JUSTICE IN A NON-IDEAL WORLD

Submitted by

Alexandre Gajevic Sayegh

For the Degree of Ph.D.

Department of Philosophy

University College London

'I, Alexandre Gajevic Sayegh confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Sections of Chapter 2 and Chapter 6 appeared at the following place:

> (2016) "Justice in a Non-Ideal World: The Case of Climate Change", Critical Review of International Social and Political Philosophy.

Sections of Chapter 3 and Chapter 5 appeared at the following place:

> (2016) "A Case of Non-Ideal Guidance: Assessing Tax Competition", Moral Philosophy and Politics.

This thesis has improved on these articles.

Para minha mãe. Que amo até acabar todas as casas.

« Je veux chercher si, dans l'ordre civil, il peut y avoir quelque règle d'administration légitime et sûre, en prenant les hommes tels qu'ils sont, et les lois telles qu'elles peuvent être. Je tâcherai d'allier toujours, dans cette recherche, ce que le droit permet avec ce que l'intérêt prescrit, afin que la justice et l'utilité ne se trouvent point divisées. »

Jean-Jacques Rousseau, *Du Contrat Social*, 1762

Abstract

This doctoral thesis, titled *Justice in a Non-Ideal World*, explores the methodology and the content of a non-ideal theory of justice. It puts forward a methodology for theorising about politics according to which ideal theory does not have analytical primacy over non-ideal theory. Non-ideal theory is not a matter of implementing ideal principles, but rather a field of inquiry that seeks balance between moral values and real world considerations that constrain the realisation of justice. I argue that although there is value in ideal theorising, the specific task of guiding political action is better carried by non-ideal theory. This thesis highlighted the contribution of a non-ideal approach to our theorising about politics and explored the action guiding potential of non-ideal theory by identifying its defining features. This contribution is exemplified in the contexts of tax competition, climate justice and carbon pricing.

This thesis explores ways through which the non-ideal theorist contributes to the normative weightlifting of political philosophy by gathering knowledge from other disciplines (in particular social sciences, economics, climate sciences) without presupposing ideal theory. I argue that non-ideal theorising provides not only non-normative supplementations to political theory, in terms of elements that counted as empirical and social scientific input, but also some of its core normative components. The problem this thesis sets out to fix is centrally a problem with the normative incompleteness of ideal theory. I found that non-ideal theory provides valuable insights to formulate an action guiding political theory – especially with regard to compliance, feasibility, fact-sensitivity and path dependence – in order to address issues such as tax competition, climate justice and carbon pricing, which require the collaboration of researchers across a range of disciplines.

JUSTICE IN A NON-IDEAL WORLD

List of figures

List of Abbreviations

OECD: Organisation of Economic Cooperation and Development

UNFCCC: United Nations Framework Convention on Climate Change

COP: Conference of Partners

IPCC: Intergovernmental Panel on Climate Change

WTO: World Trade Organisation

IMF: International Monetary Fund

IFF: Illicit Financial Flows

UT+AF: Unitary Taxation and Formulate Apportionment

GHG: Greenhouse Gas

CO_2e: Carbon Dioxide Equivalent

MBI: Market-Based Instruments

WCI: Western Climate Initiative

RGGI: Regional Greenhouse Gases Initiative

EU ETS: European Union Emissions Trading System

GCF: Green Climate Fund

BAU: Business as Usual

CCA: Command and Control Approaches

Acknowledgments

A thesis surely is a reflection of how one perceives the theoretical horizon of her or his discipline at the moment of writing it. But at the moment I am about to submit my thesis, I am pleased to remember that this was quite a collective effort. I write these lines thinking about the many occasions I had to discuss my work with my dear supervisors, at conferences, with so many excellent colleagues and some of my dearest friends. This thesis benefited from the amazing input of many people. Yet, none of them is responsible for any potential shortcomings.

I was supervised by Jonathan Wolff and James Wilson these past four years. Their supervision was outstanding. I remember well sitting in Jo's office, trying with all my might to explain what I was arguing for, listening to him for a moment, and suddenly understanding what I was talking about. James and Jo always gave me the freedom to write exactly what I wanted and provided me with the guidance to accomplish it. Their supervision method was always the method of motivation, never the one of pressure. I cannot thank them enough.

My examiners, Laura Valentini and Simon Caney, contributed with insightful remarks for my work to progress. I am grateful for their diligence on pressing me to clarify important aspects of my work.

