings* create certain claims that ought to be respected by political institutions. And self-ownership is the most straight-forward embodiment of this intuition. But being the most straight-forward is not a virtue in itself if the underlying assumptions are indefensible and the basic model untenable and very reductionist. The normative consequences and plausibility of a conception of person we choose to adopt must be much more carefully elaborated and much more vigorously defended. Moreover, this conception of person must never be the one and only foundational consideration. It needs to be supplemented by a view of society and human interaction within it.

The trouble with value pluralist justification of liberalism provided by neo-Berlinians, in sharp contrast to self-ownership, lied in mostly ignoring the role of a conception of person in liberal justification, and focusing instead on the nature of values and its consequences on human interaction. This proved to be an insufficient ground to provide any real prescriptive content – and thus failed to justify liberalism.

Neither the enquiry into the nature of values in human societies, nor the stipulation of the normative weight of human beings can carry the justification of liberalism by itself. Any plausible attempt to answer the original question needs to combine both of these factors. It needs to acknowledge both that human values are irreducibly plural and that freedom somehow retains its paramount importance as reflecting the deep nature of human beings. So far, Isaiah Berlin came closest to this position but failed to philosophically justify it.

The next chapter will be dedicated to John Rawls, a thinker that provided by far the most elaborate and interesting attempt to philosophically combine social pluralism and normative power of a conception of person into a unified theoretical conception. The analyses

present it this chapter should help in evaluating the theory of Rawls, identifying its weaker and stronger spots with regard to the main topic of this text.

CHAPTER II

1. Reading Rawls

The main question of this thesis remains the same in this chapter as it was before. I am still interested in ‘what can we philosophically say to justify liberalism as a mode of political existence of society’. After the disappointing results of the previous chapter, I ask it here in the framework of thought of the leading political philosopher of the 20th century – John Rawls.

The most important lesson from the previous chapter concerns the limits of straightforward justifications of liberalism. Neither self-ownership nor value pluralism works, as they operate in contexts that are too limited to carry the hefty justificatory burden. As a result, they end up either as simple stipulations of controversial normative principles, or as problematic assertions of metaethical truths with no clear path towards the rules of political practice. In this context, Rawls is a welcome change. He is one of the few recent thinkers with broad and systematic ambitions, keen not just to resolve particular issues but to introduce a full-fledged conception of justice for contemporary societies. The existence of such background greatly facilitates the quest for liberal justification, although it leaves substantial space for different interpretations. In this chapter I would like to navigate these interpretations, present what I take to be the most charitable reading of Rawls and his pursuit of ‘justice as fairness’, and, finally, see how his theory answers my question.

Concerning the different interpretations, Rawls’ work is most commonly divided into two distinct periods: ‘early Rawls’ of *A Theory of Justice*¹⁰⁴ tries to come up with an updated reconstruction of Kantian morality and contractualist arguments of Locke and Rousseau to fit contemporary political societies. The main task of early Rawls is to provide an alternative to utilitarianism, which was perceived as the strongest moral and political theory at the time. In contrast, ‘late Rawls’ is not interested in developing a moral theory to rival utilitarianism. In *Political Liberalism*¹⁰⁵, he wants to introduce a model for political deliberation about the principles of justice that can work with a plurality of ‘comprehensive’ theories – not just with

¹⁰⁴ John Rawls, *A Theory of Justice* (Harvard University Press, 1971).

¹⁰⁵ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993).

Kantianism and Lockean contractualism. The main areas of interest for Rawls in this period are public reason and overlapping consensus. All in all, Rawls can be seen to be engaged in two very distinct projects in his two main books: looking for justice as morality in *Theory*, while developing justice as a narrowly political exercise in *Political Liberalism*.

I hold that the view of Rawls presented in the previous paragraph is far too simplistic. Rawls does not have two distinct projects but one, as his main concern of ‘justice as fairness’ remains remarkably consistent throughout his substantial bibliography.¹⁰⁶ Even disregarding this main concern, the bulk of the content in this theory (concerning e.g. the nature of human cooperation, two principles of justice, or the idea of original position) is not radically modified from the 1958 essay *Justice as Fairness*¹⁰⁷ to the 2001 book *Justice as Fairness: A Restatement*.¹⁰⁸ Even the language used (‘restatement’) indicates more of an organic development than a radical shift.

Still, there is no denying that Rawls did change his views and shifted the focus of his writings somewhat. The most important of these shifts is indeed between *A Theory of Justice* and *Political Liberalism*. *Theory* introduces a comprehensive moral and political architecture we are supposed to agree on if we take the requisite point of view. At the end of the day, Rawls believed he could postulate an eventual agreement on broadly Kantian moral principles. However, Rawls later felt that his rendering of ‘congruence’ between the private and political principles of justice in his earlier work assumed too much. Under the conditions of freedom, the ‘requisite points of view’ always diverge as the power of human inventiveness and originality assures the truth of ‘a fact of pluralism’ in every free society. Realizing that the assumption of congruence is untenable, he wrote *Political Liberalism* as a remedy for this problem, abandoning the broader moral ambitions and focusing on justice purely in the political sphere.¹⁰⁹ However, since justice in political sphere was his primary target even in *Theory*, the shift is not as radical as it is often thought.

In my analysis of Rawls, I therefore do not carefully distinguish between his ‘earlier’ and ‘later’ philosophy. I draw from both and differentiate between them only when the issue at hand clearly calls for it. Given my question, I am interested mainly in the fundamental structural features of his liberal system and its justification, and in his conception of person. I

¹⁰⁶ For a thorough development of this position, see Catherine Audard, *John Rawls* (Acumen, 2007). She explicitly discusses the absence of radical shift in Rawls in *Ibid.*, 18–19.

¹⁰⁷ John Rawls, “Justice as Fairness,” *Philosophical Review* 67, no. 2 (1958): 164–94.

¹⁰⁸ John Rawls, *Justice as Fairness: A Restatement* (Harvard University Press, 2001).

¹⁰⁹ This is now a canonical interpretation of Rawls’ shift, originally presented in Samuel Freeman, “Congruence and the Good of Justice,” in *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge University Press, 2003), 277–315.

try to uncover his method, the starting points of his thinking and also the character and strength of his conclusions. That means this text has a slightly different focus from the majority of Rawlsian interpretative literature, which mostly analyzes either the concrete arguments he uses, or the main content of his theory – the two principles of justice. These topics will be present in this text only insofar as they are relevant for explaining the presuppositions behind the arguments and the main argumentative drive for his conclusions.

When it comes to grasping the nature of Rawlsian project, few places in his corpus offer a better insight than the very beginning of *A Theory of Justice*:

“Justice is the first virtue of social institutions, as truth is of system of thought. ... Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. ... Therefore in just society the liberties of equal citizenship are taken as settled; the rights secured by justice are not subject to political bargaining or to calculus of social interests.

These propositions seem to express our intuitive conviction of the primacy of justice. No doubt they are expressed too strongly. In any event, I wish to inquire whether these contentions are sound, and if so how they can be accounted for. To this end it is necessary to work out a theory of justice in the light of which these assertions can be interpreted and assessed.” (TJ 3-4/3-4 rev.)¹¹⁰

There are multiple ways of interpreting this passage, radically changing our perception of the arguments that come afterwards. In the first part of the quote, Rawls presents the basic features of what he intuitively considers to be the role of justice in politics. In the second part, he states his intention to somehow theoretically justify it. In a crucial phrase, he wishes “to inquire whether these contentions are sound, and if so how they can be accounted for”. An interpretation of this phrase may hold the key to understanding the whole ambition of Rawlsian project. However, its meaning crucially rests on what we understand under the word ‘sound’. If we read it in a weak sense, then Rawls only wants to see whether our intuitions (contentions about justice) are reasonable and coherent, so they can actually ground a decent theory. But, when reading ‘sound’ in a strong sense, Rawls wants to inquire whether our intuitions concerning justice are in fact defensible. That would presuppose him trying to justify them independently of their pre-theoretical intuitive appeal. So, in the second reading,

¹¹⁰ A remark on the adopted notational conventions: all references to Rawls will copy the established form in Rawlsian literature since the publication of the revised edition in 1999: TJ stands for *Theory of Justice*, PL for *Political Liberalism*, CP for *Collected Papers* (John Rawls, *Collected Papers* (Harvard University Press, 1999).). When citing *A Theory of Justice*, the first number represents a page from the original edition (Rawls, *A Theory of Justice*, 1971.). The second number represents the revised edition (John Rawls, *A Theory of Justice*, Revised edition (Cambridge, Mass: Harvard University Press, 1999).).

mere coherence is supplemented with a drive for critical verification or justification. That is a significant shift of meaning, though it is far from clear at this point what this sort of verification should consist in.

In the rest of the chapter, my aim is to clarify this fundamental uncertainty regarding the basic ambition of the Rawlsian project and the nature of justification in it. At the very heart of it is the role of intuitions, which present the starting point of the argumentation. Understanding their role is thus a key factor in overall interpretation of Rawlsian liberal project and it should help me see to what extent a Rawlsian theory can answer my question.

A quick caveat: the argumentation in this chapter moves in a considerably slower pace than the previous one. Here, I am not interested in the cogency of singular arguments and isolated positions. I evaluate a theory that is complex and multifaceted in a way that the previously discussed strategies of liberal justification simply are not. I try to carefully unearth the sources of Rawls' liberalism and the way he works with them. In the end, I aim to present a complex interpretive view of the parts of Rawls' thinking that are relevant to my project.

In the next section, I present the two methods of justification that Rawls has developed for the purposes of his theory: reflective equilibrium and original position. I analyze their role within the argumentative structure, together with their mutual relation. In the third section, I look at the content of Rawlsian 'intuitions' and their (very important) role in his argumentation. I argue that when the methods of justification connect with the intuitions that provide their basic content, they suffer from the same interpretive dichotomy as the above-quoted passage from TJ. Therefore, in the fourth section I develop two distinct ways to understand the Rawlsian project as a quest for liberal justification – the deliberative and descriptive interpretation – and I analyze their respective strengths and deficiencies.

In the final analysis, my conclusion will again be negative: Rawls *cannot* plausibly respond to the justificatory challenge put forward in this thesis. His answer to the question 'why liberalism?' is either incomplete or implausible, on both interpretations. However, he still gets much further than the conceptions I analyzed in the previous chapter. His analysis of the role of pluralism in contemporary societies is persuasive and his conception of person, while not transparently justified, is very appealing. Those are taken to the next chapter, which presents the last attempt to answer the justificatory challenge.

2. Two justificatory strategies

The main content of Rawlsian liberalism (to indicate *what* is actually being justified) is provided by the now famous two principles of justice. Their formulations vary in the different texts, but never so much as to change the basic idea. In *Political Liberalism*, they read as follows:

“a) Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value.

b) Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least advantaged members of society.” (PL 5-6)

These two principles satisfy the liberal ‘Millite core’ – that much is not controversial. In spite of some vague formulations (‘fully adequate scheme of rights’), they do encapsulate the liberal drive for individual rights, personal freedom, social equality, and distributive justice, while condemning entrenched privilege and merely formal liberties. They suggest a positive vision of how our society should be organized, they present what is for Rawls “the most appropriate conception of justice” (PL 3) – what he calls justice as fairness.

Apart from specifying the two principles, Rawls dedicates much space and effort into explaining what makes this conception of justice admissible, or indeed “the most appropriate”. Interestingly, these are the passages where he is philosophically the most innovative. He does not merely “generalize and carry to a higher order of abstraction the traditional theory of the social contract as represented by Locke, Rousseau, and Kant” (TJ viii/xviii rev.) – he revolutionizes it. The conceptions of reflective equilibrium and the original position radically alter the traditional contractualist idea of a group of (future) citizens consenting to common rules in a state of nature. From the very beginning, Rawls aims his arguments *at us* – he tries to make our commitments to (a specific conception of) justice as clear as possible. Hence, the tools he develops, reflective equilibrium and the original position, should answer the crucial question: why we should adopt justice as fairness as the conception of justice for our political societies.

2.1. *Reflective equilibrium*

When discussing the method and character of justification in Rawls, it is only natural to start with reflective equilibrium. For him it represents the broadest scheme of justification in moral and political philosophy, and maybe even further. It is the broadest because there is nothing beyond reflective equilibrium in practical philosophy. All the presuppositions, intuitions, arguments and theories are already in it. The basic claim is simple: all possible principles of practical philosophy must, in order to gain acceptance, be supported by some sort of rational argument we can adopt upon deliberation and then incorporate it into our wider system of beliefs. To put it in other words, Rawls stipulates that if something is to be justified, it needs an argument or a reason supporting it. Moreover, the position we adopt as a result of this argumentation cannot stand isolated from other beliefs we hold. We must therefore at least check whether our beliefs (on multiple levels of generality) aren't incoherent or mutually exclusive.¹¹¹

Elaborating on the previous point, reflective equilibrium is a condition in which our intuitions and judgments about particular cases are in line with broader principles and priority rules we came to adopt. That means we can explain our intuitive beliefs thanks to our principles and our principles have intuitive appeal because they are in line with the beliefs that we find plausible. But a good consonance between particular beliefs and principles is not the only condition of successful reflective equilibrium. The sum of our principles needs to add up to a coherent whole – and also our particular beliefs must not be in conflict with each other. If all these three conditions are fulfilled, we can say we achieved the equilibrium.

But reflective equilibrium can be viewed as a process, not only as a result, a static state of affairs. In everyday language we use the term 'equilibrium' to describe a position in which different forces are so balanced that they cancel each other out. But forces coming into Rawlsian reflective equilibrium are far from stable. Our society constantly changes, its self-understanding subtly shifts, technology progresses, social relations evolve, rankings of shared values get shuffled under practical or theoretical pressures. Therefore, principles or at least their application must change in time. Reflective equilibrium is thus more about constant balancing of principles and their applications than about finding one ideal state and sticking to it.

¹¹¹ For alternative elaboration of this position, see Robert Audi, "Intuitionism, Pluralism and the Foundations of Ethics," in *Moral Knowledge? New Readings in Moral Epistemology*. (Oxford: Oxford University Press, 1996), 101–36. Audi calls this position 'ethical reflectionism'.

An important test for achieving this balance is provided by Rawls' theory of considered judgment (TJ 47-48/42 rev.). Considered judgments are intuitions that we not only view as correct – we also claim it with some confidence. They are judgments made after careful reflection without disturbing factors such as strong personal interest, uncertainty, emotional distress and so on. A few political examples of such judgments may include 'slavery is unjust', or 'collective guilt is unacceptable'. Even though these judgments were not universally endorsed throughout human history, we still feel very strongly about them and they are deeply embedded in the public culture of societies we live in. And the more strongly we feel about them, the less we are prepared to abandon them. Thus, considered judgments can test the plausibility of our principles and their application as "provisional fixed points". Reflective equilibrium is not achieved (1) if our principles do not cover the cases we feel strongly about, or (2) if our considered judgments oppose the conclusions of our principles.

The process of reaching reflective equilibrium thus "works from both ends" (TJ 20/18 rev.). On the one hand, there are principles of varying degree of generality; on the other, there are specific judgments/intuitions about specific cases. A good theory presents principles that cover and apply to all relevant specific cases in a suitable way. However, we can often find ourselves in a situation when there is a discrepancy between our considered judgments about specific cases and the adopted principles. Then we need to gradually fine-tune the principles so they can fit our considered judgments – and also modify our judgments so that it's possible to express them with one coherent set of principles. The process works from both ends, ideally resulting in harmony between the applications of a given theory, our intuitions regarding its subject matter, and all the principles taken as one coherent system.

Reflective equilibrium thus aims to model the deliberation of limited and mortal rational beings with regards to political principles. Because of the inescapable human limitations and the open-endedness of our existence, no set of principles can be universally agreed upon once and for all. Even if we shared an identical set of political intuitions, various principles could cover various aspects of these intuitions to varying degrees, which could result in a plurality of conceptions and principles. The search for 'the most adequate' set of principles is thus a never-ending process of fine-tuning and re-arranging of what we know and what we feel confident about.

The primary goal of reflective equilibrium is to eliminate confusions and incoherencies in principles and in intuitions themselves, and, in an ideal case, to show one set of principles which at the given moment best accommodates all the theoretical and practical conditions of coherence and plausibility. The resulting theory should have a well worked-out

set of principles with clear and coherent practical implications, in line with (modified) considered judgments. In summary, the method Rawls proposes as an ultimate form of justification is in fact coherentist. He wants to work out a theory where principles and their applications would be in harmony, creating one appealing theoretical whole.

If reflective equilibrium is taken as a justificatory tool, considering the role of intuitions in it presents an acute complication. Rawls would like to think that reflective equilibrium is a theoretical ideal with no pre-determined content – that all principles and conceptions have an equal chance of becoming a part of it if they prove persuasive enough. And on surface, this is certainly true. There is no trick with definitions that would exclude utilitarianism, naturalism or any form of religious morality from entering or even dominating reflective equilibrium if we take it as purely methodological tool and separate it from the rest of Rawls' theory. However, Rawls' use of reflective equilibrium is far from neutral and content-less. He fills it with intuitions and considered judgments of his choice, which helps him to develop his two principles of justice.¹¹² That moves the justificatory burden from the reflective equilibrium itself to its content – the 'intuitions' – that now do the philosophical work. Yet, to specify the precise role of intuitions in Rawls is one of the most daunting interpretive tasks within his corpus. Therefore, given the importance of this issue, different interpretations of the role and justification of the intuitions he uses may completely shift the understanding of his theory – as I will show later on.

Not getting too far ahead, in the reflective equilibrium the role of intuitions can be interpreted in two different ways:

In the first interpretation, intuitions are taken as the only content that enters the political deliberation. They are divided into two classes – those we trust more and those we trust less. After some deliberation, the more trustworthy intuitions become considered judgments and thus gain greater importance. Both classes are then conceptualized into one coherent set of principles, with considered judgments having a more important role. Subsequently, the aim is (through a process that works from both ends) to balance intuitions and principles, so that they work well together. Resulting from this process is a coherent set of principles covering the greatest possible part of our pre-theoretical intuitions.

¹¹² After all, even the most basic and seemingly innocuous considered judgment – that slavery is always unjust – does bring into the reflective equilibrium a type of content that Rawls wants it to have. If he adopted a different set of considered judgments (including for example 'traditions are to be respected'), the result would be completely different.

Richard Hare presents several very powerful objections against a theory of this sort.¹¹³ Firstly, reflective equilibrium interpreted as a vehicle for intuitions limits morality into a moral sense theory. Rawls assumes that our pre-theoretical moral feelings are mostly correct and we only need to balance them out and make them coherent, so that they can constitute a full and articulate moral theory. This conclusion is illustrated by his comparison between moral sense and the grammar sense of a native speaker (TJ 46/41 rev.). The fact that the native speaker speaks correctly is assumed. When trying to come up with the rules of grammar, the task is only to carry the concrete utterances to a higher level where they can be expressed as principles. But that, with regards to morality, is a truly controversial position and Rawls does not defend it in any way in *A Theory of Justice*.

The situation changes in *Political Liberalism*. There, Rawls drops the grammatical analogies and provides a political explanation of our intuitions – he is interested in our moral intuitions only insofar as they are implicit in the public culture of liberal democracies. The reflective equilibrium to be achieved is therefore limited to the political, no longer trying to encapsulate the totality of practical philosophy (PL 8). Still, Hare’s objection can be reformulated even in this environment. Hare could claim that Rawls presents us a theory of ‘political sense’, too readily taking the prevalent normative perceptions of desirable political relations at the face value, disregarding the possibility that they might be wrong or simply unjustified.

A second problem presented by Hare is (what he takes to be) a blind faith in the fact that intuitions will via reflective equilibrium yield only one set of principles, which would provide the best theoretical cover for them. Hare does not trust intuitions and claims that adopting this strategy would in the end most probably result in a great multiplicity of differing equilibria balancing our intuitions in different ways. There would be many options and no way how to decide between them.

In a sense, this is exactly the picture presented by Berlin and other value pluralists I discussed in the first chapter. If values are many, if we are drawn to incompatible conceptions of the good, we either find no reflective equilibrium at all (because any coherent normative position implies too many tragic choices), or, in a less dramatic understanding, we find too many of them, each catering to a different preference ordering of values. But if these specific orderings are incommensurable, we are left with a normative impasse, not able to find ‘the most appropriate’ conception of justice for political society.

¹¹³ Hare’s analysis is very critical, borderline insulting. See R. M. Hare, “Review: Rawls’ Theory of Justice–I,” *Philosophical Quarterly* 23, no. 91 (1973): 144–55.

Here, as above, *Political Liberalism* offers some solution. Rawls claims that the tragic choices between reasonable conceptions of the good and the right definitely can and do happen. However, they are confined to the personal sphere of comprehensive doctrines, while the space for reasonable disagreement is much more limited when discussing the basic political structure of society (PL 157). As a result, reflective equilibrium might just produce the one conception of justice as fairness if its application strictly limited to the political.

To sum up, in the first interpretation our (political) intuitions are the only content introduced into reflective equilibrium. Then, working from both ends, these intuitions are refined and made coherent. The result is (ideally) a full conception of justice that reveals and strengthens the normative commitments of our antecedent beliefs about justice.

In the second interpretation, intuitions do not present the given content that only needs to be sorted out, systematized, and made coherent. The method of reflective equilibrium here aims at developing the correct (or ‘sound’) moral and political theory. Quoting Norman Daniels from *Stanford Encyclopedia of Philosophy*:

„To be of interest to moral philosophy, a reflective equilibrium should seek what results from challenging existing beliefs by arguments and implications that derive from the panoply of developed positions in moral and political philosophy. Such a reflective equilibrium would be the response to considerable critical pressures on the original beliefs. This effort would have the character of *searching deliberation about what is right*“.¹¹⁴

Reflective equilibrium under this interpretation achieves, as above, a coherent set of principles. But it no longer aims at only that – the resistance to criticism becomes also very important. Resulting principles must not only express our intuitions well, they must also be defended and argued for against alternative principles. Only then it is possible to see the result of reflective equilibrium as not only coherent but indeed ‘the most appropriate’. Under this interpretation, the fact that our intuitions may be misleading or even outright inaccurate no longer poses a problem. They need to be justified to take part in reflective equilibrium. In this interpretation, the equilibrium in fact presents the general requirements for any potentially successful political theory: it needs to be coherent; it needs to cover the important cases, and it also needs to be able to withstand critique and prove to be better than alternative theories.

¹¹⁴ Norman Daniels, “Reflective Equilibrium,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2013, 2013, sec. 3.1, <http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>. Emphasis added.

If the above-enumerated requirements are fulfilled, the resulting theory would be quite robust. However, the deliberative process of achieving this sort of reflective equilibrium would certainly be quite complex and definitely not reducible to the simple ‘working from both ends’ strategy. The given theory of justice would need to transcend the basic requirements of coherence and intuitive plausibility and offer positive and persuasive arguments in favour of the intuitions it comes to adopt. How it should do that is, strictly from the perspective of reflective equilibrium, impossible to say.

To conclude this section, Rawls’ presentation of reflective equilibrium leaves the role of intuitions crucially under-explained. That is why we can read it, to use Thomas Scanlon’s terminology, *descriptively*, as a generalization of our intuitions into principles while making them coherent; or *deliberatively*, as a critical process of searching for the most appropriate theory.¹¹⁵ Both these interpretations have their own problems and neither presents reflective equilibrium as the full theory of justification that can stand on its own and generate an answer to my question. Rawls needs something more to support his justice as fairness. That, in my reading, is the role of the original position.

2.2. *The original position*

Reflective equilibrium, as I mentioned in the previous section, works at two levels in Rawls. On the one hand, it presents a way of constructing a theory of justice; on the other, it is the most general form of justification of any possible theory of justice. It is a method of constructing principles through their gradual adjustment and confrontation with considered judgments, but it also shows conditions under which the adopted principles could be trusted. Reflective equilibrium shows the goal as well as the means to attain it at the same time.

The situation is quite similar with the original position and the veil of ignorance. In a hypothetical scenario of original position, people deliberate about the principles of justice applied to the basic structure of society in a situation where they do not have any information concerning their social status, conception of good, or personal preferences. The original position therefore also presents a route towards developing a theory of justice. It modifies the knowledge and the motivations of the parties involved so that it is possible for them to agree on one set of principles. It is a procedural tool which can be used to replace our messy and pluralistic reality with a situation in which some sort of consensus is achievable. The

¹¹⁵ Thomas Scanlon, “Rawls on Justification,” in *Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge University Press, 2003).

problems of value pluralism and the ensuing impossibility of full consensus are, in a sense, sidestepped, because the deliberating parties do not know their conceptions of good. The parties behind the veil of ignorance are thus expected to agree on a conception of justice that is clear of any distortions caused by inequalities of power, or morally arbitrary preference for one's own lifestyle and values. A justice as fairness is thus the product of a process in which at the first stage the conditions of deliberation are modified and only after that the deliberation itself takes place, ideally providing a single and specific-enough result.

But the original position does not aim to be only a vehicle for developing the principles of justice. For Rawls, the sheer fact that some principles would be adopted in it constitutes a reason why these principles are valid or correct. The conditions of deliberation behind the veil of ignorance represent the basic conditions of fairness, as they prevent partiality arising from the knowledge of our interests and status. Therefore, once again, the role of the original position is twofold. It presents a path towards as well as a justification of a theory of justice. Following the path guarantees the achievement of the goal – the construction of the most appropriate conception of justice.

The original position and the reflective equilibrium are therefore the two schemes of justification in Rawls – and there arises a very natural question regarding their relation. The text of *Theory* does not shed too much light on it and neither do the follow-up articles. Their relation, from a strictly textual point of view, is unclear. But in spite of this, I will try, using some interpretative literature, to propose what I consider to be the best interpretation of this relation.