I would also like to thank people with whom I had the privilege to have extensive discussions on my work. I think in particular of Florian Ostmann, Lior Erez, Sara Amighetti, John Wilesmisth, Oisin Suttle, Gabriele Badano, Aaron Maltais, Dominic Roser, John Broome, Jerome Alarie, Fabio Sayegh, Shmuel Nili, Aaron James, Lisa Fuller, Jeffrey Sachs, Maud Gauthier-Chung, Thomas Ferretti, Pierre-Yves Néron, Peter Dietsch, Ryoa Chung, Jakob Schwab, and Daniel Weinstock.

There are people who contributed to this thesis in ways that certainly transcend the text at the Department of Philosophy at UCL. Thanks among other people to Henry Clarke, Ed Nettel, Ed Lamb, Alex Geddes, Ben Fardell, Vanessa Carr, Hannah Carnegy, Andrew Knox, Ashley Shaw, Jonas Vandieken, Daniel Guillery, Josephine Salverda, Showkat Ali and Edgar Phillips.

There is group of people so close to my heart that I do not know exactly how to thank. Margaret Hampson, Lea Salje, Tim L. Short, Jerome Pedro, Karine Sawan, Tom Williams and Laura Silva, I do not know what I would have done without you. Reading over my work all these years was clearly the least of things. We needed GSDC. But most of all I think we needed each other. You made these years a life experience I will never forget. Laura, you add soul to the things touch. I feel it clearly being close to you.

There are three other people who have never measured efforts to help me be where I am in life (wherever that is, I am truly happy with it). Mom, you told me I could do anything I wanted in life, but you told me I should choose above all something that made me happy. I am not likely to forget it, I assure you. And yes, thank you for being so enthusiastic and supportive with my choice to study in philosophy. Philippe, my brother, managed to keep me in check while motivating me all the time. You are certainly the best motivational coach one could find. The fact that you have such politically informed views paired with academic rigour turned up to be more than just academic support. Endless conversation material is way better. And of course Joëlle, how can I thank you for countless hours of patience and numerous clever remarks on my work? Joëlle, you are a bright star who taught me so much during all these years and have never ceased to inspire me.

I would like to thank the Department of Philosophy at UCL for providing research students with so many occasions to share their work. I benefited greatly from presenting at the Departmental Seminar and at the WIP. I would also like to thank audiences at UCL, the University of Manchester, the University of Montreal, the Université Catholique de Lille, the European Political Science Association Annual Conference, the University of Cambridge, Yale University, the University of Warwick and the University of Nottingham for many helpful comments on key parts of this thesis.

<div align="right">A.G.S.</div>

INTRODUCTION

A number central questions in political philosophy have been formulated in response to the pressing issues of contemporary society. From questions about the relationship between the state and the church in medieval Europe, such as in Augustine's *City of God*, or about limits to the authority of the state in the early modern period, as in Hobbes' *Leviathan*, to the notions about how to organise industrial life in the 19th Century in the writings of Marx, political philosophy has been influenced by the urgent problems of the day. The fact that some of these debates have travelled centuries is an indication that some problems have not vanished, and have retained a similar form.

Political philosophy is concerned with justifying and criticizing existing and possible forms of political organisation. It is a normative discipline. It tries to establish norms about how society should be organized, who gets to rule, and who gets what. In consequence, a number of debates in political philosophy are methodological. They raise questions concerning how normative judgements about politics can be justified.

This thesis connects these two thoughts. Indeed, political philosophy evolves in a way that reflects the most pressing questions of society. Today's globalized economy raises challenges of a technical complexity never observed before, such as transnational trade regulations, global tax competition and illicit financial flows, massive scale non-citizenship and migration, climate change mitigation and climate refugees. Important works in political philosophy have recently started to address these issues. And indeed, normative judgments in political philosophy often bring to the fore queries about their underlying methodology. In *A Theory of Justice*, John Rawls was explicit about the relation between these two dimensions of his theory: although they could be read independently, the arguments in favour of his two principles of justice would find support in his methodology of the original position (consider for instance the

justification for the 'liberty principle' and the rationale behind the idea of the 'maximin').

This thesis is a quest for a methodology in political theory that is more tailored to address the pressing questions of today's world. I argue that the methodology commonly used in some of the most important theories of distributive justice in the past five decades has several important shortcomings when it faces the emerging challenges of a globalized economy. The latter method I am referring to is what has been labelled since Rawls as 'ideal theory'. I will search for a different methodology by bringing to the spotlight an idea, which I believe to be underexplored in political theory: action-guidance. I will present an original version of what has been called 'non-ideal theory' and will explore its potential. The capacity to guide action is one feature of political theorising that has led to difficulties with ideal theory.