Thomas Scanlon claims that the fundamental difference between reflective equilibrium and the original position lies in the fact that whereas the original position constructs principles deductively, reflective equilibrium is always inductive.¹¹⁶ Reflective equilibrium starts with a broad range of judgments, arguments and conceptions and tries to put some order into them. The original position is deductive, because „principles of justice are justified if they could be derived (from the original position – my insertion) in the right way, institutions are just if they conform to these principles, and particular distributions are just if they are the products of just institutions.“¹¹⁷

Under these circumstances, it seems reasonable to claim that these two methods of justification present two independent ways that ultimately lead to the same result. But this contention is untenable, because in *Theory* they are introduced together and Rawls explicitly

¹¹⁶ *Ibid.*, 139.

¹¹⁷ *Ibid.*

uses the idea of reflective equilibrium when he specifies the limitations behind the veil of ignorance (TJ 20-21/18 rev.). Thus, there is no escaping from the task of clarifying their relation.

The question of the scope of the reflective equilibrium is crucial to finding the solution to this problem.¹¹⁸ Reflective equilibrium is the most general method of justification in practical philosophy and maybe even beyond. It shows us how to work with intuitions and sets requirements for the resulting theory. Viewed as such, the reflective equilibrium is the most general method of deliberation, whereas the original position is only one tool used in a limited yet important area of social justice. However, if we interpret reflective equilibrium in a more limited way, it is not a general method of justification – it only presents the interplay of principles and their applications within one specific field. With regard to the original position, Rawls shows how we can use reflective equilibrium to modify it so that “by going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted” (TJ 20/18.rev.). Reflective equilibrium here does not stand above the original position – it precedes it and helps specify it.

What is shown here is an observation critically important for the interpretation of the whole *Theory of Justice*. The original position is not in any way just independently given. It and the veil of ignorance need to be suitably specified in order to serve their purpose. But Rawls does not find any transparent deductive way of doing it. He adjusts the conditions in the “initial situation” to develop the original position, which produces the results that are in line with our modified considered judgments about justice. *The original position is thus one interpretation of the initial situation that best fits Rawls’ intuitions about justice* (TJ 118/102 rev.).

This observation puts in doubt Scanlon’s categorization of the original position as the deductive tool for the justification of principles of justice. Put more precisely, the original position has a deductive component. It helps construct principles from already specified conditions behind the veil of ignorance. However, there is a very strong inductive component in trying to set up these conditions so that the resulting principles conform to our pre-

¹¹⁸ Just three years after the first publication of *Theory* Rawls published *The Independence of Moral Theory* (CP 286-302). There he distinguishes between wide and narrow equilibrium. For the canonical interpretation of this dichotomy see Norman Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice* (Cambridge University Press, 1996).

conceived ideas about justice (in order to achieve reflective equilibrium). Our pre-theoretical intuitions specify “reasonable conditions” (TJ 20/18 rev.) of fairness behind the veil of ignorance which enable Rawls to deduce the principles of justice that are in line with our considered judgments.

Because of its intimate relation with reflective equilibrium, the original position suffers from the same ambiguity regarding the role of intuitions. The original position should specify fair conditions of deliberation about the principles of justice to guarantee a fair result. Therefore, the parties behind the veil of ignorance are set up as free and equal, reasonable and rational human beings with a sense of justice and a (yet unspecified) conception of good life.¹¹⁹ Rawls hides a lot from the parties, but at the same time he assumes a lot about them as well.

All these assumptions and omissions carry a crucial weight in the deduction of the principles of justice. They are the ones that actually enable a consensus to emerge behind the veil of ignorance. Moreover, their importance is underlined by the fact that we can, by altering the conditions of the initial situation, come to radically different conclusions. Rawls explicitly claims that “for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution” (TJ 105 rev.¹²⁰). The thought experiment of the initial situation thus does not bring in itself any answers to the question of social justice. On the contrary, it can yield basically any answer. The key task for Rawls is therefore to find and defend one best specification of the initial situation (the original position) that results in the justice as fairness as the most appropriate conception of justice. Rawls explains the task in the very paragraph I just quoted from:

“(T)he question of justification is settled, as far as it can be, by showing that there is one interpretation of the initial situation which best expresses the conditions that are widely thought reasonable to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgments in reflective equilibrium. ... The procedure of contract theories provides, then, a general analytic method for the comparative study of conceptions of justice. One tries to set out the different conditions embodied in the contractual situation in which their principles would be chosen. In this way one formulates the various underlying assumptions on which these conceptions seem to depend.”

¹¹⁹ TJ, Chapter III.

¹²⁰ This quote is not present in the original edition. A paragraph on pages 125-126 that corresponds to the place cited above states a similar, though less elaborated, thought.

The thought experiment of the initial situation is therefore used as a tool for clarifying the underlying assumptions in various theories of justice. Using it uncovers what sort of fundamental considerations lay at their core. Once this is done, it is possible to decide which theory “best expresses the conditions that are widely thought reasonable” and which one expresses our considered judgments. In short, introducing any theory of justice as a result of a consensus in an initial situation strips it down to its essentials, unearths the considerations that are really in play when choosing it, and thus helps us to develop the most suitable one – which should be the one that results from considerations, and only those considerations, that are relevant for developing a theory of justice for political societies, while excluding all the irrelevant factors. This conception of justice should then inform our political societies.

Thus, when coming up with the principles of justice, the crucial role is played not only by the procedural deduction of principles from the original position, but also by our intuitions which modify the veil of ignorance and therefore have a decisive impact on the final form of the chosen principles of justice. This impact is so profound that without adopting a concrete set of intuitions driving the specification of the conditions behind the veil of ignorance, the whole construct of original position is completely inert – it is not possible to deduce any principles from it or, for that matter, get any help with the constitution of a just or legitimate state. Only with the adoption of specific starting values is it possible to suitably set up the original position and thus be able to specify the principles of justice. Therefore, if the parties behind the veil are presented as free and equal, it is only obvious that the resulting principles would mirror the assumed freedom and equality. But if they were independent and self-owning (as in Nozick, for example), then the resulting principles would be much different.

The reason the original position provides the most appropriate theory of justice lies in the fact that its underlying assumptions are very much in line with what we view as relevant for political justice. Because of this, the original position is able to produce principles that can be in reflective equilibrium with our considered judgements about justice. The whole process of justification from the original position thus hangs on Rawls being able to show that the underlying assumptions that he adopts are themselves justifiable. And these underlying assumptions are the topic of the next section.

3. The conception of person and its role

The justificatory tools analysed above are radically incomplete. The original position and reflective equilibrium both derive their plausibility from their ability to produce principles that are in accordance with our intuitions. Strictly speaking, Rawls is moving in a circle in which the principles of justice both vindicate our intuitions and are in turn vindicated by them.¹²¹ In Richard Hare's unsympathetic interpretation, this amounts to Rawls being a moral sense theorist with no further story to tell.¹²² I do not think that such reading does justice to Rawls. Still, it is clear now that any philosophical justification of Rawlsian liberalism must run through a careful elaboration of the role, content, and grounding of the intuitions he uses as starting points of his deliberations.

When it comes to the content of Rawls' intuitions that form the original position and veil of ignorance, they involve almost exclusively his conception of person. Throughout his texts, Rawls repeatedly works with seven principal characteristics of human persons that are again and again analyzed, elaborated, compared, and derived from one another. People are presented as free and equal,¹²³ reasonable and rational,¹²⁴ possessing moral personality consisting in an ability to adopt a conception of good and a sense of justice,¹²⁵ and, finally, humans are presented as self-authenticating sources of valid moral claims.¹²⁶ These seven characteristics of persons are at the root of all the restrictions and limitations in the veil of ignorance, enabling it to produce principles that can possibly achieve reflective equilibrium.

Freedom and equality of human beings constitute the fundamental starting point of the construction of the principles of justice in Rawls. Right at the start of *Theory* in the section entitled "The main idea of the theory of justice" (TJ 13/10 rev.), he claims that just is what free and equal persons can accept under fair conditions. Therefore, justice is defined procedurally, with freedom and equality being right at the beginning of the procedure. In *Justice as Fairness: Political not Metaphysical* Rawls directly states that one of the main tasks of political philosophy is to find an appropriate expression of freedom and equality. That is why "the fundamental intuitive idea (of his project) is that of society as a system of fair

¹²¹ For this interpretation, see Paul Ricoeur's excellent article "Le cercle de la démonstration" in : Catherine Audard, *Individu et justice sociale (autour de John Rawls)* (Paris: Seuil, 1988).

¹²² Hare, "Review."

¹²³ E.g. TJ 13/11 rev.; TJ §40; PL 3.

¹²⁴ E.g. CP 316; PL 48.

¹²⁵ E.g. TJ §77; PL 81.

¹²⁶ E.g. PL 32. This last one stands a bit apart from the rest, because it does not appear in TJ, but only later. Still, it merits the inclusion into the list, as it quite well expresses what Rawls wanted to say already in TJ.

social cooperation between free and equal persons.” In the same vein, in *Political Liberalism* the main question to be answered is “what is the most appropriate conception of justice for specifying the fair terms of social cooperation between citizens regarded as free and equal” (PL 3).¹²⁷

As freedom and equality stand at the start of a procedural conception of justice, one of their key specifications is procedural as well. Persons perceive each other as equals, that means “as having an equal right to determine, and to assess upon due reflection, the first principles of justice by which the basic structure of their society is to be governed“. Similarly, people are free “in that they think they are entitled to make claims on the design of their common institutions in the name of their own fundamental aims and highest-order interests” (CP 309, cf. TJ 475 rev.). Thus, procedurally defined freedom and equality require at least some minimal form of contractualism. Citizens must be the authors of a political setup in society. If they are not, their freedom and equality are diminished.

However, a purely procedural conception of justice is impossible for Rawls (PL 192).¹²⁸ The procedural notions of freedom and equality already presuppose some non-procedural basis. For freedom to have any meaning at all, a person must have a capacity to adopt goals different from the goals of society at large. Moreover, she must be able to demand for society to respect these goals, at least to a certain extent. Freedom means that citizens need not draw their goals from any higher entity than their inner preferences, interests, and values. What is more, society has a duty to protect these goals to an extent compatible with equality.

On numerous occasions when trying to elaborate the meaning of freedom and equality, Rawls turns to his conception of moral personality consisting in a sense of justice and a capacity to adopt a conception of good life.¹²⁹ The move is put most concisely in *Political Liberalism*: “...we also think of citizens as free and equal persons. The basic idea is that in virtue of their two moral powers (a capacity for a sense of justice and for a conception of good) and the powers of reason (of judgment, thought and inference connected with these

¹²⁷ This means that, for the most part, Rawls does not ask the justificatory question with regards to liberalism as broadly as I do. A quintessentially liberal conception of person stands in the background of the very task he gives himself. Still, he needs to be very clear on what conception of person he adopts and why. If he does not do that, the adopted conception may appear arbitrary – which would signal a (partial) failure of his enterprise. On the other hand, if he succeeds in vindicating his conception of person, he answers the broader justificatory question as well.

¹²⁸ On some readings, TJ indeed explicitly strived for a purely procedural justice. However, this interpretation was later emphatically denied by Rawls, admitting his mistakes and confusing formulations.

¹²⁹ Exactly the same move is found in TJ 505/442 rev.; CP 312; CP 392; PL 19.

powers), persons are free. Their having these powers to requisite minimum degree to be fully cooperating members of society makes persons equal” (PL 19).

The capacity to adopt a conception of good presupposes that a person can embrace a rational plan of life that transcends the needs of the moment.¹³⁰ A human being does not have only instincts and desires but also interests, second-order desires, and views about wider world and its role in it. The capacity to adopt a conception of good thus presupposes a possibility for a purposeful promotion of one’s interests. Moreover, a conception of good life so adopted does not need to be identical with a conception of good life of others, which creates space for freedom and plurality.

The sense of justice is understood as a force that can limit the pursuit of individual interests when they go contrary to what is considered to be right, just, or fair. Rawls sometimes puts it even more strongly – he claims that people have a “regulative desire to act upon certain principles of right” (TJ 561/491 rev.). It would seem that such an anthropological assumption would require a complex and rather problematic psychology, but that is not the case. Rawls only claims that a normal human being considers certain acts or principles to be right or wrong when interacting with others. Her actions are thus not merely instinct- or desire-driven because she is able to apply moral categories on them, with fairness being the prime example (PL 86). Moral psychology that Rawls uses thus remains “philosophical, not psychological” (PL 86-88).

This theory of moral personality shows the fundamental presuppositions of citizenship and organized society. A capacity to adopt a conception of good life guarantees that a person can identify her and in some cases also social interests and that she can at least potentially act to fulfil them. In stark contrast, an entity without capacity to view, formulate and pursue its own interests cannot have any reasons for political participation and therefore it cannot object to even the most invasive paternalism. Society without people having and formulating at least their basic interests becomes more like a hive, with no obstacles to complete enslavement.

The sense of justice guarantees the basic possibility of social rules. Without a capacity of citizens to understand and internalize rules and practices, right and wrong, or good and bad, any society always dwells in a Hobbesian state of nature. Only after mutual acceptance of social rules and accepted forms of behaviour can a number of disjointed individuals form a society in the full sense of the word. And the sense of justice is a crucial and irreplaceable component of this move.

¹³⁰ TJ 561/491 rev.

The anthropological assumption of moral personality is therefore necessary for any form of complex social cooperation. For example, every contract (not only the social one) is an expression of some common interest (of selling something, working somewhere etc.) specified by certain conditions that need to be respected by both parties. Without a (shared) interest there is no reason to create a contract and without an elementary respect to its rules there is no point to it. The sense of justice and the capacity to adopt a conception of good life are the key components of the Rawlsian idea of society as a fair system of cooperation. They may not be sufficient conditions for its creation – but they are certainly necessary for it.

The powers of moral personality are for Rawls present equally in every citizen – and that makes them suitable as the basis of equality. This statement may sound controversial – one may argue that it is not possible to base equality on natural attributes since this is exactly an area in which we are very different from one another. Even a sense of justice can be more or less developed across different individuals (TJ 504/444 rev.).¹³¹ However, Rawls emphasizes that he is interested in “a capacity and not the realization of it” (TJ 509/445 rev.). In the end, Rawls takes the capacity for a development of moral personality to be a threshold concept. If someone passes it, it does not matter how talented he is or how he realizes it – mere possession of this capacity is all that matters. The illustration used is a geometrical property of being interior of the unit circle (TJ 508/444 rev.). The points on the plain either have this property fully or not at all – and if they have it, it does not matter whether they are closer or further away from the centre. Similarly, people either possess a sense of justice or they do not. If their sense of justice is particularly well developed, they can work for judiciary – but their rights and duties as citizens are not in any way affected by it. Therefore, a possession of a capacity for moral personality is a sufficient condition for being entitled to equal justice (TJ 505/442 rev.).

However, this picture presented by Rawls is a bit too simple and ignores some crucial problems, as was recently shown by Ian Carter.¹³² Rawls introduces a binary threshold concept of “possessing a moral personality” that should count as a basis of equality, yet even he admits that this binary property supervenes on more basic properties that exist on a broad scale. A sense of justice can be found in many shades and degrees. If that is the case, Rawls needs an argument showing why the threshold concept of possessing a moral personality is somehow primary, even if it only supervenes on the more basic concepts that are unequal. The challenge, as formulated by Carter, is to find a reason why we should discard the differences

¹³¹ Cf. CP 333.

¹³² Ian Carter, “Respect and the Basis of Equality,” *Ethics* 121, no. 3 (2011): 538–71.

between our moral powers once we reach a certain threshold – because only that can enable us to view human beings that possess moral personality as equal.

Rawls does not seem to have an answer to the challenge, and that is why Carter introduces his own argument. Its basic move, without going into too much detail, is a claim that democratic states must abstain from evaluating persons' varying capacities above the threshold level. The capacities of citizens must remain 'opaque' for reasons of respect. State ought only to pay attention to citizen's outward features as agents, not the inner workings of their personality, since these are private. This evaluative abstinence then results in the necessity to treat the powers of moral personality as binary properties even if they do admit of degrees, which makes them excellent in their role as the basis of equality (all people above the given threshold are viewed as free and equal).¹³³

As I mentioned in passing above, Rawls focuses on capacities of moral personality, ignoring the actual fulfilment. He is mostly interested in a conception of person that can be used as a foundation of a theory setting up basic rules of cooperation in society. Since every actual fulfilment of human potential is contingent on the antecedent social structure and the particular situation of a given individual, every theory working out the desirability of certain social structure must work only with capacities (PL 269). Any dependence on their specific fulfilment would risk being arbitrary from the moral point of view. Moral personality is extremely important for Rawls exactly because of this feature. It assures the broad scope of his theory and serves (as I show in a moment) as the most basic stepping stone for almost every argumentative move he makes.

Rationality and reasonableness are the last pair of key concepts for Rawls' conception of person.¹³⁴ They are to a huge extent intertwined with freedom and equality as well as with

¹³³ But even if the equality in the central cases above the threshold is solved, we still need to ask about the fate of the people that fall below the threshold. For Rawls, these people are not less human or less deserving of protection. It only means that people falling below the threshold are not taken as the constitutive elements of political community and therefore are not entitled to equal rights and freedoms. Their status would need to be sorted out in their specific circumstances, taking into consideration their particular condition.

This position may sound awful, inconsiderate, and discriminatory, but that is not the case. People below the threshold of moral personality do not constitute a race or a minority. They are scattered individuals usually suffering from some kind of serious disability or illness. In order to be incapable of understanding the right and wrong, one needs to be a psychopath. And for an inability to have any kind of concept of good life, one must be severely mentally disabled. These people are incapable of everyday independent functioning in our society and they cannot become parties to any sort of social contract. Therefore, there is no injustice in their not having equal rights and liberties. It is all right if a psychopath loses freedom of movement and a severely mentally disabled does not have political rights. They are still human, still very much worthy of our attention – but they have no claim for equal rights and liberties.

¹³⁴ The most concise definition of Rawlsian reasonableness and rationality can be found in CP 503. See also Thomas Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), 191–194.

moral personality. In a sense, they represent the two basic moves of Rawlsian philosophy and Rawls wants to keep them strictly separated. Rationality is presented procedurally, as a deliberation of parties in the original position, while reasonableness is presented as a sum of conditions and limitations of the original position itself (PL 48-54).

Rationality is understood as a maximization of personal interests without any moral or other restrictions. A person acts rationally if she tries to maximize her advantage when bargaining with a powerless or disadvantaged party that cannot but agree with the terms given. On the other hand, reasonableness assumes that an action should be determined not only by the maximization of interests-satisfaction, but also by fairness or other moral restrictions. That is why it is possible to say that “their proposal is rational, given their circumstances, but it is unreasonable all the same” (CP 503). Such proposal may well reflect the bargaining power at the given moment, but it would not be considered if the parties had more equal standing. Reasonableness is therefore presented as a restriction upon pure rationality. While the latter maximizes the given interests without moral limitations, fairness or benevolence, the former brings these factors into play by embedding them into the conditions of rational deliberation of the parties behind the veil. Thus, for Rawls the unity of practical reasoning is achieved only when reasonableness carries priority and appropriately limits rationality (CP 317).

Taking the various characteristics together, it should be clear that Rawls has a specific, close-knit and intertwined conception of person. All three pairs of equality and freedom, reasonableness and rationality, and a sense of justice and a capacity for a conception of good represent two broad poles of fairness vs. self-interest that I see as driving forward the Rawlsian understanding of human social and political nature. At one pole, rationality is implied in a capacity to adopt and pursue a conception of good, which in turn motivates freedom as a possibility to pursue these conceptions, once adopted. Being able to achieve (or at least to follow) what I want is the crucial self-interested desire. At the other pole, reasonableness presupposes moving beyond our self-interest, considering also what is right or fair. The possibility of this kind of thinking is guaranteed by our sense of justice, which, for Rawls, leads us to abandon the entrenched privilege in favour of equality.

The conception of person in Rawls thus consists in the two poles of fairness vs. self-interest being elaborated as the capacities of our reason (reasonableness and rationality), as fundamental political values (equality and freedom) and as the fundamental presuppositions of social cooperation (a capacity for a sense of justice and a conception of good). Certainly,

no one pair can be isolated from the others and Rawls uses this fact very often. As I claimed at the beginning of this section, he repeatedly analyses, elaborates, compares, and derives them from one another. However, they do not enjoy a completely equal standing. More often than not (as I shown above), the two powers of moral personality seem to be the ultimate ground from which Rawls develops and elaborates both reasonableness and rationality, and freedom and equality. This fact will become significant later on.

Finally, the last of the characteristics of humans presented at the beginning of this section stands a bit apart from the others. A human being is presented as a self-authenticating source of valid moral claims. With this complex and enigmatic phrase Rawls wants to deliver a crucial message: a person has her place in the moral world not because of any external factors, but solely because of her nature, because she is *that kind of being*. Human beings ought to be free and equal, ought to have a possibility to follow their conceptions of good life. Rawls claims (CP 330) that persons can put forward these claims not because they fulfil their obligations or because of their social status – but because they are the kind of beings they are. So the source of human moral claims is the shared humanity itself.¹³⁵

Taking human beings as self-authenticating sources of valid moral claims serves as a crucial bridge between the capacity and its realization. The sheer fact that human beings are free to a certain extent or possess a sense of justice may not have any impact on political system. But if human beings are the sources of valid moral claims, the whole theory of person gains important normative connotations. It is no longer about what a person can be, but also how these possibilities are to be protected by state and society at large. It is this very position that grounds the need to construct and implement a theory of justice that specifies and protects the normative claims stemming from the human beings themselves.

Having presented Rawls' conception of person in isolation, it is now necessary to specify the role it plays in the overall structure of his theory of justice. In this regard, one basic inference holds: the more important the conception of person is for Rawls' thinking and argumentation, the more it stands in need of justification. To anticipate a bit, in the rest of this section I will argue that a conception of person as it was presented above is absolutely paramount for nearly everything Rawls does. That makes an understanding of the appeal of Rawlsian conception of person (which is the task of the next section) a crucial part of

¹³⁵ This conclusion is also supported by terminological variations Rawls uses throughout his texts. While in PL he talks about "self-authenticating" sources of moral claims, *Kantian Constructivism* (CP 303-358) has it differently, as "self-originating" sources.

developing a charitable interpretation of Rawls. If the conception of person is well-grounded, then Rawlsian liberalism can be justified. On the other hand, if the conception of person is unwarranted and arbitrary, the whole elaborate construction that follows remains acutely unsupported.

To understand the role of a conception of person, recall the basic argumentative thrust of Rawlsian justice as fairness: Its aim is to establish a theory of justice manifested by concrete principles. The principles are constructed by consensus behind the veil of ignorance. The consensus is established when parties in the original position agree on the basic structure of society that would cover their primary goods. Resulting from this consensus are the two principles of justice that can function as the publicly acknowledged cornerstone of a stable well-ordered society.

The key turn in the process of establishing the two principles of justice comes with the introduction of the original position. The original position helps us reduce the hopeless multiplicity of possible principles of justice and more or less transparently derive the final two. For Rawls, the concept of the original position stands in between the conception of person and the ideal of well-ordered society (CP 308). The original position is the model conception, the device of representation that sets up our intuitive assumptions so that they are usable in the process of deducing the political principles (PL 25-27). This modelling of assumptions is at the end of the day very simple and it uses the conception of person presented above. It has two stages: (1) setting up the original position and (2) reasoning within the already specified conditions. Rawls claims, as I mentioned above, that the reasoning within the original position represents pure rationality – but since its exercise is constrained by the conditions in it, the final principles are reasonable as well (CP 316-319).

To explain it in a bit more detail, the representation of rationality behind the veil of ignorance is very straight-forward. Parties (understood as mutually disinterested) want to maximize the fulfilment of their interests, to get the most out of their conception of good life. Behind the veil, this pure rationality is the only motivating factor, it pushes out everything else.

But rationality cannot be the only factor behind political principles. Once pure economic rationality is applied in normal circumstances, the resulting situation will most likely be extremely unjust. If people were to push forward their interests with no restrictions, then those with more favourable starting position will be successful, while the rest will not, and very deep inequality will ensue. That is why it is necessary to limit rationality with moral

considerations such as equality or fairness – which is the task of reasonableness. Thus, unequal power and partial interests need to be removed from the deliberation.¹³⁶

The veil of ignorance represents the fundamental human equality in that it does not leave any space for parties to advance one's interests at the expense of others. No one knows where exactly he will end up, so all the rules of society are set up with an equal weight put on every one member. In the resulting society, no one is justified in complaining that his interests were overlooked or not sufficiently taken into consideration.

Freedom is represented by the ignorance of conceptions of good life behind the veil. No particular conception is privileged in any way there. Parties behind the veil of ignorance do not know their religion, hobbies or occupation. Therefore, principles agreed in such situation must be modelled so as to enable broad possibilities for very different lifestyles. The only serious limitation is that a conception of good adopted by one citizen must not obstruct the possibilities of another (though in practice it obviously cannot help influencing others).

Ultimately, all characteristics of the original position are rooted in the conception of person. Parties have an equal standing and they search for general principles able to cover any conception of good life which is possible under the conditions of equality (thus excluding slavers, paedophiles, inquisitors etc.). The construction of the veil of ignorance balances reasonableness with rationality through limiting the latter by the former. All this vindicates Rawls' claim that the construction of the original position really represents the conception of person in all its aspects (freedom, equality, reasonableness, rationality, powers of moral personality). To put it differently, *the original position is fully determined by the conception of person Rawls adopts*. Every aspect of its design is driven by it.

But that is not the end of the importance of the conception of person. In the established original position, parties deliberate about the most appropriate principles of justice. The deliberation is quite limited, since no partial interests and preferences are possible. The only features promoted by the constrained rational parties are the primary goods. They are the things that “any rational man is presumed to want” (TJ 62/54). For Rawls, cases like protection from suffering and general ability to follow a conception of good life belong under this category. In a list, these goods can be identified as “rights, liberties, opportunities, income

¹³⁶ That is also why Rawls argues that the veil of ignorance should be the thickest possible. In other words, out of many possible veils suppressing information, Rawls would pick the veil that suppresses the most of them (CP 335-336). Therefore, his veil covers not only social status and position, but also religion, power structure in society, state of technical advancement and so on. In the end, the only information that he allows are general facts about human psychology and basics of social sciences (TJ §24).

and wealth” (TJ 62/54 rev.)¹³⁷ (with respect added later on). Primary goods therefore enable formulation of rational preferences even in the original position. The most appropriate principles of justice are the ones that best cover and protect these primary goods. Without them, there would be no content to start the derivation of principles, no principles would follow, and therefore a theory of justice in Rawlsian sense would be utterly impossible.

When it comes to explaining and justifying the list of primary goods, *Theory* does not give a clear-cut answer, despite the fact that Rawls analyzes them on multiple occasions.¹³⁸ A suitable explanation is given only in *Kantian Constructivism in Moral Theory* (CP 313-315) and then in *Political Liberalism* (PL 75-76). There, Rawls claims that “to identify the primary goods we look to social background conditions and general all-purpose means normally needed for developing and exercising the two moral powers and for effectively pursuing conceptions of the good with widely different contents.” In other words, primary goods are designed to promote and enhance the basic capacities (the ‘higher order interests’) as specified by the conception of person. If Rawls did not adopt his specific conception of person, there would be no way to non-arbitrarily specify the primary goods. That means, once again, that the conception of person has a fundamental and extremely important role in the development of Rawlsian theory of justice.

When it comes to the two principles of justice deduced in the original position,¹³⁹ there is unsurprisingly a full correspondence between them and the conception of person introduced above. The freedom and equality are manifested in the first principle as a claim on the fully adequate share of basic liberties for everyone. The second principle pursues the economic reasonableness by demanding equality of opportunity and the best position possible for the worst off (TJ 302/266 rev.).

It seems that every single one of the important argumentative steps in Rawls hangs on the conception of person. Without it he would not be able to suitably construct the original position, specify primary goods or derive the two principles and his ideal of well-ordered society would have been much less compelling. And that is a very important conclusion. It

¹³⁷ Cf. PL 181.

¹³⁸ For example §§11 and 13.

¹³⁹ The deduction of the two principles was by far the most commented bit of the entire Rawlsian corpus in the decades after the publication of TJ, although the trend seems to have changed recently. And that is in spite of (or maybe exactly because of) the fact that Rawls’ argumentation there is often unsatisfactory and sometimes quite unconvincing. But since the controversies regarding the formulation of the two principles of justice are not my primary focus, I will content myself with a few general observations about the principles themselves. For a comprehensive discussion on the deduction of the two principles, see Thomas W. Pogge, *Realizing Rawls* (Ithaca: Cornell University Press, 1989).

shows that status and exact shape of ‘intuitions’ has a much more deep-lying significance in Rawls that is apparent on the first reading.

All this has radical consequences on the question of justification of Rawlsian liberalism. It means that *justification of Rawlsian liberalism, for all intents and purposes, boils down to the justification of his ‘intuitions’ (i.e. his full conception of person)*. This is quite a radical statement, but it is by no means unprecedented in the literature. To take a prominent example, Habermas, in his notable argumentative exchange with Rawls after the publication of *Political Liberalism*, claims that

“the decisive issue in the justification of the two highest principles of justice is less the deliberations in the original position than the intuitions and basic concepts that guide the design of the original position itself. Rawls introduces normative contents into the very procedure of justification, above all those ideas he associates with the concept of the moral person: the sense of fairness and the capacity for one's own conception of the good. Thus, the concept of the citizen as a moral person, which also underlies the concept of the fair cooperation of politically autonomous citizens, stands in need of a prior justification.”¹⁴⁰

In the next section, my aim is to show how this ‘prior justification’ can look like, which is by no means clear.

4. The two interpretations

So far, I have shown that the conception of person in Rawls forms the crucial piece of his theory. So much so that every key argumentative turn Rawls makes is supposed to follow from it, or is at least underpinned by it in some way. Under these circumstances, the justification of Rawlsian liberalism becomes almost equivalent to the justification of his conception of person. Of course, the passage from the conception of person to his well-ordered society governed by the two principles of justice can be (and has been) challenged as well. Still, if there are powerful reasons to adopt the conception of person as it is present in *Political Liberalism* and *A Theory of Justice* (with free and equal, reasonable and rational citizens possessing moral personality that are the sources of valid moral claims), the ensuing political regime should end up being broadly liberal even if Rawls did make some grave mistake in his derivation of the two principles of justice.

¹⁴⁰ Jurgen Habermas, “Reconciliation Through the Public Use of Reason: Remarks on John Rawls’s *Political Liberalism*,” *Journal of Philosophy* 92, no. 3 (1995): 119.

Unfortunately, even though the stakes with regards to the justification of his conception of person are really high, Rawls is less than clear on the issue, which is probably why his conception of person was so often challenged and misunderstood.¹⁴¹ In his texts, he oscillates between what I call the descriptive and the deliberative interpretations. The two of them provide very different strategies for grounding the conception of person – and thus we have two different answers to the question of liberal justification.¹⁴²

4.1. *The descriptive interpretation*

The descriptive interpretation claims that our intuitions and a conception of person indeed play an important role in the choice of the principles of justice, but there is no foundationalist argument to establish them as correct or sound. Rawls wants only to conceptualize the intuitions about social justice and make them coherent and capable of yielding sufficiently strong and specific principles. His aim therefore is not to defend liberalism, only to elaborate it – to show how, after we adopt a certain set of values that are broadly acceptable in democratic societies and transform them into the limitations of original position, a liberal conception of justice follows. If he can show that values standing behind his conception of justice are in fact in agreement with our liberal democratic considered judgments, and that his principles of justice are coherent and intuitive, his goal is achieved and the resulting theory of justice is in reflective equilibrium.

¹⁴¹ In fact, it was Sandel's critique of Rawls' conception of person that sparked maybe the most extensive debate in political philosophy for the last 50 years, between liberalism and communitarianism (Michael J. Sandel, *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1998)). The discussions between libertarians and communitarians were far-reaching and complex, but I would like to present a short thought nevertheless (without the proper argument, for which there is no space). The critique of a liberal conception of person is based on a misunderstanding. Rawls does not want to say that humans are *merely* reasonable and rational, free and equal, with a thin moral personality; that interpersonal values and identities do not reach the very core of human existence; that the self is "unencumbered". Rawls only claims that a person has a basic capacity to adopt and revise its conception of good life. Translated into the political context, this means that a political identity of a citizen does not change when his religious identity does. This is a very precarious point for communitarians. On one hand, they do not want to abolish liberal freedoms and open the door for religious oppression, but on the other they still fantasize about a "traditional" society that pushed its members very strongly to conformity. Put in the simplest terms, Rawls' conception of person is political, not metaphysical and we ought to view it as such. Still, the fact that it is political does not exempt it from the justificatory challenge I present here.

¹⁴² The deliberative and descriptive interpretations follow the above-mentioned two Scanlon's interpretations of reflective equilibrium (Scanlon, "Rawls on Justification."). They only carry these two interpretations further, understanding them as encapsulating the two strategies of justification of Rawls' intuitions/conception of person as I see them.

This interpretation relies heavily on the notion of ‘public culture’ used in the first lecture of *Political Liberalism*. The deliberation about the principles of justice is supposed to start “by looking to the public culture itself as the shared fund of implicitly recognized basic ideas and principles” (PL 7). Defining it, “public culture comprises the political institutions of a constitutional regime and the public traditions of their interpretation (including those of the judiciary), as well as historic texts and documents that are common knowledge” (PL 14). Public culture determines how we understand ourselves as citizens of liberal democratic societies. The conception of person as free and equal, reasonable and rational, etc. is the result of values shared in our common normative heritage. And this very fact presents the ultimate justification of why this conception of person and not any other has to be represented behind the veil of ignorance.¹⁴³

In the descriptive interpretation, we take the conception of person as we find it in public culture. But the crucial question here is: what do we actually describe? The notion of public culture in Rawls is quite thin, as the two quotes above are the only places I could find in *Political Liberalism* that shed some light at it. There are two options here: either we describe the real-life self-understanding of citizens, or we describe a political ideal of citizen in liberal society. In other words, the conception of person may be something that almost every citizen actually accepts, or it may be a hypothetical model we ought to approximate. There is, unsurprisingly, a world of difference between these two options.

Addressing this issue in his recent and powerful presentation of Rawlsian political liberalism, Jonathan Quong proposed an interpretative dichotomy of internal vs. external conception of political liberalism that well encapsulates this problem.¹⁴⁴ When seeking for a starting point for political liberalism, one may search for it ‘outside’ – trying to find some common ground for consensus within the worldviews held by real people in contemporary liberal democracies. That seems to be the more ambitious way. The other strategy is to focus ‘internally’ on coherence and feasibility of the liberal ideal itself. If the ideal of well-ordered society fuelled by the liberal conception of person proves to be coherent and attractive despite the fundamental disagreements between free and equal citizens, then the Rawlsian liberalism is vindicated. Quong himself prefers the internal conception – and that enables him to

¹⁴³ It is important clear out one possible misunderstanding that may result from previous paragraphs. In spite of the fact that the notion of public culture is introduced only in PL, this does not mean that TJ cannot be read descriptively. Indeed, TJ frequently operates with the intuitions and considered judgments that we presuppose and consider as correct – although it fails to explicitly elaborate on their source.

¹⁴⁴ Jonathan Quong, *Liberalism without Perfection* (Oxford: Oxford University Press, 2011).

comfortably get the growing pile of critical literature on later Rawls out of the table and to introduce a reading that is much more charitable.

Quong, in agreement with many other critical commentators¹⁴⁵, holds that the external reading of political liberalism boils down to assessing the putative empirical fact that the idea of cooperative society with free and equal, reasonable and rational citizens does really correspond with the traditions and real-life political attitudes as we encounter them in our political societies. In an early pursuit of this line of criticism, Gerald Doppelt claims that the conception of person as free and equal, which constitutes the starting point in Rawls, has no automatic priority.¹⁴⁶ It is by no means rooted deeper in our cultures than Christian virtues or capitalist ethics – it is exactly the other way around. While Christianity stands truly at the beginning of our culture and civilization somewhere deep in the Middle-Ages and capitalism catapulted Europe and later North America into the world economic dominance centuries ago, universal freedom and equality were still viewed as radical and idiosyncratic values at the end of the 19th century. And even in the 20th century there were times when they seemed to be completely abandoned. That is why they cannot claim any privileged place as the first and foremost cultural values of our civilization that are somehow automatically presupposed.

Even taken a-historically, Rawlsian freedom and equality are far from established as the principal political values endorsed across the political spectrum. If we look at what people can *really* agree on, we get a hopelessly complex picture which moreover radically varies from country to country. In the US, it probably resembles some sort of constitutional consensus on basic rights and liberties while Scandinavian countries supplement it by a robust measure of social provisions. What is more, this consensus may rapidly change in time, reflecting momentary fluctuations of national economy, international security, immigration situation, and other wholly contingent factors. Establishing a political conception of person based on what real reasonable citizens can actually agree upon is therefore a precarious exercise, which would, even when successful, yield distinctly non-Rawlsian conclusions. Therefore, political liberalism interpreted externally fails as it cannot sufficiently defend its core values, critically important in constructing the conception of justice.¹⁴⁷

¹⁴⁵ See for example Robert S. Taylor, *Reconstructing Rawls: The Kantian Foundations of Justice as Fairness* (University Park, Pa.: Pennsylvania State University Press, 2011); George Klosko, "Political Constructivism in Rawls' Political Liberalism," *The American Political Science Review* 91, no. 3 (September 1997).

¹⁴⁶ Gerald Doppelt, "Is Rawls' Kantian Liberalism Coherent and Defensible?," *Ethics* 99, no. 4 (1989): 815–51.

¹⁴⁷ This is, obviously, only a rough rendering of a few crucial arguments. For a more detailed discussion see Quong, *Liberalism without Perfection*, chap. 5.

A similar line of argument is proposed by broadly Marxist authors.¹⁴⁸ According to them, Rawls ignores meaningful work or a pursuit of collective interests, both of which are broadly valued. Thus, he does not have even a remotely representative conception of person.

The core of this argumentation is relatively simple and it stays the same across different authors and articles: When creating the conditions for deduction of principles of justice, Rawls uses *some* intuitions about person, citizens, and their place in society. The description of person that Rawls adopts is thus excessively selective, excluding other important values that are deeply entrenched in actual political thinking of many citizens. The preferential treatment of freedom and equality taken as implicit in actual political views of citizens of liberal democratic countries is therefore very much unjustified. They still may be the values we uniquely ought to pursue – but such a claim requires a completely different justification. The external reading seems hopelessly inadequate.

In spite of some textual evidence to support this reading of *Political Liberalism*, philosophers like Quong remain unmoved. He claims that the correct (or, at least, the only charitable) reading of Rawls' theory is internal, regardless of what certain passages seem to suggest. Political liberalism need not justify itself to the existing constituency of citizens by searching what sort of consensus is achievable given the radically divergent political conceptions they hold. Its method is hypothetico-deductive. It starts from a certain ideal of society, from citizens as free and equal, from the doctrine of the burdens of judgment – so the consensus that the external conception aims to achieve is already presupposed. The discussion then proceeds only within the bounds of the liberal democratic tradition and its aim is to establish the principles of justice and the ideal of well-ordered society as internally consistent and attractive. What Quong (and, in this interpretation, Rawls) in the end want is to present a picture of liberal society as stable and legitimate despite far-reaching disagreements in the conceptions of good life and pervasive pluralism of the reasonable comprehensive doctrines held by its citizens.

This strategy treats the conception of person adopted by Rawls as given (there is still no deliberation; the conception is a description of a liberal ideal). It claims that as an 'internal' conception it works only within the liberal framework and therefore does not need to justify and explain its normative starting point. The possible justifications of values are outsourced to the comprehensive doctrines, outside the scope of what all reasonable liberal citizens are

¹⁴⁸ Milton Fisk, "History and Reason in Rawls' Liberalism," in *Reading Rawls*, ed. Daniel Norman (Stanford: Stanford University Press, 1989); Andreas Eshete, "Contractarianism and the Scope of Justice," *Ethics* 85, no. 1 (1974): 38–49.

expected to endorse. Quong here takes seriously the formulation of the task Rawls gives himself at the very beginning of *Political Liberalism*, where he wants to answer the crucial question “how it is possible for there to exist over time a just and stable society of free and equal persons, who remain profoundly divided by reasonable religious, philosophical, and moral doctrines?” (PL 4). If justice as fairness provides an answer, then liberalism is internally consistent and attractive – and that may be all Rawls wants to achieve. The adopted conception of person is presupposed automatically, so to say, since we reason from within the liberal ideal. For this task, a philosophical justification of the sort I look for is not needed.

However, the claim that the Rawlsian normative starting point can be taken for granted since the discussion is internal to liberal theory has one important presupposition: the Rawlsian conception of person must pass as a commonplace liberal notion widely shared within the tradition. If that is not the case, and I will argue that it is not, then the fact that we view political liberalism as internal to the wider liberal tradition cannot exempt Rawls from properly spelling out and justifying his underlying normative conception of person and society. To put it differently, Rawls employs the intuitions “implicit in public culture” when stipulating the nature of human beings as free and equal. If it turned out that his intuitions are really *his* and are not widely shared even within the liberal tradition itself, then he cannot treat them as undoubted and given. If he still treats them as given, it would mean that he is open to the charge of arbitrariness – a claim that his theory starts with a normative starting point that is not justified and therefore inherently problematic.¹⁴⁹

To repeat, Rawls takes a person to be reasonable and rational, free and equal, possessing the two powers of moral personality, and being a self-originating source of valid moral claims. Taking Quong’s interpretation of political liberalism at face value, Rawls simply assumes this normative conception of person, asserting that it does not need to be justified as it is inherent to the liberal project itself. But this is not true, as even the most superficial comparison with other liberals can prove.¹⁵⁰ For contrast, take the example of John Stuart Mill, whom I use here as the representative proponent of the most basic liberal tenets, a ‘benchmark’ liberal when it comes to basic principles. Both Mill and Rawls of course believe in freedom, but in strikingly different ways. While Rawls, as I elaborated above, associates freedom with the two powers of moral personality, Mill associates it with what he calls

¹⁴⁹ This is a line against Rawls I discussed earlier, raised already in 1973 by Hare, “Review.”

¹⁵⁰ Of course, if one picks someone like Dworkin or Nagel for comparison, then Rawls’ take on liberal values may not stand out as too idiosyncratic. But that is only because they, as a group, exercised heavy influence upon each other - which resulted in some important similarities between their respective philosophical projects.

‘experiments in living.’ For Mill, we should be free to organize our lives as we please, because that leads to experiments in living which broaden our knowledge of what a good life may consist in.¹⁵¹ That profoundly differs from what Rawls has to say both in content and in philosophical implications. Claiming that Rawls’ notion of freedom is only a garden-variety liberal conception is thus clearly misguided.

A similar conclusion follows from the comparison of Rawls’ idea of ‘fair cooperation between free and equal persons’ with Mill. In *On Liberty*, it is the so-called ‘harm principle’ that serves pretty much the same purpose as a fair cooperation between the free and equal in Rawls – it is the basic regulative ideal which shapes the role of the state. Yet these two are quite distinct in scope as well as in ideological background. While Mill’s main object is to philosophically capture the limits of state power (which is historically maybe the most important liberal project), Rawls’ concern for ‘fair cooperation’ is much broader. In any case, Rawls and Mill take profoundly dissimilar approaches to the basic questions of liberal politics – in spite of sharing some important “morphological” family resemblances.

If we take both Mill and Rawls to be liberals, then ‘the question internal to liberalism’ must not just presuppose some quite specific normative background (be it harm principle or moral personality) and simply elaborate it. The form and justification of the ‘liberal’ conception of person is unclear even within the liberal theory itself.¹⁵² Hence, Rawlsian values cannot be assumed away at the start of the discussion as ‘implicit in democratic culture’ or, even more narrowly, ‘implicit in liberal tradition’. When compared to other liberal conceptions *they do stand out* – and therefore need some sort of argument covering them, especially given the fact that his conception of person is so crucially present at every turn of the argument Rawls makes.

As a counterargument, Quong in his book claims that to assume liberal values at the beginning does not beg the question with regards to the liberal principles derived at the end. He argues that enough is happening in the argument to prevent the charge of circularity – general liberal values are specified into a concrete theory of justice which is attractive and also stable over time. Hence, from his perspective, my argumentation can be criticized as misconstruing the nature of justification in Rawls.

The proponents of internal interpretation of political liberalism can claim that Rawls does not need any argument proving that his conception of person is somehow uniquely

¹⁵¹ See Mill, *On Liberty*.

¹⁵² Indeed, this might be the most basic presupposition of this thesis. If a form and justification of a liberal conception of person were clear, the very existence of this thesis would be utterly pointless.

persuasive and thus better than all possible alternatives. The strength of Rawls' liberalism does not come from the solidity of its assumptions. All he needs is their plausibility. He is not a foundationalist. What is important is the theory as a whole, and its persuasiveness is provided by synergy of multiple independent and by themselves partial deliberations, values and arguments that provide support for each other, so that the resulting construction is more solid than any of its parts. One of those parts is of course the conception of person, but it does not need to be bullet-proof, only mildly intuitively attractive. The real force of the overall construction is only shown when this intuitive attractivity is added to the overall coherence, to the force of the deduction of the principles, to the fairness of the original position, to the fact that the resulting society would be stable for the right reasons and compatible with human psychology, etc. The argumentation of Rawls is not one big continuous deduction starting with a conception of person, but a complex structure supported on multiple points by partial deliberations – whose real beauty only comes out when we focus on the whole.¹⁵³

Thus, if we take the holistic structure of justification in Rawls seriously (the objection continues), the fact that the adopted conception of person is not uniquely implicit in liberal tradition ceases to be problematic. It suffices for Rawls to show its plausibility, with additional justification provided by the structure as a whole. And when I push Rawls to make the support of his conception of person stronger, I only show my misunderstanding of his overall ambition. Simply put, he does not need an argument of this kind.

The presented objection to my critique is persuasive and to some extent true. Rawls really thinks (and sometimes explicitly claims) that the best argument for his theory is indeed the theory as a whole. He also believes that there is more than one source from which this theory can gain its plausibility, so an intuitive conception of person does not stand alone. This would mean that my contention from the beginning of this section (that the justification of liberalism in Rawls amounts to the justification of his conception of person) is horribly wrong.

However, the conception of person still presents an absolutely crucial piece of puzzle, and there is no way around it. Rawlsian political liberalism does start with a highly specific and surely not commonplace conception of moral personality, which is deeply connected to the basic values of freedom and equality. Rawls would not be able to derive his principles of justice without these normative assumptions.

¹⁵³ A similar reading of the method in Rawls is provided by Onora O'Neill, "The Method of A Theory of Justice," in *John Rawls: Eine Theorie Der Gerechtigkeit*, ed. Otfried Hoffe (Berlin: Akademie Verlag, 1998), 206–18.

Of course, not all philosophical work in Rawls or Quong is done by their normative assumptions. Nevertheless, for my purposes it is sufficient to claim that *enough* of the philosophical work is done by the normative conception of person for it to require a proper justification. Without this justification, there seems to be no reason to accept the elaborate theoretical structure as a whole – especially when Rawls is leaning on his conception of person at every step of the argument. At this point, the basic building blocks of Rawlsian theory are just assumed away with no polemical force and the further construction just mirrors these fundamentals – so there is nothing to persuade the critics to change their opinion.

Thus, the internal reading of *Political Liberalism* collapses. It is just not possible to read the conception of person in Rawls as straight-forwardly presupposed within the liberal culture. If that is indeed the case, there needs to be some argument for it, as the issue is very far from self-explanatory. Under the descriptive interpretation, neither Rawls nor Quong seem to have such argument. Moreover, it is also not possible to gloss over this problem claiming that the lack of justification of the conception of person is somehow supplemented by the overall plausibility of the entire structure of justice as fairness. Since, as I argued above, the whole theory of justice in Rawls is upheld by the conception of person, this step would amount to a very clear case of bootstrapping as no independent argument or broader consensus helps justify Rawls' starting point. The argument is pushed to a large extent by his (as I argued) seemingly gratuitous assumptions.

The descriptive interpretation of Rawlsian political thought (in its internal as well as external variant) is therefore neither very appealing by itself, nor it sufficiently answers my question. Throughout this thesis, I ask what we can philosophically say to justify liberalism as a mode of existence of political society. Here, Rawls has a very elaborate story to tell: we should accept his liberal principles of justice, because they are fair. This fairness lies in the principles adequately representing the idea of social cooperation between citizens viewed as free and equal. To ensure this adequate representation, a complex structure is constructed, with the original position at the centre. However, as all this (or at least much of it) rests on the conception of person, Rawlsian justification of liberalism cannot go through unless this conception is justified as well. So far, the descriptive attempt to achieve this justification failed.

4.2. *The deliberative interpretation*

In the core of the second interpretation lies an ambition not only to use the intuitions concerning the conception of person, but to defend and justify them as well. Rawls is here understood as trying to bring out the most appropriate conception of justice via vindicating his normative starting point of free and equal citizens possessing a moral personality. It is in this light that we can read quotes like: “My aim has been to indicate not only that the principles of justice fit our considered judgments but also that they provide the strongest arguments for freedom” (TJ 243/214 rev.). That would imply that Rawls wants to provide reasons for freedom, not only assume it.

In this interpretation, our ‘intuitions’ enter the search for reflective equilibrium (via the original position) not as given facts about our traditions, but as assumptions to be proved and verified.¹⁵⁴ The whole theory would be complete and the reflective equilibrium would be achieved if it became clear a) what values stand at the core of our theory of justice; b) that these values are well-established; c) that the deduction starting from these values via original position really yields the principles specified in the theory; and d) that the values, adopted principles, and their application together form a coherent system. The resulting structure would form a sort of ‘bridge’ (CP 308) through which our values and assumptions (the Rawlsian conception of person) are critically evaluated and eventually transformed into political principles. This bridge is provided by the original position behind the veil of ignorance. As should be obvious by now, in this thesis I am preoccupied mostly with a) and b). With a) analysed in the previous sections, I focus now on b).

One crucial advantage of this interpretation is its immunity to the arbitrariness objection. The intuitions concerning the conception of person play an important role but they are no longer blindly assumed as an integral part of the tradition or as being widely shared by the citizenry – they are critically assessed in the process of constructing a theory of justice. Therefore, they are not exposed to the arguments I presented above. Of course, their justification may be a source of disagreement, but at least it is not resigned upon from the start, like in the descriptive interpretation.

Another advantage, especially with regards to the main task of my thesis, lies in the fact that the deliberative interpretation aims to provide a full-blown defence of liberalism. When adopting the descriptive interpretation, a critic may always claim that she simply does

¹⁵⁴ See my analysis of the deliberative interpretation of reflective equilibrium above.

not share the normative commitments of the theory. Therefore, she has no reason to acknowledge the results deduced from them. On the other hand, she does not have this option with the deliberative interpretation – she must confront the provided argumentation. Rawlsian liberalism under this interpretation aims to be justified further than in the narrow circle of people already accepting Rawlsian assumptions.

However, there are looming difficulties with the very ambition of the deliberative interpretation, even before the argumentation starts. The trouble is that arguing for one specific conception of person as the most appropriate and thus as uniquely admissible to regulate political justice may run against the deeply ingrained pluralism of our societies. To sketch the argument (which is very important for Rawls in his later work¹⁵⁵), the modern societies are inevitably pluralistic. Therefore, any successful political theory needs to embrace this pluralism and search for a conception of justice that is solely political, not comprehensive, and, as a result, does not run against any reasonable worldviews. Thus, deep philosophical and metaphysical explanations, especially when they concern the very justification of liberal regime, are excluded from the start, as we want this justification to be open to humanists, Christians, Muslims, Kantians or utilitarians alike. The defence of the adopted conception of person is therefore a very delicate affair. It must keep strictly within the bounds of the political – and it is really not straight-forward how the argument should proceed.