Yet, I will argue here that it is only natural that ideal theories fail to guide action, in the sense I will specify below. That is because guiding action falls outside of its scope. It is not part of its function. Failing to guide action is not a failure of ideal theory *qua* ideal theory. It is a call for a methodology in political theorising that is conceptually equipped to carry out this task. The nature of the inquiry in this thesis is therefore meta-theoretical, but this meta-theoretical exploration will not prevent me from passing normative judgements about some of these political issues. Using climate justice, carbon pricing and tax competition as models, this thesis will point to a fruitful way to formulate principles of justice that can guide action in these complex contexts.

The ideal and non-ideal theory debate should be understood as a chapter of the traditional discussions about the goals and methods of political theorising. These discussions included the oppositions between realism and idealism, the moral and the political, the descriptive and the normative, the theoretical and the practical. By understanding the ideal and non-ideal debate in a narrow sense (say, by overly focused on a specific feature of ideal theory), we run the risk of losing sight of its very purpose. Although it tends to get very technical, the debate's quality remains attached to its capacity to describe the extent to which a theory relates to problems that concern us in the real world. By characterizing ideal theories as views radically disconnected from our

deep concerns about the political, we run the risk of misplacing the value of theorising about politics.

To avoid confusion, I take all theories of justice, ideal and non-ideal alike, to be aimed at reforming a feature of the world and to be making idealisations. All political theories aim at understanding aspects of the word that need reform. At the same time, we will see how important it is not to criticise ideal theory for what it is not. I will argue that ideal and non-ideal theories are analytic tools that should help understanding the relation between moral values and political action, in the world today. This debate is not internal to political philosophy. It is about how theories and values connect with political reality.

In other words, despite its theoretical function, I do not wish to characterise ideal theory as completely separated from real world problems. Ideal theory, as I am going to define it, refers to political theories that aim at the formulation of principles of justice, and do so by abstracting from or idealising about features of the current institutional order and institutions in particular. Ideal theories need not be 'fact-insensitive' in G.A. Cohen's terms, i.e. they need not to be completely detached from facts about human psychology or the functioning of institutions. They idealise certain features of society and institutions, such as, but not exclusively or necessarily, by assuming general compliance. I will explore in this thesis an understanding of non-ideal theory that does not presuppose ideal theory, one that carries out a different task, which proceeds from ongoing practices and the potential injustices they perpetrate. Non-ideal theories, as I define them, are concerned with addressing the practices that are the causes of injustices in the first place. Non-ideal theories aim at regulating specific political regimes, notably by addressing the non-ideal circumstances and empirical features of the world that explain, among other things, non-compliance and path-dependence. In short, ideal and non-ideal theory approach action-guidance very differently.

What is particular to ideal theory is a special attention to the definition of our values and principles independently from the functioning of actual practices and institutions and under the assumption of full compliance. Rawls's theory is not completely idealised and fact-insensitive. It is idealised in the sense it is aimed at

defining a theory of 'perfect justice' that assumed 'full compliance' of agents in society[1] and in the sense that it abstracts from the functioning of institutions, meaning it does not derive its principles from the functioning of the different practices and institutions that composed the basic structure of society. Other liberal egalitarian theories, such as Ronald Dworkin's theory of equality of resources or Richard Arneson's theory of equality of opportunity for welfare, are also ideal in the sense used in this thesis, because these theories are not aimed at understanding the specific features of the different practices that contribute to perpetuating ongoing injustices, and therefore are not constrained by these features.

Whether these theories can be applied to ongoing practices is only one aspect of the debate about the transition between ideal and non-ideal theory and it is not a question I wish to begin my inquiry with. I wish instead to find what the contribution is of a non-ideal theory that does not presuppose ideal theory, to conceptualise the central questions that should be of concern for a truly non-ideal theory and to test the practical significance of these questions. In other words, I wish to explore non-ideal theory as a conceptual space that is more closely related to the findings of social sciences, natural sciences, psychology, and economics.