The delicacy of the problem is underlined by the fact that the deliberative interpretation tries to unearth parts of Rawls' thought that are not always explicit, not very systematic, and sometimes difficult to follow. Rawls of course openly uses many different values, but their justification, as I noted above, is often left in the background (which is why the descriptive interpretation is textually sound and maybe even the most natural way to read Rawls). For the deliberative interpretation to work, it must uncover what Ronald Dworkin labelled "the deeper theory of Rawls".¹⁵⁶

Dworkin identifies what he calls a "surface argument"¹⁵⁷ which runs via reflective equilibrium and the original position to the derivation of principles of justice and well-ordered society. But these are for Dworkin only supplementary conclusions, which can lead us to uncover a deeper structure of values standing behind them, identified above as Rawls' conception of person. Dworkin's motivation is thus very similar to mine: he is not satisfied

¹⁵⁵ Especially in PL and in *Justice as Fairness: Political not Metaphysical* (CP 388-414)

¹⁵⁶ Ronald Dworkin, "The Original Position," in *Reading Rawls*, ed. Norman Daniels (New York: Basic Books, 1989), 26.

¹⁵⁷ *Ibid.*

with the surface argument and claims that it would make sense only if we can properly identify the elements standing behind it. However, in spite of these resemblances, my interpretation is very different from Dworkin's, whose reading of Rawls in the quoted text resembles some of his own philosophical positions to an almost comical extent.¹⁵⁸

The question that the deliberative interpretation so crucially asks, to frame it precisely, is *why we must value the moral personality so much as to place it at the very heart of political philosophy*. In other words, we have a conception of person and we need to show that it is somehow uniquely normatively relevant for political society – that it ought to guide and regulate it. If Rawls cannot show why these capacities are primary for justice, then justice as fairness is, as I have shown, quite arbitrary and inadequate for the task of justifying liberalism.

In spite of the fact that the conception of person stays virtually unchanged throughout Rawls' works, its justification, as demanded by deliberative interpretation, changes dramatically. In *A Theory of Justice*, Rawls relies on so-called Aristotelian principle (TJ §65) to provide a key account of human motivation. It claims that "other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity" (TJ 426/374 rev.). This principle purports to explain our deeply entrenched interest in developing our capacities, abilities and skills – and the more complex they are, the higher the motivation we can have to develop them.

The Aristotelian principle then allows Rawls to look at our most complex capacities and pronounce that they constitute 'the highest order interests' of human beings (CP 334). These capacities, the ability to adopt a conception of good and to cultivate a sense of justice, are for Rawls absolutely necessary for any exercise of practical reason taken as an ability to shape and influence one's life, make authentic decisions, embrace goals and deliberate about means for their fulfilment. Their development is necessarily demanded, because it is presupposed in all other possible developments of genuinely human capacities.

Rawls therefore concludes that human powers of moral personality, which stand at the heart of his conception of freedom and equality, are morally (and therefore politically) imperative. Their development stems from the deepest ('highest order') interests present in

¹⁵⁸ This is one more instance of an often repeated case when one great philosopher interprets another. In such circumstances, the resulting interpretation is often very interesting, but almost never balanced and strictly true to the text (the prominent case being Kripke's Wittgenstein). The interpreting philosopher often imports into the interpreted text a lot of her own original ideas, which often pushes it far beyond the original author's intentions.

every human being (TJ 159). Indeed, these powers stand at the very core of what it means to be a human being. Therefore, not providing options for realization of these capacities for some person or group equals denying their humanity. If we are to treat people as people, we need to respect them by respecting their powers of moral personality.

The Kantian allegiances of the above position are more than clear. The most important attribute of human beings is their moral and personal autonomy. It is this autonomy that needs to be developed and supported in our political societies. If it is not, a great moral wrong is done.

A relatively recent exposition of this strand of Rawlsian thinking came from Robert Taylor.¹⁵⁹ His *Reconstructing Rawls* is an interesting attempt to prove that Rawls makes best sense if he is understood as a profoundly Kantian thinker, while once he turns his back on Kant and tries to develop political liberalism, his theory quickly loses plausibility. Taylor holds that the best way to understand moral personality is in a strict analogy with Kantian autonomy. In his view, Rawlsian justice as fairness is a systematic development of Kantian morality into political philosophy.

I will not attempt to directly confront Taylor's interpretation of Rawls, as it would take me far from my central purpose. The Kantian strand in Rawls' thinking is certainly strong, especially (but not exclusively) in his earlier work. Nevertheless, I think Taylor underestimates the factors that drive Rawls to abandon this 'comprehensive' justification for justice as fairness. Rawls claims (and I fully agree) that a modern pluralistic society cannot have political foundations that depend so much on a comprehensive philosophical doctrine, especially as divisive as Kant's. Moreover, it is not clear if the Kantian conception of person can really become plausible outside of the system of transcendental idealism, one that scarcely anyone supports today.¹⁶⁰

If Taylor is right, then any proponent of liberalism faces a dilemma: either she takes pluralism seriously, as later Rawls tries to do, and must therefore relinquish providing a proper justification of liberalism, or she provides a proper justification of liberalism knowing that it will contain assumptions and positions that not every reasonable person could subscribe to. Taylor emphatically chooses the latter option.

However, in my quest for liberal justification I refuse to accept Taylor's dilemma just yet. I think that Taylor underestimates the options Rawls has at his disposal that are strictly

¹⁵⁹ Taylor, *Reconstructing Rawls*.

¹⁶⁰ Interestingly, Taylor views the controversial nature of the Kantian conception of person as an advantage and claims that a non-contested, non-comprehensive justification of liberalism is not possible. For him, any theory that justifies liberalism must have deep metaphysical roots.

political, not metaphysical. In *Political Liberalism*, Rawls *does* try to justify the pre-eminence of moral personality in considerations of social justice. For that purpose, he uses ‘the idea of social cooperation’. Indeed, just as freedom and equality are anchored in the moral personality, the moral personality seems to be anchored in Rawls’ notion of social cooperation.

Social cooperation for Rawls “is distinct from merely socially coordinated activity. ... (It) is guided by publicly recognized rules and procedures.” “Cooperation involves the idea of fair terms of cooperation: these are terms that each participant may reasonably accept, provided that everyone else likewise accepts them.” Also, “the idea of social cooperation requires an idea of each participant’s rational advantage, or good” (PL 16). So the concept of person is not arbitrary, since “it has been understood, in both philosophy and law, as the concept of someone who can take part in, or who can play a role in, social life, and hence exercise and respect its various rights and duties. Thus, we say that *a person is someone who can be a citizen*, that is, a normal and fully cooperating member of society over a complete life” (PL 18, emphasis added). In other words, human cooperation in its very core contains an idea of fairness (that is, cooperation is understood from the start as being subjected to moral categories). Also, cooperation needs to be motivated by some kind of interest, a sense of gain broadly conceived. Without these two factors, the Rawlsian notion of cooperation is impossible. This allows him to claim that “since persons can be full participants in a fair system of social cooperation, we ascribe to them the two powers ... a capacity for a sense of justice and a capacity for a conception of good” (PL 19).¹⁶¹

Rawls thus does not present a general conception of persons and their nature – he limits himself to the claims about persons insofar as they cooperate. This conception of person is thus strictly political, confined to the realm of human institutional interaction.¹⁶² Here, Rawls cannot speak of ‘the highest order interests’ of human beings. Nevertheless, he asserts that “since citizens are regarded as having the two powers (or moral personality), we ascribe them two corresponding *higher* order interests in developing and exercising these powers” (emphasis added). “Someone, who has not developed and cannot exercise the moral powers to the requisite minimum degree cannot be a normal and fully cooperating member of society over a complete life” (PL 74). Hence, Rawls tries to identify what features the people possess are necessary for their role as citizens. When he identifies them as the two powers of moral

¹⁶¹ Essentially the same argument (perhaps even more clearly and more concisely put) is presented in PL 300-302.

¹⁶² The word ‘institutional’ here is understood as ‘being guided by rules and procedures’ – see the quote from PL 16 above.

personality, he can assert that the corresponding conception of person is indeed unavoidable for the requisite form of cooperation – and therefore cannot be arbitrary. If this argument is sound, we can conclude that Rawls is justified in using this conception of person as a driving force behind his construction of justice as fairness.

However, Rawls' argument basing moral personality on the idea of social cooperation seems to face a quite conspicuous problem: the idea of cooperation seems to be tailor-made to fit (and demand) the two powers of moral personality. In the quote above (PL 16), Rawls distinguishes “merely socially coordinated activity” and “the idea of social cooperation”. The latter seems to presuppose a relatively clear picture of how a fair political interaction should look like – it already contains a normative ideal. It is still relatively imprecise, but it does incorporate all that Rawls needs: a conception of person that should guide our view of justice understood as fairness in cooperation.

This means that the justification of moral personality via social cooperation can be viewed as merely another unilluminating normative move. Rawls seems to kick the ball one step further, without really meeting the challenge. What seems to hang in the air now is the above-presented idea of social cooperation. Why should we accept this specific notion of cooperation? Why is “socially coordinated activity” not good enough? These are the questions that Rawls does not answer.

It seems we have come to the end of the Rawlsian road. I could not find any further justification of his position. It stops with the idea of social cooperation and the two powers of moral personality mirroring each other.

Since the descriptive road of taking these concepts as implicit in democratic culture is not open to Rawls, as I argued in the preceding section, the deliberative interpretation is the only way to read him as trying to justify liberalism. However, at this point it seems that his reasoning proceeds from a liberal idea of cooperation towards a liberal conception of person, through a liberal device of representation (the original position), finally ending with the two liberal principles of justice. In the end, we are back to the Ricoeur's circle of demonstration, where we both start and end with a normative ideal of social cooperation, which is ‘only’ developed and purified in this process.¹⁶³

Still, what Rawls offers us is very impressive. It is a full-blown liberal conception of justice proposing at least some answer to most traditional problems of political philosophy. When it comes to my fundamental question (“what can we philosophically say to justify

¹⁶³ Paul Ricoeur, “Le Cercle de La Demonstration,” in *L'individu et Justice Sociale (autour de John Rawls)*, ed. Catherine Audard (Paris: Seuil, 1987).

liberalism as a mode of existence of political societies?”), his answer is robust and elaborated. He claims that liberal justice answers the call for fairness implicit in human cooperation, that it reflects the way we think about ourselves as free and equal citizens. Nevertheless, his answer is incomplete. He does not discuss the reasons why we should think of ourselves politically first and foremost as free and equal citizens possessing the two powers of moral personality (a sense of justice and the capacity to adopt and follow a conception of good), or why we should subscribe to a clearly normative ideal of human cooperation that presupposes the specific combination of fairness and self-interest that is so heavily present throughout Rawls’ work. All these are very much unfinished questions.¹⁶⁴

In the next chapter, I will try to prove that the normative ideal of social cooperation mirroring the two powers of moral personality is not an argumentative dead-end. If supplemented by additional arguments and considerations, albeit ones which are not explicitly present in Rawls, this liberal starting point can be persuasively vindicated, thus answering my original question. With that, my aim for this thesis will be complete.

¹⁶⁴ For interesting alternative elaborations of these (or roughly similar) Rawlsian problems, see Brian Barry, “John Rawls and the Search for Stability,” *Ethics* 105, no. 4 (1995): 874–915; Leif Wenar, “Political Liberalism: An Internal Critique,” *Ethics* 106, no. 1 (1995): 32–62; Sylvia Burrow, “Reasonable Moral Psychology and the Kantian Ace in the Hole,” *Social Philosophy Today* 17 (2001): 37–55. Most authors in this list seem to agree that the problems Rawls faces with the insufficiently defended conception of person are due to his implicit espousal of Kantianism. However, I agree with (later) Rawls against Barry in thinking that a reversal back to explicit endorsement of comprehensive universalist Kantianism is not a good strategy for justifying liberalism in our pluralist societies. The next chapter strives to show an alternative way forward.

CHAPTER III

1. The argument so far

This thesis is largely preoccupied with a single big task, trying to find ‘what can we philosophically say to justify liberalism as a mode of political existence of society.’ While not offering any specific answer so far, I did make some headway. One important consideration established in Chapter I was the importance of a relatively developed background behind any answer that wants to be credible and persuasive.

It proved to be extremely problematic to justify liberalism straight-forwardly, as value pluralists aimed to do. Isaiah Berlin, unlike his value pluralist successors, seems to have acknowledged this fact. He presented considerations on the issue that are irreducible to a simple argument from pluralism to liberalism, in spite of the fact that he is often read reductively, as trying to do exactly that. However, while the overall picture of complex interconnections between freedom and plurality is attractive, Berlin still did not offer a discernible and persuasive argument answering my question. The vital context is missing – we do not learn exactly how the normative situation of political societies can be changed by the status of human beings.

Self-ownership, on the other hand, provided the solution to the last problem that was simply too easy. A *fiat* stipulating a massive political importance of human self-relation in the end proved to be neither credible nor persuasive. Even though many philosophers including myself (as will be evident in this chapter) believe that a normative self-understanding of human beings should play a role in trying to find an appropriate form of political society, a simple stipulation is not enough. What is needed, again, is a developed train of thought covering and explaining this stipulation so that the acceptance of a normative conception of person does not become a mere subject of an empty vote of confidence, a blind call of intuition. That would represent a failure of philosophy to help with the justification of liberalism.

The question of how a conception of justice can stop being a blind call of intuition was the main subject of Chapter II. The crucial cornerstone of the whole Rawlsian enterprise, as I

argued at length, is his use of moral personality, for which he does not seem to present a clear and persuasive justification. I identified two strategies, both textually supported, that could provide some sort of argumentative cover for his specific and well-developed conception of person.

One is descriptive, trying to establish that his conception of person is embedded in liberal societies as we find them. The second is deliberative, trying to provide an independent argument supporting it. Ultimately, both are unpersuasive. The descriptive interpretation, in spite of being favoured and strengthened by contemporary political liberals, is unable to bear the weight of Rawls' rather specific use of moral personality. I could not find what the conception of person is supposed to describe as both the actual worldviews of citizens of liberal societies and the generic liberal ideal substantially differ from it.

The deliberative strategy does not fare much better. Grounding the conception of moral personality on a normative conception of social cooperation is an interesting move, but Rawls does not develop the idea further. As it stands, the normative idea of social cooperation can either slide back into the descriptive interpretation ('liberal societies understand cooperation thusly'), or it remains an unwarranted normative *fiat*, similar to the one used by the friends of self-ownership ('we should understand social cooperation in this way').

Overall, Rawls does provide us with a complex picture of moral personality and social cooperation that is needed to for a potentially successful justification of liberalism. However, he is less than clear on the sources of this picture, which casts doubt on the whole enterprise that is built upon it. If we read his theory as a justification of liberalism, it is certainly incomplete.

Having described such a bleak picture in the previous chapters, I will attempt to create a happy ending to this one. The subsequent text can be read in two ways. Either it can be viewed as an original argument, merging some Rawlsian insights with some distinctly un-Rawlsian, mostly continental material. Alternatively, it can be read as a broad and independent reinterpretation of Rawls, working with the promising features identified above and supplementing them on a few occasions with an original argument or reflection. Either way, what follows is my take on what I view as the most promising answer to the question of justification and defence of liberalism as a mode of functioning of political society.

This chapter should provide a justification of liberalism that fulfils the desiderata as they were presented throughout the text so far. It should justify a normative status of individual resulting in some conception of rights; it should explain how this normative status

translates into wider political relations; it should show the way for liberal societies to respect the pluralism within them; it should explain why liberalism is the best mode of existence for political societies across the world. This is obviously a very tall order. However, I hope that the critical and analytical work I have done in the previous chapters will make it easier.

In the next section, I try to establish what I take to be the core conception of person that we necessarily subscribe to every time we talk about justice. To this end, I employ two distinct argumentative strategies ultimately sharing the same aim. One, the ‘recursive justification’ can be described broadly as Kantian, while the other is hermeneutic and broadly Heideggerian in nature. With the core conception of person established, the third section is devoted to analysing the nature of human interaction. I argue that interactions between humans inherently contain a normative component; they are moralized from the start. This insight, carried to the political level, means that human social relations take the form of ‘the struggle for recognition’. The context of this struggle provides me with additional desiderata (beyond the conception of person established earlier) that need to be reflected in every acceptable conception of justice. In the fourth section, building on the previous ones, I argue that these established desiderata in the context of modern societies do not admit of a political form of society that would not be liberal or at least they make the situation of illiberal regimes precarious both normatively and in reality. That completes my argumentation.

2. Interpreting the question of justice

As analysed in the previous chapter, neither the descriptive nor the deliberative line of argument was very persuasive in justifying the liberal political regime. The descriptive strategy does not stand up to its billing as providing an uncontroversial basis assumed in liberal democratic societies while the deliberative argument ends up in a stipulated impasse. In this section, I offer a different strategy which I adopt and tailor for my purposes from the two philosophers that, at least in my reading, undertook to navigate in between the simple ‘deliberative’ or ‘descriptive’ options. Generally speaking, I take it that both Immanuel Kant and Martin Heidegger (1) wanted to avoid operating with stipulated metaphysical categories while (2) trying to articulate some shared experience of human beings that is deep, relevant, and quite general. None of the two wanted just to describe – but they were very keen not to work within an uncritically assumed normative or metaphysical framework.

However, my fragmented use of Kant and Heidegger should not be taken as a full attempt to interpret what these two thinkers actually wanted to say – far from it. My use of them is radically selective and purpose-specific, as I only analyse one particular methodological moment in each of their respective broad and deep philosophical conceptions. Moreover, I do not even use this methodological point in the way they intended, as I take it out of context of their work and apply it ‘my’ problem of understanding justice.

Generally speaking, my strategy here is to develop an argument that can justify the Rawlsian picture of moral personality while not falling into the pitfalls of deliberative and descriptive arguments analysed above. The recursive justification and the use of hermeneutic method enable me to uncover some of the structural features of our asking about justice which may remain hidden when we analyse justice conventionally. The results of my argument should thus provide the cornerstone for further analyses in the sections that follow.

2.1 Justice and Kantian recursive justification

It is almost a platitude to say that Kant is a well-known and popular figure in contemporary political thought – some even complain that the theoretical discussions in this area have become so kantianized that there is barely room for much else.¹⁶⁵ In spite of that, Kant is almost never taken on board without reservations. His transcendental idealism is universally shunned as supremely problematic and therefore the inspiration of contemporary thinkers with his thought is always somewhat loose and eclectic.¹⁶⁶ What is often deemed attractive are his accounts of moral autonomy, the priority of right over good, respect for common humanity, or his resolute rejection of paternalism and despotism. Methodologically, there has been an intense debate concerning the constructivist interpretation of categorical imperative and moral philosophy in general.

Here, I am not interested in any of the above. In what follows, I present what I take to be the most common form of Kant’s arguments, disregarding the content and sidestepping the constructivist questions. Interestingly, the argumentative strategy I use here is far from

¹⁶⁵ See for example William A. Galston, “Moral Personality and Liberal Theory: John Rawls’s ‘Dewey Lectures,’” *Political Theory* 10, no. 4 (1982): 492–519.

¹⁶⁶ It must be said that more often than not contemporary philosophers have a tendency to build distinctly un-Kantian premises into their Kant-inspired framework, often resulting in theories that are “Humean with Kantian face” (Sandel, *Liberalism and the Limits of Justice*, 13.) A clear example of what Sandel has in mind is the version of Kantianism by Korsgaard. See Christine M. Korsgaard, *The Sources of Normativity* (Cambridge University Press, 1996).

commonplace in contemporary literature, despite being so important. As a result, its application to the issues of political justice may yield interesting and unorthodox results.

I borrow the account of Kant's argumentative strategy from Katrin Flikschuh.¹⁶⁷ She calls it "a strategy of recursive justification":

"The basic strategy is to regress from a non-contentious, universally affirmed experiential premise to the contentious, non-experiential, subjectively necessary condition of its possibility and from there to a vindication of the objective validity of that presupposition. Recursive justification thus conceived involves the following three steps:

- 1) Affirmation of a non-contentious (generally accepted) experiential premise.
- 2) Analytic regress from that premise to the non-experiential subjective presupposition of its possibility.
- 3) A critical demonstration of that presupposition's objective validity in relation to the experiential premise."¹⁶⁸

Kant uses this form of reasoning in the most famous and celebrated arguments he develops. For example, his argument for pure categories in *Critique of Pure Reason*¹⁶⁹ starts from the premise that we experience outside objects. Kant then uses this premise to argue for the categories as the necessary condition of possibility of such an experience. In the final step, he demonstrates the categories as necessary for us to form a conception of outside objects. Similarly, the main argument in *Groundwork*¹⁷⁰ starts from our consciousness of the demands of moral law. Kant then argues for the idea of transcendental freedom as necessary presupposition for us being the subjects of moral law. In the third step, he attempts to demonstrate transcendental freedom as the ultimate ground of morality.¹⁷¹

The basic aim of this argumentative strategy is to make sense of our experience without superimposing some pre-existing metaphysical structure. Generally speaking, our experience, our behaviour, our perceptions are used as the starting point for further analysis, which is then carried forward by whatever implicit content is found as presupposed in this experience, may it be the categories of perception or an idea of unconditional moral duty. In

¹⁶⁷ Katrin Flikschuh, *Kant Contra Cosmopolitanism: Assessing the Global Justice Debate* (Cambridge: Cambridge University Press, forthcoming).

¹⁶⁸ *Ibid.*, chap. 2.

¹⁶⁹ Immanuel Kant, *Critique of Pure Reason*, ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1999), chap. II.

¹⁷⁰ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor and Jens Timmermann (Cambridge: Cambridge University Press, 2012).

¹⁷¹ *Ibid.*, chap. 3.

an ideal case, the presupposed content carries enough force not only to explain the phenomena, but also to provide a theoretical framework in which the phenomena are best captured and understood. As I said, in the first *Critique* and in *Groundwork* Kant aspires to do just that with regards to moral law and our perception of space and time. Whether he succeeds is, unsurprisingly, open to dispute.

Here, I am interested neither in moral law, nor in space and time. My question is how a strategy of recursive justification would work on the subject of political justice, disregarding Kant's actual views on the matter which are heavily comprehensive and difficult to maintain as strictly political. For that, I first need a non-contentious experiential premise. The experiential premise might be the simple: "I am concerned about political justice", "I care about political justice", or even "political justice matters to me". As an experiential premise, it is of course first personal – there is no presumption of any objective or intersubjective validity this point. It is only stating my consciousness of the issue of political justice coupled with its perceived importance for me. As a starting point, it is therefore relatively modest.

The second step is the place where this argumentative strategy truly unfolds. The question is, 'What are the non-experiential subjective presuppositions of possibility of me caring about political justice?' The adjectives are important here. The question is not about my personal history – my preoccupation with political justice may have developed from the family tradition of heated political discussions over Sunday lunch, but that is beside the point.¹⁷² The strategy wants to explore the conditions of the very possibility of the stated experiential premise. It therefore severs the connection with my own empirical existence. Roughly speaking, we search now for the largest common denominator for all the subjects that might share this type of experience with me.

I will use the term 'social reason' as a roof concept for this common denominator – as a technical term for the set of individual capacities that enable me to care about justice. The basic condition of possibility for my caring about political justice is thus me possessing a social reason. Without social reason, justice would never arise as an issue.

Firstly, a person possessing a social reason in the context of political justice must be able to cooperate with others. Only cooperating individuals can bring about the situation when people care about political justice. Presumably, an imaginary Rousseauian savage does not and cannot have these preoccupations. Only a sustained human collaboration can cause them.

¹⁷² Similarly, Kant's preoccupation with moral law may stem from his particular Pietist upbringing but that is not what he is concerned about in *Groundwork*. Hence, reading this book only as a picture of Kant's personal understanding of morality would completely miss the philosophical points he tries to make.

Secondly (and connected with the first point), a person possessing social reason must be able to create and obey rules and understand their point. These might not take the form of a contemporary sophisticated legal system, but some notion of rules, even if implicit, is necessary nonetheless. If I care about political justice, then this justice must take some form. It might regulate the issues themselves (in case of theft, you do X), or delegate powers (in case of theft, a person Y will be the judge), but some such form must be there. Justice as a constant disorderly flux is unimaginable. Therefore, as an agent that cares about justice I must be ready to think about possible rules for society and be able to submit to them myself and obey them.

Thirdly, if I really care about political justice, I must acknowledge that I am not the only person around. Others must be able to create, understand, and obey the rules as well. As subjects of rules and regulations, they must therefore have some legal and moral status I should respect. Not taking others as morally relevant equals not caring about political justice – if I take others fully instrumentally, the question ‘what is just’ becomes meaningless. Some level of recognition is therefore necessary. A further argument would be needed to prove that this recognition must in fact imply an *equal* status, but for the present purposes it is enough to state that the recognition of moral and legal status of my possible co-citizens is a condition of possibility of me caring about political justice.

Fourthly, if I care about political justice, there must be something motivating me, implying that I or someone else has something to lose. Otherwise there is no reason why I should be interested in the whole project. And this “something to lose” stipulates some notion of moral and practical interests or goals I have when cooperating with others. If I possess social reason, I not only act, I act for some reasons or to further my interests. In other words, if I care about something, this caring cannot appear out of nowhere, although at this point it is irrelevant whether this interest is moral or material, egoistic or altruistic, or some combination of all of the above. What matters is that I must be a being capable of formulating and pursuing a conception of good in cooperative situations.

The list of the four capacities above is by no means arbitrary. I take it that every single political philosopher must accept them – and every political philosopher did so in the past. The four necessary conditions of possibility of us caring about political justice that I labelled together as possessing social reason (the capacity to cooperate, to create and understand rules, to recognize others as politically and morally relevant, and to pursue so far unspecified interests) are also preconditions of meaningful political theory. They explain how the question of political justice arises and why it matters to us. If there are multiple agents that recognize

each other and pursue their interests together, there necessarily arises a possibility of a clash of interests. This sets up the very question of political justice, asking “what each is due” – which is a primordial definition of justice present already in Plato’s *Republic*.¹⁷³ If any one of these conditions is not fulfilled, the question of justice never arises. If the agents do not cooperate, are not able to follow rules, do not recognize each other, or do not have interests, political justice ceases to be an issue and asking “what each is due” loses its meaning. So an understanding of persons as possessing social reason in this sense is necessarily presupposed in us caring about political justice.

However, the argument is not complete before the last and the most problematic step – the critical demonstration of objective validity of this conception of person in relation to the original experiential premise. This step often proved to be tricky even for Kant himself. Indeed, when it comes to this third step in the *Groundwork*, “there is virtual unanimity that the attempt fails”¹⁷⁴ supported by the fact that Kant himself abandons the argument in his later work. The question, then, is how one can vindicate the above-established presuppositions about social reason. I think the best way to do it is to shift focus from personal interest in political justice and from objective presuppositions of political theory to the society at large: roll the argument back, so to say, and ask if these four capacities of social reason are not also objectively necessary for functioning of any human political society.

To give a quick answer, they are. No conceivable human society can possibly function without its subjects manifesting all four capacities listed above. Most obviously, some cooperation is needed to form and sustain society in the first place. Secondly, no organized form of human cooperation can function without acknowledged and at least partially observed rules. Even a minimal public order presupposes a formation of certain practices that in turn help to create some degree of mutual expectations that are regularly fulfilled. They need not be coercive, they need not be written down, but some such norms are necessary nonetheless. Therefore, societies must be able to form the norms and citizens must be able to understand and obey them. Otherwise there would be no moving on from the isolated Rousseauian savages.¹⁷⁵

Thirdly, if society is a system of cooperation between agents, then there needs to be some kind of recognition between them. If one does not acknowledge a moral status of any of his co-citizens and this sort of attitude is widespread (so it is not the case that this person is

¹⁷³ Plato, *The Republic*, trans. Desmond Lee (London: Penguin Classics, 2003), 331e.

¹⁷⁴ Henry E. Allison, *Kant’s Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 214.

¹⁷⁵ The fact that a human society consists in cooperation guided by norms is a commonplace sociological assumption. See for example Kathy Stolley, *The Basics of Sociology* (Westport, Conn: Greenwood, 2005), 43–47.

only an isolated lunatic), then it is extremely difficult to imagine how it may be possible for such society to function. We need to see our fellow others as capable of advancing claims that ought to be respected by us. Of course, it might not be necessary to view all the other humans as possessing this sort of status. History is full of societies that consisted of a small minority of full right-bearing citizens coexisting with large groups that were acknowledged only barely, if at all. Ancient Greek city-states are probably the best known examples. Still, what enabled even these states to function was the mutual recognition of the core of the citizens. Without it, these societies would not be possible.¹⁷⁶

Lastly, a society of beings without interests would not be a human society. We cannot imagine a human society as a sustained cooperation between creatures that furthers the goals that none of them grasps or identifies with. Such entity may be more appropriately described as a hive – and it has nothing to do with the human collaboration as we know it. Therefore, members of any society must manifest a purposeful behaviour informed by their interests and goals.

Again, I take the view presented above – that the four capacities of social reason are necessary for any human society to emerge – as unproblematic. To give just one illustration, all social contract tradition crucially rests on the citizens sharing these capacities. After all, not only a social contract but any possible contract is inconceivable if the two parties do not want to cooperate and do not pursue some good – and there is no point in it if the parties are unable to understand the obligation that follows from it and do not recognize each other as parties of an agreement. The absence of the first two conditions is almost a conceptual prerequisite of any cooperation, while the absence of the last two conditions would make it utterly impossible for people to escape the state of nature as Hobbes conceives it.¹⁷⁷ Generally speaking, human society as we understand it is unthinkable without us possessing social reason.

With this step, the recursive justification is complete. My caring about justice (the first step) was shown to presuppose four features that I jointly labelled as ‘social reason’ (the second step). Then, social reason understood in this way was shown to be necessary for any human society to exist and function (the third step). Therefore, social reason which manifests itself in my caring about political justice seems to be vindicated as a *sine qua non* of political society.

¹⁷⁶ I analyse the aspect of recognition in human societies later on, in subsection 3.3.

¹⁷⁷ Hobbes, *Leviathan*, chap. 13.

These results are important because they go some way towards filling the weak spots previously identified in Rawls's justification of liberalism. To remind, they were his struggle to justify his conception of person grounded in moral personality, and his seemingly *ad hoc* stipulation that human cooperation must not be merely a "socially coordinated activity" but must always include some notion of fairness.¹⁷⁸ Now I have means to at least partially address both of these problems.

The recursive justification has shown that social reason must always be presupposed when we discuss human political relations. As a result, Rawls' conception of moral personality is not merely arbitrarily stipulated – presupposing it is necessary within the very context of political justice itself. After all, if social reason is a necessary condition of possibility for the question of political justice to arise, then it makes sense to presuppose it in the context of this enquiry. And social reason as I used it in my argument is very close to Rawlsian moral personality. Dropping the conditions of cooperation and recognition for now, the remaining two features of social reason are a capacity to understand and abide by the rules – which comes very close to Rawlsian sense of justice – and, secondly, a capacity to formulate and pursue my interests as an agent – which is equivalent to Rawlsian capacity to adopt a conception of good.¹⁷⁹ It also shows human cooperation as normatively loaded from the start, not merely as "coordination".¹⁸⁰

Thus, the recursive justification applied to the question of justice comes close to vindicating Rawls' moral personality by showing that moral personality is necessarily presupposed in the very context of enquiry about justice. This shows that we can operate with a partially normatively loaded conception of person without presupposing some given metaphysical or comprehensive account of human beings and without trying to describe a specific cultural ideal. The strategy is to find out what needs to be presupposed once we accept the experiential premise of being interested in justice (still staying within the bounds of the political).

As to the second problem I identified in Rawls, if we presuppose human cooperation as an activity conducted by beings possessing social reason, it is much easier to claim that when these beings work together, it is not a "mere coordination", but a social cooperation as Rawls understands it. Social reasoners as I defined them have a capacity to create rules and act by them in cooperative settings. As a result, the questions of fairness and justice always

¹⁷⁸ See Chapter II section 4.

¹⁷⁹ Rawls, *Political Liberalism*, 81.

¹⁸⁰ *Ibid.*, 16.

emerge in these situations, simply in virtue of the types of beings that participate on them – they create rules and are normatively expected to follow them. Therefore, Rawls is justified in stressing the necessary normative undertones of fairness in human cooperation generally understood.

In this subsection, I followed the Kantian recursive justification strategy applied to the question of political justice. I started with an experiential premise of me caring about it, and then I spelled out the very conditions of possibility of such experience. I ended up with a conception of person possessing a social reason that I hold is objectively necessary in any context of human society. If this can be vindicated, it will be a good start, as it provides two pieces of puzzle that were manifestly missing in Rawls: a justification of his conception of moral personality and a tool to distinguish between human cooperation and a merely socially coordinated activity. However, before I continue with the hermeneutic side of this argument, I have to strengthen this one with a few additional considerations – some parts of the reasoning I presented invite swift and powerful objections.

2.2 Interlude: Developmental psychology and political justice

There is a possibly very persuasive response that can be mounted against the recursive justification argument as I presented it above. It can be said that I smuggle normative considerations into human cooperation as awkwardly as Rawls, not considering the alternatives. If I am a western-raised political philosopher, of course I care about political justice. After all, it is in my job description. However, this fact is not on par with perceiving external objects as spatially and temporarily extended – political justice may be just a particular obsession connected with my profession and it is certainly not “non-contentious and universally affirmed”. Going back to Hobbes, it might be claimed that the only thing people want from politics is survival and furthering their own interests, with no normative considerations present at the outset. Non-self-interested normative considerations are only an internalization of the rules created by the sovereign to maintain order and keep its power. An experience of caring about justice and fairness is highly specific and derivative of a certain type of character living under a stable liberal regime. People cooperate and their interests are real, yet it is a struggle for power, not struggle for justice, which takes precedence.

As a result, the four features of social reason I introduced can be read completely without the ‘normative bias’. We can cooperate, create and enforce rules, acknowledge the

role of others, and pursue our interests without any sort of morality entering into it. We do it for self-regarding reasons, to further our interests and inclinations. We submit to rules out of fear or out of desire to belong somewhere and reap the benefits.¹⁸¹ Indeed, it can be claimed that the four characteristics of social reason I introduced are transparently manifested not only in humans, but in animals as well. A pack of wolves certainly cooperates and it does have its hierarchies, so the members know their place and submit to the ‘rules’ of the group. Moreover, the different members are recognized according to their status and the fact that they belong to the pack in the first place is an expression of their desire to survive and procreate – living in a pack does further their interests. Yet Rawls would never claim that wolves possess moral personality. They merely coordinate their actions when they hunt – a true cooperation with the subsequent demands of fairness and justice does not take place.

All this means that the recursive justification seems to do nothing to show the necessity of moral personality for human collaboration – and we still do not know why this collaboration should be as normatively loaded as Rawls wants us to view it. The Rawlsian weak points remain weak. As a result, we still do not have a conception of person and social cooperation needed to advance a justification for liberalism.

However, the first-person experiential premise of “I care about political justice” and a generally normative character of human cooperation may prove to be more resilient than this counterargument allows for. Michael Tomasello, an American developmental psychologist, presents a well-developed case for a normative nature of human cooperation and us caring about justice. In his book *Why We Cooperate*¹⁸², he put together an impressive amount of research on very young (often pre-verbal) human children and compared their behaviour to our closest primate relatives. The general aim was to put a finger on possible differences between human beings and animals when it comes to our cooperation and interaction and then attempt to explain these differences.

Tomasello’s experiments usually involve a certain cooperative task and a subsequent sharing of spoils. The research showed several remarkable results. First of all, the normative understanding of cooperative activities (‘how something is to be done’) is almost instantaneous in even the youngest human infants. This is supported by the fact that these children participate in enforcing the rules of a given game almost immediately after they

¹⁸¹ For a contemporary exposition of this view see David Gauthier, *Morals By Agreement* (Oxford: Oxford University Press, 1986).

¹⁸² Michael Tomasello, *Why We Cooperate* (Cambridge, Mass: The MIT Press, 2009).

grasped them and began to play.¹⁸³ Such behaviour is simply not found amongst our closest primate relatives, not to mention other animals. Secondly, while chimps tend to look at maximizing their own individual interests even in more complex cooperative activities, often fighting over the spoils, humans are inclined to agree on sharing the spoils to a much greater extent. The distributive pattern is very often egalitarian or proportionate to the amount of ‘work’ put into accomplishing the cooperative task, with participating kids often surprisingly declining the reward if it was deemed as too low – an implied protest against unfairness.¹⁸⁴ Again, this is something that other animals simply do not do. This impressive ability to cooperate with a sense of fair terms and reciprocity (to use the Rawlsian vocabulary¹⁸⁵) is all the more interesting as the children in the studies are so young that they could not conceivably have learned it from their parents – most of the kids in Tomasello’s studies were preverbal. The drive for this sort of cooperative behaviour “comes naturally.”¹⁸⁶

This innate inclination and ability to cooperate is importantly helped by the fact that human beings from a very young age understand cooperation not only from their own perspective but from bird’s eye view, so to say.¹⁸⁷ Tomasello designed a very simple experiment to prove it. It consisted in a cooperative activity where two participants worked together but had to perform different tasks. He took chimps and very young children and first taught them how to perform one set of tasks. Then they played the game with their partner for some time, until they understood it completely. When he switched the tasks, he discovered that chimps took approximately the same time to learn the second task, while the children learned it considerably faster. Tomasello claims that it is a result of our much better capability for understanding role-playing. When children play a role in any given activity, they have a holistic grasp of the whole activity, as they understand that the second person is also playing a role, discerning the role played from the person playing it. Even the youngest children thus understand and recognize the person they cooperate with and her goals and motivations. As a result, we can thus observe a much better ability to get in the second person’s shoes and, consequently, cooperate with the mutual recognition of each-other’s needs.¹⁸⁸

¹⁸³ Ibid., 36–39.

¹⁸⁴ Ibid., 33.

¹⁸⁵ Rawls, *Political Liberalism*, 16.

¹⁸⁶ Tomasello, *Why We Cooperate*, 4.

¹⁸⁷ Ibid., 40–42. Tomasello often uses Nagel’s metaphor of a ‘view from nowhere’ to explain the type of understanding we have of our cooperation. Moreover, he sees his research as furthering the philosophical points developed Nagel in Nagel, *The View from Nowhere*; Thomas Nagel, *The Possibility of Altruism* (Princeton, NJ: Princeton University Press, 1979).

¹⁸⁸ Tomasello, *Why We Cooperate*, 60–68.

All this has enormous implications for our understanding of human interactions. When humans cooperate, we do not merely perceive our own interests. We also understand the roles and motivations of the other person. All this enables us to share goals, have much more complex mutual expectations and even create obligations.¹⁸⁹ As a result, when spoils are being shared we perceive not only what we want but also what the other wants from our shared activity – an activity we have undertaken together in order to get a reward for both of us, where a certain pattern of sharing is already (implicitly or explicitly) deemed as just or fair. That completely changes the cooperative landscape. The perspective of fairness and justice (of “what do we owe to each other”) seems to be an inescapable component of human cooperation and even the youngest infants perceive it.¹⁹⁰

Here, I am not interested in metaethical, evolutionary, or biological aspects of this story. I only want to show how humans perceive their cooperation, grounding my first-personal experiential premise from the recursive justification argument. As it turns out, the research by Tomasello suggests that the self-understanding of our social selves necessarily involves considerations of fairness. When we cooperate, we cannot escape the Plato’s question of giving each his/her due. Obviously, no hard questions about the content of justice are settled at this point. After all, most human cooperation these days is more complex than two people shaking a tree to get a coconut. But it does mean that justice and fairness are implicitly present every time we work together and we cannot ignore it.¹⁹¹

Given the necessarily normative nature of social cooperation, the original experiential premise “I care about political justice” may just prove to be non-contentious and universally affirmed, at least in practical human behaviour. When people cooperate, some treatment and distributive patterns will always be considered unfair and unjust, while others are fair and acceptable. Human cooperation is never only about a maximization of preference-satisfaction. The role of the other is always acknowledged and rules are more than mere arbitrary instructions – from the earliest age we naturally perceive them as something that ought to be

¹⁸⁹ Ibid., 57–58.. For an analysis of human shared intentionality, see Margaret Gilbert, *On Social Facts* (Princeton University Press, 1992).

¹⁹⁰ See Scanlon, *What We Owe to Each Other*. Indeed, over the course of the book Scanlon makes a persuasive philosophical case for the same conclusions that are supported by Tomasello: that human cooperation has a form of an implicit normative contract manifested by certain mutual duties and obligations.

¹⁹¹ Arguing for this point, Plato’s Socrates seems to claim that some form of justice is necessary for any cooperative enterprise – even an inherently unjust one: “We don't speak the complete truth about those men who we say vigorously accomplished some common object with one another although they were unjust; they could never have restrained themselves with one another if they were completely unjust, but it is plain that there was a certain justice in them which caused them at least not to do injustice to one another at the same time that they were seeking to do it to others.” Plato, *The Republic*, 352b–c.

respected. Hence, normativity is deeply embedded in our cooperation – and therefore we cannot but care about justice in society.

The deeply embedded moral character of human cooperation goes against the picture of human social relations presented by David Gauthier.¹⁹² Prisoner's dilemma or any other interest-maximizing scenario is not a good model of human cooperation simply because humans do not and maybe even cannot normally perceive their interactions in this way. Cooperating with someone, recognizing someone, creating and following rules, all contain the normative element presupposed by the social reason as I developed it. As a result, wolves are no longer an apt example and Rawlsian moral personality necessarily results in more than a merely socially coordinated activity.

Since we cannot but cooperate as moral persons, moral personality is justified as an assumption – as I claimed above, it is internal to the enquiry about justice. Having a moral personality is the reason why we strive for rules in the first place. Therefore, the strategy of recursive justification supported by the considerations about the nature of human cooperation seems to have strengthened the weak spots in the Rawlsian thinking I identified above.

Building on the insights just presented, the fact that we are beings that cannot but view their cooperation normatively immensely changes the requirements of justification and legitimacy for our societies. Human societies face the task of being set up so as to express and satisfy our yearning for justice resulting from our shared moral personality. To illustrate this point, there is no right or wrong way to organize a society of robots. Some ways are more efficient than others with respect to some possible goals one might have, but that is all that can be said. However, the situation is completely different when it comes to human beings. The considerations of justice always creep in – simply in virtue of the nature of beings that take part in those societies and the nature of their cooperation. Ruling *merely* with force is not an option for human societies. Even the most brutal and barbaric military junta always provides some sort of justification for why they need to be in power and why the brutal things they do are necessary or at least somehow warranted (and therefore, in the final analysis, just). Moreover, every government, even the most inept one, always presents itself as bringing some sort of good to the people, furthering their interests. Admitting the opposite is politically unthinkable and amounts to an open call for revolution or abdication. The moral personality of citizens therefore presents the ultimate framework in which every political regime must

¹⁹² Gauthier, *Morals By Agreement*.

operate. The right and the good for citizens and political society are the horizon of our political thinking.

In any case, the recursive justification argument about political justice seems to be well-supported by the contemporary scientific evidence and it points to a conception of politics that is deeply responsive to the normative questions of human cooperation. Now it is time to look at Heidegger, whom I use to convey a similar point – but from a somewhat different perspective. I then connect the dots at the end of this section.

2.3 Justice as a mode of our social self-understanding

The argument I developed in the preceding two sections is crucial for the progress of this thesis. It establishes the normative conceptions of person and social relations that move my search for liberal justification an important step forward. In this section, I develop a similar argument using a broadly hermeneutic methodology inspired by Heidegger. This should ensure that the Rawlsian assertion of moral personality is really sound in its role as the basic element of political theorising.

Before starting, let me stress again one important caveat: the content of Heidegger's philosophy falls outside the scope of this thesis and I cannot delve more deeply into it. I only take his method of enquiry when formulating the quintessentially Heideggerian question about the nature of Being (which is in itself a very limited part of his methodology) and reformulate it as a non-Heideggerian question about justice. This mode of asking about justice can then reveal some considerations that are lost or not recognized in the descriptive or deliberative approaches towards the justification of liberalism. In what follows, I start with a short overview of Heidegger's argumentative strategy from the first six sections of *Being and Time*,¹⁹³ and then proceed with the argument itself.

Heidegger starts his *opus magnum* by stating his utter dissatisfaction with the way in which philosophers seem to have forgotten about the philosophical importance of the question of Being. After the first attempts in Plato and Aristotle, philosophers had a tendency to designate Being as the most general and therefore the emptiest of concepts that in virtue of this fact resists any possible definition – which was reason enough not to pursue the issue further. For Heidegger, that is a mistake. As some understanding of Being is presupposed in

¹⁹³ Martin Heidegger, *Being and Time* (Oxford: Wiley-Blackwell, 1978).

everything we do, in order to do philosophy (or science) properly we need to make our understanding of Being explicit and ask the question of Being once again.

However, we not only do not know the answer to the question of Being, we do not even know the question. Therefore, as there are multiple types of beings, we must first find out “in which entities is the meaning of Being to be discerned? ... to take its departure? Does some particular entity have priority when it comes to work out the question of Being?”¹⁹⁴ Only once we have found out which type of being we need to concentrate on it is possible to start the enquiry into the nature of Being.

And here comes the crucial turn that determines the drive of the entire book that follows. Heidegger writes: “Looking at something, understanding and conceiving it, choosing access to it – all these ways of behaving are constitutive for our inquiry, and therefore are modes of Being for those particular entities which we, the inquirers, are ourselves. Thus, to work out the question of Being adequately we must make an entity – the inquirer – transparent in his own Being.”¹⁹⁵ In tackling the question of Being, it is the Being of the entity that asks the question that gains the utmost importance. And Heidegger calls this entity *Dasein*. As he specifies it, “*Dasein* is an entity which does not just occur among other entities. Rather, it is ontically distinguished by the fact that, in its very Being, Being is an issue for it. ... It is peculiar to this entity that with and through its Being, this Being is disclosed to it. Understanding of Being is itself a definite characteristic of *Dasein*’s Being.”¹⁹⁶

At the first glance, these passages may seem unclear or even unintelligible, but the thrust of the argument is clear enough. Heidegger is interested in the nature of Being. However, the issue is complicated and it is not entirely clear where he should start. Nevertheless, from his analysis it turns out that an understanding of Being is presupposed in the very enquiry about its nature. And at the very heart of this pre-theoretical understanding stands the enquirer – someone for whom ‘in his very Being, Being is an issue for him.’ From then on, the Being of the enquirer – of *Dasein* – becomes the primary concern of Heidegger. That is where the first part of the *Being and Time* ends – with subsequent analyses being far beyond the present topic.¹⁹⁷

As I claimed above, emulating Heidegger’s strategy may prove to be very fruitful when it comes to understanding the basic shape of the question of justice, especially in

¹⁹⁴ *Ibid.*, 26.

¹⁹⁵ *Ibid.*, 26–27.

¹⁹⁶ *Ibid.*, 32.

¹⁹⁷ For a very instructive source on Heidegger’s philosophical concerns and motivation, see Stephen Mulhall, *The Routledge Guidebook to Heidegger’s Being and Time* (London: Routledge, 2013).

connection with an ambition to justify a liberal political regime. Justice, like Being, is a very broad concept with an unsure scope and meaning – even though it is not the most general and the emptiest of concepts. Still, how to approach it is an open question with no clear-cut answer. We know that there can be just institutions, just actions, just people. But when trying to capture the essential features of justice, we often do not know where to start. We think about justice, analyse it, but it still eludes us.

If we step on the same path Heidegger adopted, one crucial aspect of our deliberating about justice becomes immediately apparent: When trying to figure it out, we already have some sort of understanding of what justice is. We see various actions as fair or unfair, experience moral outrage or call of duty. Nevertheless, our pre-theoretical understanding of it is of course not good enough to support a fully developed conception of justice by itself – such conception is not simply given to us, we do not directly perceive it. Yet the fact that we all *sort of* know what justice is about is of crucial importance. We do not construct a conception of it out of nowhere.

Transposing Heidegger's answer to the question of justice, the paraphrase of the crucial bit could go as follows: 'Understanding something as just, judging, praising, acting virtuously – all these ways of behaving are constitutive for our inquiry, and are modes of being for those particular entities which we, the inquirers, are ourselves. To work out the question of justice adequately, we must therefore make an entity – the inquirer – transparent in his own yearning for justice.' Hence, 'Human being is an entity which does not just occur among other entities. Rather, it is ontically distinguished by the fact that, in its very being, justice is an issue for it. ... The self-understanding of humans as beings who care about justice is a definite characteristic of their life.'¹⁹⁸

To break it down without paraphrasing Heidegger, the fact that human beings deeply care about justice should become fundamental in our enquiry about the political conception of justice and its nature in general. Only once we take this fact into consideration and give it a proper place within the theory we will be able to correctly conceptualize justice. And with that, hopefully, its role in political society will become much clearer. To put it differently, it is

¹⁹⁸ Paraphrase of Heidegger, *Being and Time*, 26,32. The problem of these paraphrases lies in a fact that when Heidegger talks about the Being of a being, there is a crucial resonance between those terms. Our Being is presupposed in everything we are. However, our status as beings capable of justice does not go that deep into our Being, so this resonance is not quite achieved. But still, the same turn of argument is possible on a less 'ontically fundamental' level. 'Being capable of justice' is a core and deep-reaching mode of human life. What is more, understanding and interpreting this capacity is imperative when trying to construct a theory of justice. Therefore, I take this Heideggerian strategy to be relevant in the context of justice, even if Heidegger himself might not have endorsed it.

clear that justice constitutes a part of our human life. Raising the questions of justice is a mode of human conduct that is important for us and we cannot completely escape it. If we want to come up with a specific conception of what justice is and how it works, we must regard our attitude towards it as profoundly relevant. Indeed, as with Heidegger's Being, the analysis of our instinctive attitude towards justice may provide us with a key to unlock the subject as a whole.

Therefore, in the first step of an analysis of our relationship with justice we have to ask a few fundamental questions: Why is justice important? Why do we care about justice at all? Or, even more specifically, why do we care about justice in politics? What I want to do here is to (hermeneutically) try to make transparent our own desire for justice. That would uncover the motivational structures, which lead us to pose the question of justice in the first place.

There are two answers to the questions above. Firstly, justice is a natural and necessary response to the rise of the sphere of intersubjective interactions of human beings. When a group of agents tries to live together in an organized society, justice becomes an issue. There is always more than one possible distributive arrangement for the material and, importantly, moral goods (money, status, recognition, etc.), so some decisions need to be made – and we can understand and discuss the different arrangements as being more (or less) just and fair. The preservation of property as described by Locke¹⁹⁹, the fundamental self-interest by Hobbes and Gauthier described above, these are all conceptions motivated in a large part (in case of Locke) or wholly (Hobbes and Gauthier) by our preoccupation with justice as a device of intersubjective interest-satisfaction.

However, the interest-satisfaction in itself does not provide the full answer to the questions above, it does not fully disclose our preoccupation with justice. We also care about justice, about right and wrong, because *that is who we are*. This point may seem to be hopelessly circular – ‘we care about justice because we do’ – but that is not completely so. Justice is important for us because we are beings capable of perceiving it and also acting according to it. Justice therefore presents us, at least from the first-person perspective, with an option of behaving in a certain way. And a possibility of this kind of behaviour is an important part of human life, an important part of how we understand ourselves.

The results of the considerations above are valuable because they give us a necessary context of any inquiry about justice – a conception of person that we inevitably operate with.

¹⁹⁹ Locke, *Two Treatises of Government*, 1988, § 138.

This conception consists in the two central features of human beings that motivate their preoccupation with (political) justice. Firstly, it is the crucial fact that we as agents find ourselves in situations where we depend on other agents (and vice versa) for resources, recognition, and simple help – and our interests may cross. Every conception of justice must therefore take into account the intersubjective challenges that human agency necessarily faces in coordinating diverse and often clashing interests. Otherwise it would just disrespect the very conditions from which any enquiry about justice emerges in the first place. However, we also possess (to use a Rawlsian phrase) a sense of justice. Some institutional setups can be just, while some others might not be – and we are capable of spotting the difference and it matters to us in a way that is irreducible to simple pursuit of interests. As Tomasello's toddlers, we are driven to refuse an unfairly low reward as an affront to our sense of justice even if our interests are to suffer.

This twofold conception of person comes, again, relatively close to the one adopted by Rawls. Using the Heideggerian method, I shifted the focus from the abstract questions of justice to the person that is asking the question. That revealed an underlying conception of person that needs to be present in every enquiry about justice. One component in this conception is a capacity to act, form, and follow interests while the other is a sense of justice. This distinguishes humans from robots on one hand and animals on the other, and is a condition *sine qua non* for establishing any sort of moral community. In any case, the Rawlsian moral personality, consisting in a sense of justice and a capacity to adopt a rational plan of life, seems to be quite well established once the Heideggerian preoccupation with the nature of being interested in justice is put into place.

Restating the whole strategy with a bit more hindsight, what I proposed in this subsection was carried by a broadly hermeneutical method adopted from Heidegger. The starting point of an enquiry is justified when it is shown to be inherent to the enquiry itself. The initial problem was that the question of justice does not provide us with some easy and self-evident point of entry. There are multiple possible subjects of justice, conflicting intuitions, plurality of possible strategies. But the central stage amongst these difficulties is always taken by the human being whose yearning for justice is the true source of all the questions and problems. Therefore, we must direct the first focus of our theoretical endeavour to this very human being, disclosed in its desire for justice. Such analysis will unearth the basic impulses for coming up with a theory of justice, answering the fundamental questions of motivation: what do we want, why, and what are the limitations of it.

The argument I developed here does two important things. First, it justifies the broadly Rawlsian approach to justice, starting with the conception of person as having a moral personality. Secondly, it fits well the desiderata that I am working with throughout this text. The content of the conception of person acquired by this method is not descriptively taken as something we happen to believe in. Nor it is just stipulated as normatively relevant for our political societies. There is no comprehensive metaphysical story behind it either. It thus can sail through the problems that marred the descriptive and deliberative attempts to justify (political) liberalism.

Putting the two lines of the argument together, what I tried to do in this section is to present a picture of human beings that is implied in our preoccupation with justice. The two methods applied were inspired by two seemingly very different philosophers, but they involved a very similar move. Both refused to start with some metaphysical postulates, yet both still aimed at universal validity. What was different from the approaches described in previous chapters was the primary concern for justice not as an abstract cultural practice or a transcendental set of rules, but as a subject experienced and grasped by human beings. Both the preoccupation with the conditions of possibility of my caring about justice and the focus on human understanding of justice as a mode of being proved to be fruitful in setting the most fundamental context of any enquiry about justice.

In a sense, both strategies represent a shift of perspective, a small ‘Copernican revolution’. Instead of trying to figure out what is just or what is the best political regime directly, I shifted the focus on the entity that is asking the question. The hope is that this shift of focus can help in answering the original questions. There are few important inroads already made. First of all, the strategy did provide some basic content and context of an enquiry about the best political regime. The human cooperation is understood as normative from the start and guided by our innate sense of justice. Moreover, it was established that the question of political justice can be answered only by taking into consideration and working with human beings understood as sharing a social reason. All these are important qualifications taking the enquiry forward.

Moreover, shifting the focus helped sidestep the problems that proved to be fatal for the attempts to justify liberalism I analysed above. Asking what conception of person is implied in our quest for justice is neither a description, nor a stipulation. The problem of a descriptive interpretation of Rawls and of value pluralism was the inability to specify what was being described and what are the normative consequences of this description. The

metaethical reality, social ideals, or political opinions of citizenry were unsure targets unable to securely ground the normative commitments that supposedly followed. As a result, the whole edifice of liberal justification that was intended to rest upon them proved to be quite doubtful. The arguments above do not exhibit this flaw as they do not describe and do not take a conception of person from somewhere – they only ask what must be true if persons care about justice (which we do, as I showed in the section on Tomasello’s developmental psychology).

The arguments I presented also do well when compared to the ‘deliberative’ strand of liberal justification. This strand proved to be problematic because it always ultimately rested on some sort of normative *fiat* which was assumed but not explained or defended. Self-ownership asked us to adopt a picture of property relations dominating the normative landscape, while Rawls interpreted deliberatively never moved beyond stipulating that moral personality and freedom and equality should be the cornerstones of our enquiry into the principles of justice. In contrast, the Kantian and Heideggerian strategies go beyond a mere stipulation – again, they ask what follows from the human understanding of and preoccupation with justice.

However, the task of justifying liberalism is far from done with setting up the context of the enquiry and the basic conception of person to go with it. The content I presented may still be prescriptively barren and not tell us much about what sort of political regime is right or can be justified for human political societies. That is why I want to build on these insights and advance into more complex questions of human self-understanding in the next section, where my strategy to justify liberalism starts to take a more definite shape.

3. Seeing ourselves as normatively relevant political beings

Establishing moral personality as necessarily implied in any enquiry about justice is a crucial step in my pursuit to philosophically justify liberalism. However, there are few more steps needed. As I analysed in the last chapter, the passage between moral personality and freedom and equality is less than clear in Rawls.²⁰⁰ Even though his conception of moral personality undeniably supports the further steps he takes, moral personality in itself is not sufficient to get to freedom and equality, and therefore to fully establish a liberal regime as the best alternative to govern a political society.

²⁰⁰ See Chapter II, section 3 and TJ §77.

In the last chapter, I commented on Ian Carter's interesting contribution as he attempted to tie moral personality with freedom and equality using respect.²⁰¹ However, Carter's solution cannot help answer my question. He is using respect as a core liberal value that supplements the Rawlsian moral personality and thus vindicates other core liberal values, freedom and equality. But the fact that liberal values mutually reinforce and support each other does not by itself answer the question of liberal justification. Carter's strategy is extremely interesting, but he does ask us to accept respect and democracy as the postulated starting point, a normative *fiat* not unlike the ones I analysed in the first two chapters. I suspect that if Carter was to elaborate his position further, it would fall into similar problems as the deliberative interpretation of Rawlsian justice as fairness. Therefore, I cannot rely on his analysis when pushing my argument forward – I need another path to get from moral personality to freedom and equality.

This section elaborates on moral, social, and political implications of the conception of person as it was presented in the last chapter. The intermediate goal is to provide an additional content behind an important and insightful yet a bit underdeveloped Rawlsian assertion of human persons being “self-authenticating sources of valid moral claims.”²⁰² In the first step, I discuss the consequences of us and others sharing a moral personality in normal everyday interactions. The inescapable conclusion here is that people and their actions morally matter to us. I adopt the famous Peter Strawson's article *Freedom and Resentment* as a guide to this issue. Later, I move to the subject of recognition where the political aspects of us possessing a moral personality can be seen the clearest. It is indeed the issue of recognition that unlocks what I understand to be the critically important part of liberal justification.

3.1 Resentment and seeing the other as morally relevant

When discussing human cooperation, Rawls jumps directly from “the idea of social cooperation” to fairness, reciprocity, and justice between the free and equal.²⁰³ I take it that the move is too fast and a few important insights can be gained laying out the possible intermediate steps. After establishing moral personality in the previous section, I therefore want to explore the space in between moral personality and fairness, something that Rawls does not do and that is crucial for the justification of liberalism I present here.

²⁰¹ Carter, “Respect and the Basis of Equality.”

²⁰² Rawls, *Political Liberalism*, 32.

²⁰³ *Ibid.*, 16.

What happens and what does it look like when the agents with moral personality cooperate? That is the question I am interested in now and that is also the subject of Peter Strawson's famous essay *Freedom and Resentment*²⁰⁴ as I read it. One of the basic phenomena we observe when agents with moral personality interact is a possibility of them adopting 'reactive attitudes' towards one another like resentment, indignation, blame, or gratitude. This is where the underlying normative nature of human cooperation best manifests itself.

If we want to understand the limits of normative nature of human cooperation, we have to ask under what conditions do we stop having reactive attitudes. One prominent way to achieve this is to call into question the very agency of the subject involved. 'He is only a child' or 'he is a schizophrenic' are examples of perfectly valid answers. We are thus invited to suspend our ordinary reactive attitudes towards the person involved, changing the way in which we see him because the situation is somehow different from our normal intersubjective interactions.²⁰⁵

When we stop seeing someone as a normal partner for discussion or interaction and approach him for example as mentally deranged, "all our attitudes tend to be profoundly modified."²⁰⁶ Applying a simplifying dichotomy to a complicated issue, Strawson contrasts "the attitude (or range of attitudes) of involvement or participation in a human relationship, on one hand, and what might be called the objective attitude. ... To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject of what, in wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained."²⁰⁷ We always adopt this attitude towards severely mentally disabled and often towards children. We see them as entities to be managed, not as authors of their own projects. Moreover, we can adopt this attitude to fully-grown human beings as well. We may analyse their situation dispassionately, try to think what could work in motivating them, or how to discourage them from certain behaviour. In short, objective attitude rejects the normative element of human cooperation (manifested by reactive attitudes) by denying or at least disregarding the moral personality of the subjects involved.

²⁰⁴ Peter F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 1–25.

²⁰⁵ *Ibid.*, 8.

²⁰⁶ *Ibid.*

²⁰⁷ *Ibid.*, 9.

Yet, adopting a dispassionate objective attitude at all times towards everyone is “practically inconceivable.”²⁰⁸ Even the coldest CEO cannot constantly see her colleagues as simple ‘human resources’. We are committed to personal human relationships, to seeing others as partners and subjects on their own right, not as simple obstacles to be managed and overcome. It is an important part of our human world that it contains people with whom we can truly interact and not just use/train/handle them. Our personal involvement with other people is an indispensable part of our living world and the absence of it is impossible for us to imagine. At the end of the day, even though we are capable of both objective and personal attitude towards others, it is the personal attitude that is unavoidable in human interaction and it constitutes the baseline of how we approach each other.

A personal involvement is a necessary precondition for experiencing reactive attitudes. Not only the negative ones in form of resentment and indignation, but also gratitude, shame, forgiveness or “the sort of love which two adults can sometimes be said to feel reciprocally, for each other.”²⁰⁹ When one adopts objective attitude, all these are unthinkable – we need to truly engage with others in order to experience them.

When agents with moral personality cooperate, the moral stance towards others (including the attributions of responsibility, which is Strawson’s main concern) is taken for granted. Strawson claims that “the existence of the general framework of (moral – my insertion) attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external ‘rational’ justification.”²¹⁰ What matters here are the *experiential facts of humans interacting with each other*, necessarily raising normative concerns embedded in their reactive attitudes. The philosophers therefore ought not to strive for a complete detachment of these human attitudes to reach an ideal of third-personal objectivity. If they do, something important slips out of their hands and they are no longer capable of correctly viewing the important facts of our common humanity.²¹¹

When it comes to the political consequences of this picture, it establishes several crucial points that I want to use for my purposes. The first is a claim that we cannot stop being personal in our intersubjective relations, which therefore become and remain deeply normative. We do not rationally choose such mode of interaction; it reflects “the fact of our natural commitment to ordinary inter-personal attitudes. This commitment is part of the

²⁰⁸ Ibid., 11.

²⁰⁹ Ibid., 9.

²¹⁰ Ibid., 23.

²¹¹ The issue is raised and thoroughly explored in Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, Mass.: Harvard University Press, 2009).

general framework of human life, not something that can come up for review.”²¹² Personal as opposed to objective attitude to others is thus deeply engrained in the world we experience and live in. If the previous section established that we are justice-seeking moral agents, this one showed that no moral agent can live on its own normative island – the moral status of the other is necessarily acknowledged in our mutual interactions.²¹³

Our relations towards others are necessarily moralized from the start. “The personal reactive attitudes rest on, and reflect, an expectation of, and a demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves; or at least the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard.”²¹⁴ Resentment, indignation, blame, praise, shame, gratitude and other reactive attitudes all contain deep normative structures inherently present in them. In feeling them, we take others as normatively relevant and want, in turn, to be acknowledged as normatively relevant ourselves. These normative claims do not need an independent rational argumentative support; they are supported phenomenologically, so to say – as natural parts of the experience of human interaction. They are a necessary part of life for every human being.

Moving now outside the Strawsonian vocabulary, we can claim that our attitudes towards others reveal that we always see ourselves and others as Rawlsian holders of “valid moral claims”²¹⁵. We raise and accept normative claims, recognizing others as beings that raise and accept these claims as well. This is based on the ‘practical inconceivability’ of regarding others objectively, as mere hindrances or resources helping or obstructing our interests. What is established here is thus an interesting necessary feature of human self-understanding. We understand ourselves and others as normative beings right from the start, in every interaction with our fellow human beings. We raise moral demands upon others and we struggle for moral recognition from others. And we cannot do otherwise, as both of these actions are a necessary part of the experience of human interaction.

The argument just presented brings my search for liberal justification one step further. It builds on the Copernican fact of me caring about justice – moral personality is presupposed in the enquiry about justice and we necessarily understand human cooperation as having an intrinsic drive towards fairness. Furthermore, as I have shown now, any conception of justice needs to take into consideration a normative relevance of human beings and their claims. In

²¹² Strawson, “Freedom and Resentment,” 13.

²¹³ For a development of this point, see Emmanuel Levinas, *Totality and Infinity: An Essay on Exteriority* (Pittsburgh: Duquesne University Press, 1969).

²¹⁴ Strawson, “Freedom and Resentment,” 14.

²¹⁵ Rawls, *Political Liberalism*, 32.

human interactions, political or not, we cannot but attribute moral relevance to other people's actions – we praise and blame them, attribute responsibility, are grateful or outraged. If someone claims a right towards something, he does not merely inform us of his interest, he makes a second-personal normative plea that we may disagree with but cannot just shrug off. Politically, the basic reaction to these findings is an attribution of respect and recognition – topics that Strawson does not explore directly but that cannot be avoided in my quest for liberal justification.

3.2 Recognition as a political acknowledgement of moral personality

My Strawson-inspired analysis of resentment showed that interacting human beings possessing moral personality view others and themselves as normatively relevant, as subjects of moral reactive attitudes, as moral actors that can be praised and judged in virtue of the moral status they possess. But (and this is a crucial assertion for my argument) if people view themselves as holders of valid moral claims, then this fact needs to be mirrored and appreciated by the political regime under which they live. And this appreciation of moral standing of persons is what I understand under the label recognition.

To clarify, I use the term recognition in its most basic political sense – recognition happens when the government and political society in general acknowledges its members as agents with moral personality and, as a result, as holders of certain valid moral claims – as entities that are not to be governed solely in rational-utilitarian (objective) way because they possess a normative status over and above that.²¹⁶ As I am focused on finding a justification of a basic liberal form of government, I am not interested in the undoubtedly important intra-liberal disputes about what forms of recognition a liberal state can/ought to/must adopt to sufficiently appreciate its inner cultural diversity and to protect its minorities from the oppressive tendencies of the dominant culture.²¹⁷ These further forms of recognition are built on the basic one I maintain here – in this important philosophical debate, what is asked is not whether the state should recognize its citizens and minorities as holders of valid moral claims,

²¹⁶ I use the term 'holders of valid moral claims' instead of the original Rawlsian 'sources of valid moral claims'. As I show later on in this section, understanding citizens as *sources* of valid moral claims is already a specific liberal form of political recognition that is not necessary (and not even possible) in every political society. 'Holder of moral claims', on the other hand, aims to be a general expression not tied to any specific political regime or conception of justice.

²¹⁷ In this debate, the key piece of literature is Charles Taylor et al., *Multiculturalism: Examining the Politics of Recognition*, ed. Amy Gutmann, Expanded Paperback edition (Princeton, N.J.: Princeton University Press, 1994).

but how extensive these claims are and what would full recognition practically entail. The basic recognition is presupposed on both sides.

At this point, I do not claim that the political recognition of citizens as normatively relevant must be equal, extensive, or entailing some specific individual and political rights. I only claim that it should be there, at least in its most basic form. Every political regime should acknowledge that its citizens morally matter, just as we in the basic interactions acknowledge that our friends and colleagues and random people that happen to stand in a queue just in front of us morally matter – we acknowledge their moral status and, as a result, not jumping the queue is the least we can do in this respect (and we also feel resentment towards the people that do it).

Political regimes that do not recognize the fact that their citizens morally matter are in an important respect inhuman. What they do in politics is what Peter Strawson claims is ‘practically inconceivable’ in the realm of intersubjective relations. Human beings that are not recognized as morally relevant are viewed fully instrumentally – as an inventory, as pieces of cattle that need not be informed, consulted, or even taken into consideration when the issues regarding their lives are being decided.

A withdrawal of basic recognition is wrong because it asks us to politically deny something that is almost impossible to deny interpersonally. Under normal circumstances, my collaboration with another adult human being enables me to have certain emotional responses towards this human being that I would not have if it were a small child, a mentally disabled person, or a cat (unless I heavily and self-delusionally anthropomorphise the cat). I understand these reactions (resentment, blame, etc.) as symptoms of basic recognition of moral personality. I view the person as having its own interests yet possessing also a capacity to understand the moral features of the situation – which makes her blameworthy if she chooses the wrong option but in this very respect also makes me recognize her and grant her a status of a morally relevant being, a holder of potentially valid moral claims. Thus, I take it that every political withdrawal of basic recognition (most prominent in slaver societies but also in Nazi Germany with regards to the Jewish population) is an abomination. It asks us to collectively deny something that we individually know is true.

A denial of basic recognition to some group or nation or race often occurred in practice, but there was always some false consciousness involved. Stephen Darwall in his *Second Person Standpoint* argues that our natural human interactions force us to view our

fellow human beings as free and equal.²¹⁸ I am not that optimistic and I do not think that Darwall's argument to this extent works. Still, the grain of truth in this position lies in the fact that a political denial of basic recognition to a normal adult human being does require a non-trivial dismissal of the fundamental experience of human interaction.

The second problem with a political denial of basic recognition (which I understand throughout this text as a denial of moral personality and a status of a holder of possibly valid moral claims) lies in the fact that basic recognition is an elementary presupposition of any political relation. There is nothing political in a relation between a shepherd and her flock. It is an instance of domination, pure and simple. Similarly, a master does not have a political relation with a slave.

I am inclined to think that an absence of recognition in political relations is conceptually impossible. Politics starts with recognition. It need not be extensive and need not be equal, but it is necessary. This is true even for slave-owning societies. In Ancient Greece, the political society consisted of the citizens that mutually recognized each other even though it did contain also other, non-recognized people. Yet without this elementary recognition among citizens, the existence of city-states would be impossible.

To summarize, the basic recognition is a necessary condition for a minimally acceptable political regime. First of all, it is conceptually required as no political society can exist without recognition among its members. Secondly, a denial of recognition for certain classes of non-citizens (mostly slaves and unwanted minorities) manifests itself as false consciousness. A normal natural interaction between human beings results in a recognition of the other's moral personality and, consequently, a recognition of her status as a holder of potentially valid moral claims. The Strawsonian reactive attitudes are the clearest manifestation of this tendency. The denial of recognition for political reasons always asks us to ignore this basic fact of human interaction, it asks us to pretend that the others do not possess a moral personality, that they do not or even cannot share the same moral universe with us. But this is manifestly wrong.

A need for recognition is built upon moral personality as I presented it earlier in this chapter. Recognition is the political manifestation of the fact that we possess moral personality and thus see our cooperation as normative, as involving justice and fairness. Yet if every minimally acceptable political regime needs to recognize its citizens, then this need for

²¹⁸ Darwall, *The Second-Person Standpoint*, 268.

recognition can tell us something in favour of the regimes that indeed do provide this recognition. And, as I will argue shortly, liberalism is such regime.

The adoption of recognition as an important political category goes against one important strand of political thinking – Hobbesian. The social struggle perceived by Hobbes consists in clashes of interests. Conflicts over resources are the defining feature of human interaction, with some kind of contract required to coordinate the competing parties and enforce compliance with the agreed rules of conduct. However, the Hobbesian picture does not tell the whole story. As noticed already by Hegel and elaborated by Axel Honneth, who is alongside Charles Taylor the most important proponent of recognition in political philosophy, Hobbes completely ignored the fact that an unease concerning intersubjective recognition of our identity is built into social life from the very beginning – and it strives for political manifestation. Therefore, furthering our interests is not the only (and perhaps not even the most significant) point of human interaction. Our interests are surely important for us, but focusing only on them distorts our understanding of the nature of social conflicts, which is normative from the start. For Honneth (and Hegel), “practical conflict between individuals is understood as an ethical moment within a collective social life.”²¹⁹ Recognition is not only implied in the building of political society. It is also subjectively very important and this fact has far-reaching political consequences

Social life cannot be defined as a pure clash of interests; it is a ‘struggle for recognition’ where people seek the intersubjective acknowledgement of their position as morally relevant beings. As we cannot but get personal with others, we want to be understood and valued as persons. The self-understanding of human beings has an important social and political layer. Only in the eyes of others can I self-affirm myself as a being that I aspire to be. In other words, if I want to view myself as a holder of valid moral claims, this attitude needs to be mirrored in the political regime regulating social relations.

The main task of this thesis, to repeat it once again, is to find ‘what can we philosophically say to justify liberalism as a mode of existence of political society.’ So far in this chapter, I have established a few important conditions that any ‘mode of existence of political society’ needs to comply with in order to reflect and respect the type of beings living under it. A political society is a community of moral persons. This, in itself, is extremely important. The questions of justice, of fairness and legitimacy, necessarily arise in it and need

²¹⁹ Axel Honneth, *The Struggle for Recognition: The Moral Grammar of Social Conflicts*, trans. Joel Anderson (Cambridge, Mass, 1996), 17.

to be answered somehow, with the success and stability of the given political society depending on no small measure on the success of the response. Moreover, given that the social interactions of human beings are manifestly normative from the start, the agents seek for a vindication of their status as morally relevant – they seek for recognition. Therefore, every political society faces the task to create conditions under which this recognition can be given. Every mode of existence of political society should entail recognition. The alternative is frustration and, inevitably, violence.

3.3 Forms of political recognition

An open tyranny is regrettably always an option for human societies. If one part of society finds itself in a strong position of power (and it does not matter whether it is a majority, a minority, a limited group, or indeed a single individual), they might be able rule this society by brute force. This means that the ruling class might not give even the basic recognition to certain groups of people living in society. More often than not, the false consciousness entailed in such a move would produce a certain kind of legitimation story purportedly justifying the given status quo. However, such legitimation stories are bound to be dishonest at least on some level and therefore ultimately unsatisfactory, revealing the naked hubris and violence behind them. In that vein, the subservience of black slaves in American south was justified by an unsubstantiated ‘white man’s burden’ of being superior in every way over the black population. Similarly, the treatment of Jews by Nazis during WW2 was rationalized by a ludicrous claim of their posing an existential threat to the German nation. Finally, in an example favoured by Michael Walzer in his *Spheres of Justice*²²⁰, Helots in Sparta were formally prisoners of war – enemies vanquished in battle, brought back to Sparta. Yet this status was strangely passed on from one generation to the next. Surely, a substantial dose of false consciousness is required to believe that a girl that was born and spent all her life in Sparta cannot be politically recognized because she lost this right in battle as an enemy of Sparta and is still formally in a state of war with her captors.

The question, then, is how we can socially and politically deal with the requirement of recognition once we cannot ignore it. At this point, it is important not to jump to conclusions. Claiming that some form of liberalism is needed for a real political recognition would be

²²⁰ Michael Walzer, *Spheres Of Justice: A Defense Of Pluralism And Equality* (New York: Basic Books, 1984), chap. 11.

absurd. This answer would imply that during virtually all human history the political societies were in a state of denial of what I identified as a most natural way of human interaction.

Indeed, there exist multiple ways for a political society to recognize its members. In what follows, I comment on what I take to be the two most important instances of non-liberal rendering of recognition. That sets the stage for my final arguments in favour of liberalism.

Historically, recognition naturally finds its place in family, where emotional ties and common lineage provide a solid basis for mutual acceptance. Familial and tribal societies have no problems with giving recognition to their members. Since parties see each other as sharing the same ancestry, it is unproblematic to acknowledge a common moral nature. A member of family is always viewed within the family as a 'holder of valid moral claims'. As a source of recognition, this is thus quite satisfactory for societies that function as an aggregate of family units or live in a (still ultimately family-based) tribal system. I am recognized within my family and my family is recognized within the society at large. Therefore, I always carry a certain social and moral status and the relative prestige of my status is determined by my position within the family and the position of the family within the broader society.

Of course, recognition within a family is never equal and is compatible with an oppression of the weaker members, especially women. I would go even further and claim that equality within family is simply impossible. It cannot even be understood as a value in tribal or family-based societies. The members are too close to each other, the roles are too entrenched, and the differing status is too well-established. The difference in between the generations could never be wholly abolished. As a result, in family one never counts just for one (as is blatantly obvious to everyone that wanted to use some democratic procedure to settle an important issue within the family – like where to go on vacation). So even though the inequality within family may not always take the form of strict Roman *paterfamilias*, some forms of precedence and domination are bound to endure.

But despite this potential for oppression of weaker members, the family-based recognition comes quite naturally and proved to be remarkably stable. The small and simple frame of a tribe or a family means that everyone is given a role. Moreover, the shared bonds, common identity, and personal proximity provide a good environment for mutual recognition of moral personality. In these types of societies, nothing more is needed.

However, family as the major source of recognition becomes burdensome and impractical when societies grow larger in size and complexity so that family ties can no longer be the bond that holds society together. This observation is relatively uncontroversial and was emphatically confirmed by Henry Sumner Maine, a prominent English 19th century

legal theorist.²²¹ In his analysis of legal development of ancient societies (prominently Rome), Maine notes that it took centuries for them to develop a notion of civic relationship that was independent on family. In the meantime, ancient Roman society created often quite dubious and complicated rules for adoption and inheritance that supported what was a pure legal fiction of common ancestry. Even when stripped of all emotional ties, the introduction of purely civic relationship in place of this washed down familial one was a slow and arduous process. A recognition of a person as a ‘son of X’ was too deeply entrenched to yield easily. After all, a civic relation as a source of recognition is too complex a notion to develop overnight. Nonetheless, and this is the important point, the rising complexity of ancient Rome made it inevitable and the whole society was, as a matter of practical necessity, pushed away from a family-based conception of recognition.

The same goes for any society that reaches a certain level of sophistication. While some shared recognition will always occur within family and in the mutual relations of families, this form of recognition is unable to remain the backbone of political relations in any society with more than a very limited population and a straight-forward political establishment. A modern society, inevitably counting millions of members and complex web of interrelations, must base its recognition on something else than familial ties.

Besides family, shared beliefs, shared identities, a shared place in a certain picture of human world are the most powerful means of recognition in human societies. We are social and cultural animals and our identity can never be strictly individual. We are always part of something that transcends us and these shared motives can and often do manifest themselves politically, which makes them excellent as carriers of recognition.

A group of people sharing a complex worldview does not need familial ties to recognize each other, especially when the worldview itself delivers some content regarding the role of human being in society, its value, and requested/recommended forms of behaviour. Inevitably, complex worldviews contain just that. In a society where some such worldview is shared, recognition therefore comes from being a part of that common world. Politically, this may enable the regime to give recognition to every member of society.

An example of a (bit idealized) European medieval kingdom may illustrate the case. In it, the majority of almost powerless peasants was ruled and guided by church and feudal hierarchy, with a hereditary monarch at the top. There were rules governing the status, rights, and privileges of each of the classes and they gave the members of aristocracy and clergy a

²²¹ Henry Sumner Maine, *Ancient Law* (New York: Henry Holt and Company, 1906).

much higher and more protected status than that of the peasants. Indeed, peasants were tied to their feudal master and politically disenfranchised.

Thus, the medieval Christian world order contained three classes: those who pray, those who fight, and those who work, with the working majority unmistakably at the bottom. The ruling status of those who fight was justified by adopting a version of an argument elaborated famously by Robert Filmer and later attacked by John Locke.²²² The king had a God-given right to rule the earthly matters while the spiritual realm was to be firmly managed by the Church. In between them, they shared the power and it was the role – or indeed a duty – of peasants to submit to this power on pain of disrespecting God’s commands and the God-intended world order.

Nevertheless, it would be wrong to say that peasants in a medieval kingdom did not receive any recognition in form of an acknowledgement of their moral personality. The recognition they got was deeply embedded in the world order they subscribed to. First of all, they were members of a class that was as indispensable to the world order as the other two. Secondly, they were members of the church, manifested by their participation on common rituals, from baptism to funeral. Their moral personality was acknowledged when they admitted their nature as sinners seeking for salvation. Lastly, the Christian story of salvation testifies that even the lowest humans do matter in God’s eyes. When it comes to self-understanding of the lower classes in this medieval Christian society, they saw their indispensable role; they understood their transcendental nature as manifested by the possession of an eternal soul; and they believed in their worth for the Creator that had sent his only begotten Son to save them. Hence, there could be no grounds for complaint based on a structural lack of recognition. The (presumed) unity of purpose made liberalism unnecessary without undermining the given society. The monarch and the church were given a licence to rule the political and the spiritual life while even the lowliest of peasants were recognized as agents with moral personality.

The sort of recognition coming from a shared worldview is unproblematic when used within confined associations. However, when this recognition is to become a backbone of political regime, a state, it requires a non-trivial unity of purpose and a shared self-understanding throughout the given society. Unsurprisingly, if people do not share a value horizon anymore, basing recognition on it becomes impossible. Any such move under the conditions of plurality is exclusionary, as it in the end denies recognition for large portions of

²²² Robert Filmer, *Patriarcha and Other Writings*, ed. Johann P. Sommerville (Cambridge: Cambridge University Press, 1991).

society that happen to endorse a different set of values or a different worldview – an atheist or a Muslim cannot draw recognition from being a part of the medieval Christian world order and they do not have an option to freely adopt a politically expedient culture.²²³

Most if not all modern societies find themselves in a situation of not having a single conception of world order to draw upon when giving recognition to their citizens. For example, Europe has since the Reformation certainly lost what is now often presented as a sense of unity and fellowship springing from the shared value horizon.²²⁴ This may of course cause lamentations and deep nostalgia, but the ‘good old times’ are not coming back. If contemporary societies wanted to achieve the high level of unity required for the ‘Middle Ages type’ recognition to go through, it would necessitate an unprecedented amount of coercion and brainwashing, if it is possible at all.²²⁵ That means that the types of recognition that were widely available for most societies throughout the history are not available now, for us. This situation, as I claimed above, started to develop with the Reformation and it has been ‘worsening’ (to adopt the nostalgic spirit for a moment) ever since. There are no common rituals for people to perform, there is no general membership in a religious institution, no shared belief in a universal order where everyone finds his/her place. The shared world order cannot be a source or recognition anymore.

The recognition coming from a shared world order proved to be as limited as the one coming from family. This type of recognition depends crucially on the worldview being ‘shared’. If it is not, then it cannot help us. Thus, under the conditions of plurality we must seriously ask what means do we have to establish a scheme of recognition and respect for moral personality in a political society.

Generally speaking, recognition raises a normative problem whose solutions always require certain ‘conditions of fit’ to be fulfilled. If these conditions are not met, the recognition cannot be successful. The success of political recognition therefore depends on two factors – the content of the recognition-giving strategy and the suitability of this strategy for a given society. After all, many recognition narratives (but also normative reasons in general) require the agent in question to hold some prior beliefs or to be in a specific situation.

²²³ For a more extensive developments of these thoughts, see Will Kymlicka, *Multicultural Citizenship: A Liberal Theory of Minority Rights* (Oxford: Clarendon Press, 1996), 84–93. A similar point to mine is also developed by Michael Sandel: “The nation proved too vast a scale across which to cultivate the shared self-understandings necessary to community in the formative, or constitutive sense.” Michael J. Sandel, “The Procedural Republic and the Unencumbered Self,” *Political Theory* 12, no. 1 (February 1, 1984): 93.

²²⁴ Alasdair C. MacIntyre, *After Virtue: A Study in Moral Theory* (Notre Dame: University of Notre Dame Press, 1981).

²²⁵ Rawls calls it “the fact of oppression”. See Rawls, *Political Liberalism*, 37.

In this vein, the family-based recognition for political society requires an uncomplicated structure of social relations, and the shared-value-horizon recognition requires, well, a shared value horizon and a political representation that matches and expresses this value horizon. If these conditions of fit do not obtain, the political societies in question fail to give recognition to their members.

It is a straight-forward consequence of the last paragraph that some modes of giving recognition that were successful in the past may not be successful later on. If a situation of a given society or beliefs of its members change – if the conditions of fit no longer hold – then recognition becomes a problem. In that case, the grounds on which the members used to view themselves as holders of valid moral claims disappear. The society in question thus faces the task to either change the type of recognition it provides or to suppress the struggle for recognition in the class of its members that lack it and subjugate them.

3.4 Liberal recognition

So how does a liberal regime work with recognition? The important point to notice is that in liberal regimes recognition is not given through a third external entity – through family or church membership, race or nation. The citizens are recognized individually and directly. This much is clear with regards to every liberal philosopher, from Locke through Kant and Mill to Rawls or Dworkin. Citizens are recognized as ‘self-authenticating sources of valid moral claims’ which means that there is no outside entity (like family or church) that gives them recognition – it comes from the internal value/dignity/normative standing of the citizens themselves.²²⁶ Indeed, it is often the case that the recognition of citizens is the cornerstone upon which the state is built (typically, it takes a form of a constitution or a declaration of rights around which the fundamental political institutions revolve).

As I argued at the very beginning of this thesis, a liberal regime is defined primarily through its inherent toleration of different forms of life on one hand and a robust system of human rights on the other, so it certainly respects its citizens.²²⁷ Citizens maintain the authority to do important decisions and the state consciously and emphatically abstains from

²²⁶ I take it that this is the main reason why Rousseau, otherwise an integral part of the social contract tradition, is not a liberal. In his theory, the General Will takes precedence as the source of valid moral claims and the citizen takes recognition only through it. See Jean-Jacques Rousseau, *Du Contrat Social* (Oxford, Clarendon Press, 1972).

²²⁷ See ‘Introduction’ and Freedman, *Ideologies and Political Theory*, 144–155.

interfering into large parts of their lives.²²⁸ A system of strong human rights signals that a state is unwilling to intervene in individual's life even if it would have been desirable from a certain viewpoint (if, for example, a higher social or even individual utility would be created for the citizen(s) involved), signalling that the citizen is not a mere holder but indeed a 'self-authenticating source of valid moral claims'.²²⁹ The recognition of citizens thus comes from their protected status as possessors of stringent and extensive rights and from the fact that they are equal, when even the necessary privileges of power are open to fair competition. In short, a liberal regime ticks all the boxes I presented in this chapter so far – it gives recognition, respects moral personality and plurality, and can thus very well accommodate citizens that care about justice.

Liberal regime conveys the respect for the individual as a source of valid moral claims through legal status and the subsequent institutional protection the citizen receives. The impersonal rights, rules, and laws replace family or church membership as the general sources of social respect.²³⁰ This has several advantages vis-à-vis the previously defined sources of recognition. Firstly, the size and complexity of the given society ceases to be a problem as legal relations require neither emotions nor a familial or personal bond. Recognition becomes more structured, better defined, and institutionally enforced. Secondly, respect given by means of legal status is importantly belief-independent. A citizen does not have to endorse any particular identity to receive legal recognition. In a liberal state it does not matter whether one identifies as Russian, catholic, African-American, farmer, or football fan. The basic institutional protection does not depend on any of these identities.²³¹ The political rights are reserved for citizens and the basic rights and institutional protection is granted to every human being within the jurisdiction of the state. Even though there are grey areas in these assertions (some liberal states are notoriously failing to grant citizenship or even any kind of legal status to some of their long-term immigrant inhabitants), the general theoretical presumption behind every liberal regime is one of protection of individuals and their liberty via working institutions and rule of law. As a result, the 'conditions of fit' for recognition given by a liberal regime are in this respect much more permissive than ones given by family or a shared

²²⁸ This line of reasoning is most forcefully argued for in Mill, *On Liberty*.

²²⁹ Rawls, *A Theory of Justice*, 1999, §1.

²³⁰ The importance of 'rights as trumps' for a liberal regime is most emphatically argued for in Ronald Dworkin, *Taking Rights Seriously* (Cambridge, Mass: Harvard University Press, 1978).

²³¹ Again, I focus here on 'basic' institutional protection. It might be the case that certain broadly liberal regimes (such as Canadian Québec) adopt some measures over and above the basic general protection of rights that are identity-dependent. The identity of francophone Québec is protected so that both the French and the English speakers face some restrictions as to what should be an official language of a big company, etc. But all this happens with a solid background of liberal institutions firmly in place.

value horizon. They do not require neither simplicity of political relations nor social unity – which is a significant advantage.

However, liberal recognition carries one implication that should not be overlooked – its entrenched individualism. For liberalism it is impossible *not* to be individualistic as the absence of external recognition-giving entities (which make plural recognition possible) and the respect for a normative standing of an individual leaves it no other choice. This does create a few issues for liberalism. As Jürgen Habermas persuasively argues, the modern legal frameworks have difficulty expressing collective rights and interests.²³² Even when they are acknowledged, they are articulated as a series of individual claims (typically, social exclusion of certain classes of citizens is solved through general welfare programs delivered as a series of benefits to be claimed by individuals, the minority language rights are fulfilled with an introduction of a second official language, giving the individual members of the given minority a right to speak their native language in official communication). Thus, individualism is built very deep into the liberal forms of recognition, at least when it comes to the core liberal rights and principles. As the liberal regime abstains from drawing on any particular identities, the general focus on individual is all that remains.

As a result, there is one important condition of fit for liberal recognition that must not be overlooked. Liberal recognition cannot be successful in societies with strong and unyielding collective identities. If the members of society understand themselves primarily and predominantly through the lenses of their familial, tribal, or religious identity, then the individual liberal rights and entitlements they receive might not make sense to them and might not convey the intended recognition. Under these conditions, liberalism fails.

After establishing the firm place of Rawlsian moral personality in asking about justice, this section elaborated on the direct moral and political consequences of this assertion. A Strawsonian analysis of reactive attitudes served as a platform to move the argument forward. I argued that there is no need for a fundamental justification of a normative nature of human social and political relations as its denial is ‘practically inconceivable’ and humans inevitably view each other as morally relevant and possessing a moral personality. I argued that this factor is politically manifested as a struggle for recognition. The normative aspect of human social relations resurfaces as a need for an arrangement that will enable every member of a

²³² Jürgen Habermas, “Struggles for Recognition in Constitutional States,” *European Journal of Philosophy* 1, no. 2 (August 1, 1993): 128–55.

society to be and feel normatively relevant. The alternative is naked oppression and despotism.

I surveyed what I take to be the two recognition-giving structures dominating human political relations for virtually all known history: family and shared identity. They are both powerful and, when successful, prove that liberalism is not a universally compelling mode of existence for human political society and there may be circumstances under which it is uncalled for. However, they both possess important structural weaknesses. The family-based recognition works on a political level only when the society in question is simple enough for a family to be a meaningful political unit. The shared identity works in giving recognition only when they are truly and (almost) universally shared. All these are very demanding conditions.

Liberalism requires neither unanimity nor does it place limits on the size of society. The recognition of moral personality and worth of its citizens is built into its very identity. These are very powerful considerations in favour of liberalism, although they need an elaboration to provide a compelling philosophical case. That is the aim of the next section.

4. The uniqueness of liberalism

In this section, I develop two arguments that aim to prove that liberalism is the only feasible option for modern political societies. Both arguments stand on the points developed throughout the thesis, especially in this chapter. They take as given that moral personality is presupposed when dealing with the question of political justice. What is more, it is necessary for political regimes to acknowledge moral personality of their members (unless, of course, they opt for naked oppression). If successful, the arguments I develop here provide a strong answer to the question that interested me from the very beginning: what we can philosophically say to justify liberalism as the mode of existence of political society.

4.1 An argument from pluralism

The direct neo-Berlinian arguments from pluralism to liberalism I analysed in the first chapter did not work. Pluralism remained an elusive notion and its normative consequences were unclear, to say the least. On the other hand, Rawls also uses pluralism heavily – but not as a part of his argument. For him, ‘a fact of pluralism’ is a circumstance we need to work

with and it underlines his commitment to political as opposed to comprehensive liberalism. I take it that it is possible to construct a persuasive argument justifying liberalism as the best mode of existence for a modern political society somewhere in between these two positions, using the tools I developed in this chapter.

The basic assertion is simple: under the modern conditions of pluralism, no political regime *except liberalism* can provide the requisite respect for moral personality so as to give a sufficient amount of recognition for its citizens. To elaborate this assertion (already sketched in the previous section), the reason for the uniqueness of liberalism is the fact that it provides the same recognition whether a given citizen believes in it or not. Even illiberal citizens have rights, are robustly protected, and are free to organize their lives. The state gives them the same amount of respect; they get the same amount of freedom of expression. All in all, citizens are all equal whether they agree with the establishment or not. Respect in a liberal state is independent from private beliefs and therefore can function well even under the conditions of pluralism. On the other hand, the non-liberal legitimation stories, like the one provided by the medieval society, are heavily belief-dependent. A lowly peasant that happens to be an atheist can get no consolation from the fact that his role is indispensable for a God-given world order, or that he could be a full member of Catholic Church. That is why non-liberal recognition cannot function in modern pluralistic society – this society simply does not provide it with the necessary conditions of fit. On the other hand, liberalism copes with pluralism very well.

Of course, this does not mean that all citizens must be satisfied with the recognition they get. To use a fashionable example, many Muslim citizens of liberal countries seem to be offended by the liberty of others to mock their religion. But, again, I take this to be an intra-liberal dispute about balancing the freedom of expression with the (perhaps sometimes justified) sensibilities of some parts of population. The laws prosecuting Holocaust deniers are a good example of circumstances where recognition of a historical trauma may prevail over the freedom of expression. However, the basic recognition is untouched by these considerations as all citizens still get a full set of rights, complex institutional protection, and an equal legal status regardless of which side of the debate we think is right. The basic recognition remains in place.

Minor issues aside, the fully spelled out argument goes like this:

- 1) Human beings care about justice and the political arrangements of society are always susceptible to the question whether they are in fact just;

- 2) The question of justice inherently implies the picture of human beings as possessing moral personality – as beings that think normatively and possess a set of personal interests, a conception of good;
- 3) Moral personality that human beings possess inevitably manifests itself in interpersonal, social, and political relations (reactive attitudes being a prime example);
- 4) Moral personality manifested politically takes the form of a struggle for recognition, a pursuit of political acknowledgement of persons as holders of moral claims that every state must answer to somehow – either by accommodating it or by an overt oppression;
- 5) Avoiding oppression, recognition can take multiple forms and historically was closely connected with family and particular religious and social identities;
- 6) However, under the modern conditions of plurality a political recognition cannot work with particular identities as they are too many, too complex, and, most importantly, they are not generally shared – the conditions of fit for these forms or recognition do not obtain;
- 7) Therefore, the only acceptable option is to recognize human beings as human beings, focusing on their basic moral and political status, their moral personality, and their rights as equals;
- 8) Point 7 implies liberalism – so liberalism is the only meaningful option for political societies under the modern conditions of pluralism.

To put the argument in short: What can we say to justify liberalism as the mode of political existence of society? We can say that liberal regimes are able to be successful even if the conditions of fit do not obtain for other types of regimes – and in case of pluralism, they do not. Pluralism therefore favours liberalism.

4.2 An argument from inwardness

However, liberalism is not the most appropriate system for contemporary political societies solely because of pluralism. I want to advance a more complex argument, claiming that the way we understand ourselves in modern times is inimical to non-liberal forms of recognition. I label the specific modern self-understanding of human beings as a belief in ‘inwardness’ of value. Thus, I claim that apart from pluralism or complexity of society the inwardness of value becomes another strong condition of fit, excluding many forms of

political recognition, in the end making liberalism the only feasible political option, even though there were other options open in the past.

To explain, there is a virtual unanimity between historians of ideas that there has been an important change of self-understanding in our societies sometimes between Middle-Ages and the 20th century. It can be best described as a turn towards the inwardness of value. This process was most famously described by Charles Taylor in his *Sources of the Self* in the story of subtle internalization from Augustine to Descartes to Locke and culminating in the age of enlightenment.²³³ I cannot possibly hope to trace the historical developments of human self-understanding in this thesis – it is a colossal task very far beyond the scope of the current enquiry. Also, it is not my ambition to present a historical argument. I only want to provide what will hopefully sound like a plausible account of our modern self-understanding and then show how well it fits liberalism.

All the general politically relevant characteristics of human beings I identified above (having moral personality, being a holder of valid moral claims, seeking for recognition) are now almost automatically understood as springing from the depths of the human being taken in itself – although it was not always like that, as was famously captured by Constant.²³⁴ He described the ancient societies as giving recognition to citizens externally, via a political establishment, with no independent internal value understood as belonging to the human beings as individuals. The individuals might have had a political standing, but they did not have ‘inalienable’ individual rights in the sense we now take almost as a given. The rights were always maintained and guaranteed externally. So even though the ancient citizens did not have an independent ‘inalienable’ standing as individuals, as independent moral persons, they did receive recognition via the state – although they could have been stripped off it any time.

It is immediately apparent that our self-understanding is nothing like this. We the moderns are all descendants of Kant in this respect, looking with awe at internal normative standing of human being. We situate value firmly inside a person, with no necessary connection to her holding a particular worldview or to her belonging to a given church or *polis*. Our respect for moral personality and political recognition in general do not come from an external source. Human beings possess value in themselves, it is not given to them by any

²³³ Taylor, *Sources of the Self*, 127–177.

²³⁴ Benjamin Constant, in *The Liberty of Ancients Compared with that of Moderns* (Cambridge: Cambridge University Press, 1988), 309–28.

institution or authority.²³⁵ As a result, the agent in this model is strictly speaking not ‘given’ recognition from some institution or law. Law only recognizes the agent for what he already is – a self-authenticating source of valid moral claims. The Rawlsian language is very telling here. It was always the case that human beings could claim or assert some right, possessions or status. However, only in modern times humans became the proper, ‘self-authenticating’ sources of these claims, with their recognition and respect for moral personality unmediated externally.

Now it may be claimed that what I do here is that I smuggle comprehensive Kantianism back into the justification of liberalism, introducing yet another philosophical normative *fiat*, declaring something to be valuable in a specific way. However, that would be a misleading way to put it. My claim is that we the moderns cannot but understand human beings as valuable *simpliciter*, regardless of their external relations. Obviously, this is compatible with the Kantian views on dignity and moral autonomy. However, it is as compatible with a host of other manifestly non-Kantian doctrines, including the self-ownership theory of Locke and the theory of Berlin, both of which I analysed in this thesis.²³⁶ And I do not simply declare the value of human being. My claim is hermeneutic, about our self-interpretation. It is perhaps clearest understood in its negative form: as we understand human beings, they (their lives, their suffering) can never cease to morally matter. This seems to be the consensus across the nations and cultures in modern world.

Compared with the internal value of human beings, the external connections (belonging to a certain family, class, polis, or religion) are, politically, of little normative importance. This was not always the case but it is certainly the case now. Of course, the external connections matter in real life as they have enormous consequences on fortunes of a given individual. But they do not change our moral and political value. To use the simplest example, religion, class, or nationality really should not matter when a ship hits an iceberg and the captain must decide whom to put on the limited number of lifeboats. These factors are, to use the Rawlsian vocabulary once again, “arbitrary from a moral point of view”²³⁷ – at least from the perspective of us, the moderns, who understand value as internal to human beings and non-contingent on race or nationality.

²³⁵ Maybe the clearest contrast here can be provided by the medieval practice of excommunication. In that process, all grounds of recognition were taken away from the person. It was an absolute devaluation of status, a complete social death. The external relation of membership in Catholic Church was thus a necessary condition for any recognition. When it was withdrawn, there was almost nothing left – which is very much unlike the prevailing contemporary understanding of human being and its intrinsic value.

²³⁶ See Chapter I.

²³⁷ Rawls, *A Theory of Justice*, 1999, 14.

However, if a human being is regarded as a proper source of value independently from the external connection and the worldview adopted, it importantly transforms the whole dynamic of social relations. Basically, it creates a necessary requirement, a condition of fit, for political recognition. It must work from within, from the value that the human being possesses in itself. The external entities are no longer normatively powerful enough to be bestowing the political status. As a result, it drastically limits the application of political categories I developed in this chapter. A political recognition, an acknowledgement of moral personality require an appreciation of an inner value of human individual.

The resulting model of social relations is a one of a community of beings possessing inner value with no outside force discriminating between them and providing a justification for the subsequent unequal treatment. The citizens are understood as bearers of moral personality and, in virtue of this fact, are recognized as sources of valid moral claims. Indeed, this basic model is exactly the one that Rawls starts his enquiry with. The appreciation of the inner value of human beings leads to the understanding of political community as a system of cooperation between the free and equal. Thus, “the first fundamental question of political society” concerns “the most appropriate conception of justice for specifying the fair terms of social cooperation between citizens regarded as free and equal.”²³⁸

In this model, a liberal regime is clearly the one that is called for, while the illiberal options are put off by its civic and egalitarian nature (so family-based recognition is excluded) and by its refusal to rely on external political sources of value (so shared-worldview-types of recognition are excluded as well). The basic liberal values of freedom and equality have already been affirmed. They only await a more specific decontestation. Simply put, no other form of recognition can live with the intrinsic value of human beings understood as possessors of moral personality.

Let me put the second argument a bit more rigorously. It shares the initial five steps with the first one:

- 1) Human beings care about justice and the political arrangements of society are always susceptible to the question whether they are in fact just;
- 2) The question of justice inherently implies the picture of human beings as possessing moral personality – as beings that think normatively and possess a set of personal interests, a conception of good;

²³⁸ Rawls, *Political Liberalism*, 3.

- 3) Moral personality that human beings possess inevitably manifests itself in interpersonal, social, and political relations (reactive attitudes being a prime example);
- 4) Moral personality manifested politically takes the form of a struggle for recognition, a pursuit of political acknowledgement of a person as a holder of moral claims that every state must answer somehow – either by accommodating it or by an overt oppression;
- 5) Avoiding oppression, recognition can take multiple forms and historically was closely connected with family and particular religious and social identities;
- 6) However, the modern understanding of a moral and political value of human being as independent from external connections makes it impossible for these forms of recognition to be successful, as they rely exactly on these external connections;
- 7) If human beings are thought to be inherently valuable and if external sources of recognition are rejected, then there is no way to normatively distinguish between citizens from a general political perspective;
- 8) When citizens are viewed as normatively relevant yet undistinguished, what follows is a firm presumption in favour of equality and freedom.
- 9) The presumption in favour of freedom and equality inevitably results in liberalism.

In a sense, my argument ends exactly where Rawls starts – with the basic ideas of fairness, freedom, equality, and general worth of human beings. This standpoint necessarily leads to liberalism as I understood it, resulting in the acceptance of the ‘Millite core’. If this standpoint is inevitable for modern political societies, then it is a strong consideration that can be mounted in favour of liberalism: A political regime needs to recognize its members, yet the only available form of recognition is one that respects them as individuals with an innate value. This makes for a general political presumption in favour of freedom and equality, implying liberalism.

The two arguments I proposed can be summed up as two general conditions for a normative appropriateness of a political regime. If a society comes to view human beings as internally valuable, the external forms of recognition become ineffective. A similar situation obtains for pluralism. If a society is diverse, then the external forms of recognition become inadequate. If these conditionals (and their premises) hold, liberalism is the only regime that can treat its citizens with institutional respect and recognition they ought to get as agents with moral personality, as internally valuable beings regardless of their other identities. These are quite weighty considerations favouring liberalism as a mode of existence of political society.

With the main arguments done, I want to take a look at two relatively natural and obvious lines of criticism. Firstly, it is possible to claim that by operating with an inwardness of value, recognition, and sources of moral personality, I certainly introduce metaphysical stipulations in the Rawlsian sense and therefore I am untrue to his basic ambition to be political, not metaphysical. But I strongly disagree with this characterization. In developing a strategy for liberal justification, I focused on the nature of citizens as political actors implied in their preoccupation with political justice. I elaborated on strictly social and political aspects of this preoccupation – on the struggle for recognition. The first stages of my argument (where I pursued the general characteristics of human beings as moral agents seeking for recognition) thus ground the moral personality and the struggle for recognition in purely political aspirations of human beings. No non-political comprehensive notions are present.

In the later stages of my arguments, I maintained that no political regime can be justified only from natural, moral, or political characteristics shared by all human beings. We need additional *a posteriori* elements, the conditions of fit. I identified two of them: the pluralism of contemporary societies and our self-understanding that pervades our thinking about value, locating it inside human being. Pluralism is clearly a political consideration. When it comes to the intrinsic value of human beings, it certainly is a heavyweight moral statement. However, as I already indicated above, the acceptance of this intrinsic value is not a unique feature of one specific comprehensive doctrine like Kantianism. In fact, I take this presumption to be a firm part of the overlapping consensus in contemporary societies, both amongst real people and amongst philosophers. Barring the extremist and fundamentalist views, all major philosophical doctrines and religions take the inner value of human beings as an unproblematic part of their worldview. It may be true that the source of this value for many religions lies strictly speaking outside of the person – it comes from the status of human beings as parts of God's creation. Yet, this status is (in all non-extremist religions) belief-independent and is not contingent upon membership in a church or a nation. Therefore, a human being does not require an external recognition from some institution or church for its internal value to be acknowledged. The presumption of an intrinsic value of human being is thus political in the sense that it is a part of an overlapping consensus, and it plays a crucial role in type of political recognition that inevitably results in liberalism, broadly speaking.

The second objection takes issue with the conditional character of my argument. Isn't it the case that I need to prove that the conditions really hold, that we *should* understand human beings as inherently valuable? I believe that this type of objection is misguided as the argument it asks for is both impossible and unnecessary for my conclusion. Our self-

understanding is the account we give of ourselves, our place and role in the universe. It cannot be proved or disproved by a simple argument, it is a hermeneutic fact. Its role is to make sense of the world around us and in this sense it is a precondition of every argument. A picture of humans as beings with internal depth and normative standing does fulfil this task. We understand art as an expression of a unique individual insight coming from within; we hold that innocent human suffering is an evil regardless of the circumstances and should be stopped, if possible; we maintain that we are entitled to at least a basic respect from others, not in virtue of any external factor but because we are human and that (morally and politically) counts for something. None of this is automatic or necessary – but these are the pillars that hold the modern world, universally shared in great majority of contemporary societies, supported both secularly and religiously. And I see no indication of a paradigm shift that would alter these convictions and thus change the central hermeneutic fact.

Surely, our self-understanding *may* change. There might come a time when a different self-conception will make more sense to us. But until then, liberalism is bound to be the only acceptable political regime for societies that aim not to suppress their citizens but to recognize them. The thought (although centred more narrowly on freedom) is best expressed in the last paragraph of Isaiah Berlin's *Two Concepts of Liberty*:

It may be that the ideal of freedom to choose ends without claiming eternal validity for them, and the pluralism of values connected with this, is only the late fruit of our declining capitalist civilisation: an ideal which remote ages and primitive societies have not recognised, and one which posterity will regard with curiosity, even sympathy, but little comprehension. This may be so; but no sceptical conclusions seem to me to follow. Principles are not less sacred because their duration cannot be guaranteed. Indeed, the very desire for guarantees that our values are eternal and secure in some objective heaven is perhaps only a craving for the certainties of childhood or the absolute values of our primitive past. 'To realise the relative validity of one's convictions', said an admirable writer of our time, 'and yet stand for them unflinchingly is what distinguishes a civilised man from a barbarian.' To demand more than this is perhaps a deep and incurable metaphysical need; but to allow such a need to determine one's practice is a symptom of an equally deep, and more dangerous, moral and political immaturity.²³⁹

²³⁹ Berlin, *Four Essays on Liberty*, 172.

Conclusion

This thesis followed two main lines of enquiry. One was centred on pluralism and its role in political societies and the second one focused on a conception of person present or somehow assumed in political philosophy. Although the two lines inevitably interacted and were united by the common concern for a credible justification of liberalism, it is nonetheless possible to follow them separately.

The first, pluralist, line started with Isaiah Berlin, who was the first to draw a connection between the inescapable variety of human values and political freedom. His account does not lack sophistication, but it does lack a certain philosophical rigour. That proved to be a challenge for his followers, who tried to clear out his position but happened to throw away its sophistication as well. Presented as a simple argument, the passage from pluralism to liberalism has an air of a paradox, as it tries to make a move from a plurality of values to a restricted range of them. Thus, on neo-Berlinian account, the role of pluralism is hopelessly problematic and the argument for liberalism does not work.

In Rawls, pluralism does not play a direct role in the argument for justice as fairness. Its role is in the background, explaining why he is interested in political liberalism in the first place. However, as I argued in Chapter II, pluralism presents a difficulty for Rawls. His conception of person, of freedom, equality, and moral personality is either assumed as unproblematic when it in fact is not, or is inadequately argued for. Rawls therefore has no means to defend his conception of person as the only adequate one to anchor a conception of justice when confronted with alternatives inevitably existing in a pluralist society. As a result, his argumentation either does not work or is incomplete.

My aim in Chapter III was to overcome the problems with pluralism while still trying to constructively use it as a part of a justification of liberalism. I argued for a conception of person necessarily implied in any pursuit of justice. This necessary implication makes the conception of person immune from pluralist charge of arbitrariness. After all, if a conception of person is implied in the pursuit of justice, it cannot be only one of many possible conceptualizations of human being. It is the appropriate one for a philosophical inquiry into justice.

Armed with this conception of person, I argued that even though an infinite variety of political regimes is possible for human societies, there are certain conditions that limit this number – and a pluralism of worldviews is one of them. Under the conditions of pluralism, a range of possible political regimes is restricted to the ones that are able to give a belief-independent recognition to their citizens. And so far, liberalism is the only type of regime that we know is able to fulfil this condition.

The second line of enquiry followed the possibility of establishing a normative conception of person that would directly set constraints for a political regime. In Chapter I, I analysed popular theories adopting the most straight-forward normative conception of this type – self-ownership. I found it wanting on multiple accounts. Specifically, I argued that such a paper-thin conception of person (consisting in one normatively-heavy assertion) is unable to be refined-enough to be persuasively justified and must either over-rely on its alleged intuitive plausibility or function as a normative *fiat*, an article of faith.

It is one of the great virtues of Rawls that he consciously tries to avoid the heavy normative proclamations of the type used by the proponents of self-ownership. Nevertheless, his arguments do move in a circle, where liberal assumptions vindicate liberal conclusions. My interpretative attempts to break the circle either by grounding the assumptions in the public culture or by producing an independent argument for them failed. The normative conception of person used by Rawls remained unjustified, in spite of its considerable refinement and theoretical sophistication.

In Chapter III, my aim was to break the Rawlsian circle by the ‘Copernican’ arguments presenting moral personality as an unavoidable feature of human beings interested in justice. After its establishment, I searched for its political manifestations in the struggle for recognition. In my final argument, I stipulated what I consider to be the crucial feature of our contemporary understanding of value – the worth of human beings is recognized independently of their external identities and connections. This final premise enabled me to show liberalism as uniquely satisfying the presented normative ‘conditions of fit’ for contemporary political societies – when holders of moral personality seek for political recognition and they do not see external entities granting it, the remaining source of value is internal to the individual. We see a human being as an entity to be respected by itself, as a holder of rights and entitlements against everyone, including the political institutions. And that implies liberalism.

I would like to end by making a gesture towards some more practical implications of my arguments, as they are directly relevant to the major political and international concerns of today. I started this thesis with a worry that there seems to be a wide gap between the relative confidence of liberal societies in their basic principles as opposed to a perceived lack of philosophical support for them. I presented a few arguments that might help close that gap. However, there still are quite a few illiberal countries in the world today and some of them seem to thrive or at least be very stable. Russia, China, and Saudi Arabia are the best-known examples. This naturally raises the question of the relevance or at least the scope of my arguments. If liberalism is the only option left for contemporary societies, how are we to understand the existence of the successful illiberal ones?

I hold that the arguments I presented are actually quite useful for analysing the situation of the illiberal countries listed above. They all face the task to recognize their citizens (as holders of moral personality). In case they fail, their only alternative is naked oppression. All three of these countries make a vigorous effort to provide sources of recognition to their inhabitants from illiberal sources: Saudi Arabia from religion and Russia and China from triumphalistic nationalism. Ideally, their politically powerless citizens could find their value and dignity in being a part of a larger story of heroic and dominant nation or a true religion, whilst lacking the rights and institutional protection common in liberal democracies. However, these sources of recognition are very difficult to maintain, exactly for the reasons I identified in Chapter III.

Pluralism becomes an enemy these illiberal countries can fight but never defeat. A theocracy like Saudi Arabia makes it impossible for non-Muslims to identify with the regime that makes Koran and Sharia law to be the cornerstones of political society (Koran in declared to be a constitution of Saudi Arabia). The same goes for Russians and Chinese that are cosmopolitan or not part of the ethnic majority. With regards to these citizens, a more-or-less overt oppression is the only strategy available for illiberal countries to pacify them, as is more than evident in Russian Caucasus and Chinese Uyghur province (and recently in Hong Kong). The systematic reliance on censorship and heavy-hand police tactics reveals great problems for the regimes involved. They know that the forms of recognition they are able to provide are inherently limited and unsuccessful at least with respect to certain parts of population. Thus they can only hope for containment, so that doubts and dissatisfaction do not spread so far as to cause serious social unrest. The very use of this tactics must be viewed as an important deficiency of the regime unable to win the obedience and allegiance of its own citizens. These

regimes therefore fail in justifying themselves to their own citizens – and the only strategy to put an end to this failure is for them to become liberal.

The impact of the inwardness of value is far more abstract and therefore much more difficult to show on concrete examples – but I believe that this motive is very much present as well. Of course, the self-understanding of human beings as self-authenticating sources of valid moral claims is most prevalent in the traditional western liberal democracies. The argument I developed is therefore most convincingly applied to them – there simply is no alternative regime for these societies. No grand narrative could ever justify the state to strip the liberal rights from its citizens while still somehow maintaining their recognition.

However, it would be wrong to claim that the inwardness of value is not present in the illiberal countries as well. In the basic civic activism, in protesting injustice and abuses of power, many inhabitants of these countries feel that the state should respect them more and be accountable to them. Unless we believe (as Vladimir Putin seems to) that all the numerous pro-democratic movements in the last 15 years were the product of sophisticated CIA plots, the inevitable conclusion is that there is a very broad appetite towards liberal freedoms and a transparent state respecting its citizens. Thus, the picture of human being as inherently valuable (with all the political consequences that follow) seems to become popular in many traditionally illiberal countries. And if a significant part of citizens understand themselves as self-authenticating sources of valid moral claims, they are bound to sooner or later clash with the regime that does not treat them as such and does not give them the recognition they think they deserve.

As with pluralism, the illiberal regimes are engaged in an uphill and hopeless battle against what they try to present as rotten and decadent (Russia), un-Chinese (China), or sinful (Saudi Arabia) West. However, it is very difficult for them to directly challenge the basic idea of respect and power for citizens. This self-understanding seems to be attractive for modern societies across different cultures. The coercive measures and the subsequent fear in the population may of course work for a long time in suppressing the liberal drive, but the pretention of general acceptability of illiberal regimes is necessarily lost.

Adopting a vocabulary from Rawls' *Law of Peoples*, there used to be a category of political regimes that might be labelled 'decent but illiberal'.²⁴⁰ One of the conclusions of this thesis should be that this category is hopelessly lost because of plurality and the modern understanding of value. The political regimes in this category either become liberal and

²⁴⁰ John Rawls, *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 2001), 63.

maintain their decency, or remain illiberal and lose it. When we ask “what can we say to justify liberalism as a mode of political existence of society”, the answer put in this way is that liberalism is *the only decent* mode of existence of political society. And that is a weighty consideration.

Bibliography

- Allison, Henry E. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.
- Audard, Catherine. *Individu et justice sociale (autour de John Rawls)*. Paris: Seuil, 1988.
- . *John Rawls*. Acumen, 2007.
- Audi, Robert. "Intuitionism, Pluralism and the Foundations of Ethics." In *Moral Knowledge? New Readings in Moral Epistemology*, 101–36. Oxford: Oxford University Press, 1996.
- Barry, Brian. "John Rawls and the Search for Stability." *Ethics* 105, no. 4 (1995): 874–915.
- Berlin, Isaiah. *Concepts and Categories: Philosophical Essays*. London: Pimlico, 1999.
- . *Four Essays on Liberty*. London: Oxford Paperbacks, 1969.
- . *Russian Thinkers*. London; New York, N.Y.: Penguin Classics, 2008.
- . *The Crooked Timber of Humanity: Chapters in the History of Ideas*. 2nd edition. Princeton: Princeton University Press, 2013.
- . "The Decline of Utopian Ideas in the West." In *The Crooked Timber of Humanity*, 20–48. Princeton: Princeton University Press, 2013.
- . "The Pursuit of the Ideal." In *The Crooked Timber of Humanity*, 1–19. Princeton: Princeton University Press, 2013.
- Brenkert, George G. "Self-Ownership, Freedom, and Autonomy." *Journal of Ethics* 2, no. 1 (1998): 27–55.
- Burrow, Sylvia. "Reasonable Moral Psychology and the Kantian Ace in the Hole." *Social Philosophy Today* 17 (2001): 37–55.
- Carter, Ian. "Respect and the Basis of Equality." *Ethics* 121, no. 3 (2011): 538–71.
- Chang, Ruth. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, Mass.: Harvard University Press, 1998.
- Christman, John. "Self-Ownership, Equality, and the Structure of Property Rights." *Political Theory* 19, no. 1 (1991): 28–46.
- Cohen, G. A. *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press, 1995.
- Constant, Benjamin. In *The Liberty of Ancients Compared with that of Moderns*, 309–28. Cambridge: Cambridge University Press, 1988.
- Crowder, George. *Liberalism and Value Pluralism*. New York: Continuum, 2002.
- . "Two Concepts of Liberal Pluralism." *Political Theory* 35, no. 2 (2007): 121–46.
- . "Value Pluralism, Diversity and Liberalism." *Ethical Theory and Moral Practice*, forthcoming, 1–16.
- Daniels, Norman. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge University Press, 1996.
- . "Reflective Equilibrium." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2013., 2013. <http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>.
- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, Mass.: Harvard University Press, 2009.
- Doppelt, Gerald. "Is Rawls' Kantian Liberalism Coherent and Defensible?" *Ethics* 99, no. 4 (1989): 815–51.
- Dworkin, Ronald. *A Matter of Principle*. Cambridge, Mass.: Harvard University Press, 1985.
- . *Taking Rights Seriously*. Cambridge, Mass.: Harvard University Press, 1978.
- Eshete, Andreas. "Contractarianism and the Scope of Justice." *Ethics* 85, no. 1 (1974): 38–49.
- Feser, Edward. "Personal Identity and Self-Ownership." *Social Philosophy and Policy* 22, no. 2 (2005): 100–125.

- Filmer, Robert. *Patriarcha and Other Writings*. Edited by Johann P. Sommerville. Cambridge: Cambridge University Press, 1991.
- Fisk, Milton. "History and Reason in Rawls' Liberalism." In *Reading Rawls*, edited by Daniel Norman. Stanford: Stanford University Press, 1989.
- Flikschuh, Katrin. *Kant Contra Cosmopolitanism: Assessing the Global Justice Debate*. Cambridge: Cambridge University Press, forthcoming.
- Freeden, Michael. *Ideologies and Political Theory: A Conceptual Approach*. Oxford: Oxford University Press, 1998.
- Freeman, Samuel. "Congruence and the Good of Justice." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 277–315. Cambridge University Press, 2003.
- . "Illiberal Libertarians: Why Libertarianism Is Not a Liberal View." *Philosophy and Public Affairs* 30, no. 2 (2001): 105–51.
- Gallie, W. B. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56, no. 1 (1955): 167–98.
- Galston, William A. "Moral Personality and Liberal Theory: John Rawls's 'Dewey Lectures.'" *Political Theory* 10, no. 4 (1982): 492–519.
- Gaus, Gerald F. *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project*. London: Sage Publications, 2003.
- Gauthier, David. *Morals By Agreement*. Oxford: Oxford University Press, 1986.
- Gilbert, Margaret. *On Social Facts*. Princeton University Press, 1992.
- Gray, John. *Berlin*. London: Fontana Press, 1995.
- . *Post-Liberalism: Studies in Political Thought*. New York: Routledge, 1996.
- . *The Two Faces of Liberalism*. Cambridge, UK: Polity Press, 2000.
- Habermas, Jürgen. "Reconciliation Through the Public Use of Reason: Remarks on John Rawls's Political Liberalism." *Journal of Philosophy* 92, no. 3 (1995): 109–31.
- Habermas, Jürgen. "Struggles for Recognition in Constitutional States." *European Journal of Philosophy* 1, no. 2 (August 1, 1993): 128–55.
- Hardy, Henry, and George Crowder. *One and the Many: Reading Isaiah Berlin*. Amherst, N.Y.: Prometheus Books, 2007.
- Hare, R. M. "Review: Rawls' Theory of Justice—I." *Philosophical Quarterly* 23, no. 91 (1973): 144–55.
- Heidegger, Martin. *Being and Time*. Oxford: Wiley-Blackwell, 1978.
- Hobbes, Thomas. *Leviathan*. Oxford: Oxford Paperbacks, 2008.
- Honneth, Axel. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Translated by Joel Anderson. Cambridge, Mass, 1996.
- Kant, Immanuel. *Critique of Pure Reason*. Edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1999.
- . *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press, 2012.
- . *Kant: The Metaphysics of Morals*. Edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1996.
- . *Lectures on Ethics*. New York: Cambridge University Press, 2001.
- Kekes, John. *The Morality of Pluralism*. Princeton, N.J.: Princeton University Press, 1996.
- Klosko, George. "Political Constructivism in Rawls' Political Liberalism." *The American Political Science Review* 91, no. 3 (September 1997).
- Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.
- Kymlicka, Will. *Multicultural Citizenship: A Liberal Theory of Minority Rights*. Oxford: Clarendon Press, 1996.
- Levinas, Emmanuel. *Totality and Infinity: An Essay on Exteriority*. Pittsburgh: Duquesne University Press, 1969.
- Lippert-Rasmussen, Kasper. "Against Self-Ownership: There Are No Fact-Insensitive Ownership Rights Over One's Body." *Philosophy and Public Affairs* 36, no. 1 (2008): 86–118.
- Locke, John. *Two Treatises of Government*. Cambridge: Cambridge University Press, 1988.

- . *Two Treatises of Government*. Reprint edition. Clark, New Jersey: The Lawbook Exchange, Ltd., 2010.
- MacIntyre, Alasdair C. *After Virtue: A Study in Moral Theory*. Notre Dame: University of Notre Dame Press, 1981.
- Maine, Henry Sumner. *Ancient Law*. New York: Henry Holt and Company, 1906.
- Mendus, Susan. *Justifying Toleration: Conceptual and Historical Perspectives*. Cambridge: Cambridge University Press, 2009.
- Mill, John Stuart. *On Liberty*. Oxford: Oxford Paperbacks, 2008.
- Mill, J. S. *Utilitarianism*. Edited by Roger Crisp. Oxford: OUP, 1998.
- Moore, Matthew J. "Pluralism, Relativism, and Liberalism." *Political Research Quarterly* 62, no. 2 (n.d.): 244–56.
- Mulhall, Stephen. *The Routledge Guidebook to Heidegger's Being and Time*. London: Routledge, 2013.
- Nagel, Thomas. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press, 1979.
- . *The View from Nowhere*. New York: Oxford University Press, 1989.
- Narveson, Jan. "Libertarianism Vs. Marxism: Reflections on G. A. Cohen's Self-Ownership, Freedom and Equality." *Journal of Ethics* 2, no. 1 (1998): 1–26.
- Nozick, Robert. *Anarchy, State and Utopia*. Basic Books, 1974.
- O'Neill, Onora. "The Method of A Theory of Justice." In *John Rawls: Eine Theorie Der Gerechtigkeit*, edited by Otfried Hoffe, 206–18. Berlin: Akademie Verlag, 1998.
- Otsuka, Michael. *Libertarianism Without Inequality*. Oxford University Press, 2005.
- Otsuka, Michael, Peter Vallentyne, and Hillel Steiner. "Why Left-Libertarianism Is Not Incoherent, Indeterminate, or Irrelevant: A Reply to Fried." *Philosophy and Public Affairs* 33, no. 2 (2005): 201–15.
- Pateman, Carole. "Self-Ownership and Property in the Person: Democratization and a Tale of Two Concepts." *Journal of Political Philosophy* 10, no. 1 (2002): 20–53.
- Plato. *The Republic*. Translated by Desmond Lee. London: Penguin Classics, 2003.
- Pogge, Thomas W. *Realizing Rawls*. Ithaca: Cornell University Press, 1989.
- Quong, Jonathan. *Liberalism without Perfection*. Oxford: Oxford University Press, 2011.
- Rawls, John. *A Theory of Justice*. Harvard University Press, 1971.
- . *A Theory of Justice*. Revised edition. Cambridge, Mass: Harvard University Press, 1999.
- . *Collected Papers*. Harvard University Press, 1999.
- . "Justice as Fairness." *Philosophical Review* 67, no. 2 (1958): 164–94.
- . *Justice as Fairness: A Restatement*. Harvard University Press, 2001.
- . *Political Liberalism*. New York: Columbia University Press, 1993.
- . *The Law of Peoples*. Cambridge, Mass.: Harvard University Press, 2001.
- Ricoeur, Paul. "Le Cercle de La Demonstration." In *L'individu et Justice Sociale (autour de John Rawls)*, edited by Catherine Audard. Paris: Seuil, 1987.
- Rorty, Richard. *Objectivity, Relativism, and Truth (Philosophical Papers, Vol. 1)*. Cambridge ; New York: Cambridge University Press, 1991.
- Rousseau, Jean-Jacques. *Du Contrat Social*. Oxford, Clarendon Press, 1972.
- Sandel, Michael J. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1998.
- . "The Procedural Republic and the Unencumbered Self." *Political Theory* 12, no. 1 (February 1, 1984): 81–96.
- Scanlon, Thomas. "Rawls on Justification." In *Cambridge Companion to Rawls*, edited by Samuel Freeman. Cambridge: Cambridge University Press, 2003.
- . *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press, 1998.
- Sobel, David. "Backing Away From Libertarian Self-Ownership." *Ethics* 123, no. 1 (2012): 32–60.
- Stolley, Kathy. *The Basics of Sociology*. Westport, Conn: Greenwood, 2005.
- Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.
- Talisse, Robert. *Pluralism and Liberal Politics*. Routledge, 2013.

- Talisse, Robert B. "Value Pluralism and Liberal Politics." *Ethical Theory and Moral Practice* 14, no. 1 (2011): 87–100.
- Taylor, Charles. *Sources of the Self: The Making of the Modern Identity*. Cambridge, Mass: Harvard University Press, 1992.
- Taylor, Charles, Kwame Anthony Appiah, Jürgen Habermas, Stephen C. Rockefeller, Michael Walzer, and Susan Wolf. *Multiculturalism: Examining the Politics of Recognition*. Edited by Amy Gutmann. Expanded Paperback edition. Princeton, N.J: Princeton University Press, 1994.
- Taylor, Robert S. *Reconstructing Rawls: The Kantian Foundations of Justice as Fairness*. University Park, Pa.: Pennsylvania State University Press, 2011.
- . "Self-Ownership and the Limits of Libertarianism." *Social Theory and Practice* 31, no. 4 (2005): 465–82.
- Tomasello, Michael. *Why We Cooperate*. Cambridge, Mass: The MIT Press, 2009.
- Vallentyne, Peter, and Hillel Steiner. *Left Libertarianism and Its Critics: The Contemporary Debate*. Palgrave Publishers Ltd., 2000.
- . *The Origins of Left Libertarianism: An Anthology of Historical Writings*. Palgrave Publishing Ltd., 2000.
- Waldron, Jeremy. "Nozick and Locke: Filling the Space of Rights." *Social Philosophy and Policy* 22, no. 1 (2005): 81–110.
- Wall, Steven. "Self-Ownership and Paternalism." *Journal of Political Philosophy* 17, no. 4 (2009): 399–417.
- Walzer, Michael. *Spheres Of Justice: A Defense Of Pluralism And Equality*. New York: Basic Books, 1984.
- Wenar, Leif. "Political Liberalism: An Internal Critique." *Ethics* 106, no. 1 (1995): 32–62.
- Williams, Bernard. "Liberalism and Loss." In *The Legacy of Isaiah Berlin*, edited by Mark Lilla, 91–103. New York: New York Review of Books, 2001.
- Ypi, Lea. "Self-Ownership and the State: A Democratic Critique." *Ratio* 24, no. 1 (2011): 91–106.
- Zakaras, Alex. "A Liberal Pluralism: Isaiah Berlin and John Stuart Mill." *The Review of Politics* 75, no. 01 (December 2013): 69–96.