The value of fundamental research lies in its quest for truths and for the laws of nature. There is no denying that there is value in fundamental research and that its path should not be dictated by imperatives of politics or the markets. Is research in philosophy fundamental in this sense? I think it is. But is it not also the case that many fields in philosophy benefit from integrating empirical findings, such as philosophy of mind and findings in neuroscience and political philosophy and findings in social sciences? I believe this is also true.

Research of the utmost importance in political philosophy has been undertaken in the past centuries of which theoretical value is its prime aim. But theoretical endeavours have been criticized for not having sufficient practical value. The response to these critics has often been that practical questions of feasibility and implementation are

---

[1] I use 'perfect justice' as short for 'perfectly just society' as coined by Rawls in *A Theory of Justice* (Rawls: 1999b).

20

beyond the scope of political philosophy. Perhaps this response is motivated by advancing political philosophy 'as a tradition'; an historical and social perspective with certain goods that constitute it (MacIntyre 1981: 222). It is worth noting that works such as Locke's *Two Treatises*, Hobbes' *Leviathan* or Mill's *On Liberty* clearly have a practical aim alongside their theoretical aim. Perhaps the pursuit of theory in political philosophy outside of an obvious practical payoff is something that became more common with the professionalization of political philosophy.

I believe that framing the debate in binary terms according to which political philosophy should be either theoretical or practical is a mistake. There is value in the philosophical endeavour of being only truth-seeking and at the same time there is value in wishing one's theory to guide action in the real world. My claim is that elucidating the connection between the two terms of this distinction is likely to cast light on their respective values. In other words, the question is not whether political philosophy should be theoretical or practical. It is to show how we can approach problems that political philosophy will need to address, regardless of the nature one may attribute to it.

The role of the theoretical endeavour, which I call ideal theory, is indeed to arrive at correct answers to normative questions. The role of the practical endeavour, non-ideal theory, is to bring the world closer to a normatively justified state of affairs (Wilson 2014, Wiens 2015).[2] The 'apparent paradox' of ideal theory, as Laura Valentini calls it, is that what is described by the ideal state of affairs does not help moving towards it (Valentini 2009). This is the central tension explored in this thesis. We should see that part of the confusion in answering this paradox stems from objections according to which a theoretical endeavour should have practical implications. Although the distinction between the theoretical and the practical aims appears to be clear, the connection between the two is much less so. A central task in this thesis is to examine the connection between these goals.

The literature on the ideal and non-theory debate, which is concerned to show whether ideal theory guides or fails to guide action has, with a few exceptions, neglected

---

[2] Wilson (2014) makes a similar distinction in the context of bioethics.

to explore the structuring features (the characteristics, the function) of a theory that guides action. The claim that ideal theory is not doomed to failure because, although it fails to guide action, it contributes to political theorising, is only one response to the paradox. The other response is to explore what kind of political theory could overcome the paradox and present the conceptual components for political theory to guide action. This is indeed the task of non-ideal theory. How to do non-ideal theory, with some exceptions, has received comparatively little scrutiny in the political theory literature.

Defining theory and ideal

It would be helpful to offer a brief definition of the notions of 'theory' and of 'ideal'. Following Zofia Stemplowska, a theory can be understood as "a systematic account of our knowledge about a given dimension of reality (where the latter is broadly construed) that satisfies the criteria of what constitutes knowledge appropriate for that dimension" (Stemplowska 2008: 323). Theories have inputs, outputs and rules from deriving the output from the input. They allow us to progress from one set of statements, to another set of statements, in order to increase our knowledge about a certain problem or phenomenon. Normative theories are specific in that they contain principles, expressing a position about one or more values. They also contain "elements that bind these principles together and account for their relevance" (*Ibid.*). These are models, with a descriptive, explanatory or predictive function, and arguments concerning the relationship between these principles. Some normative theories have even an additional output that Stemplowska calls recommendations. These are proposals for action, policies, or reforms that are able to achieve improvements, measured by the principles. The process by which we can arrive at formulating action-guiding recommendations is the focus of this thesis.

In other words, a theory can be a theory about methods or outputs, or both. As mentioned, in Rawls's *A Theory of Justice*, these two aspects are clearly developed (Rawls 1999b). The methodology is known as the 'Original Position' incorporating choice made under the 'veil of ignorance'. The outputs are composed of his three principles on the

distribution of basic liberties, equal opportunities and the difference principle. They can very well be read independently in the sense that we might fully agree with one aspect and not with the other.

The term 'ideal' is commonly used today in several different ways (Stemplowska and Swift 2012: 374 provide an overview). It can refer to the idea of perfection, as in 'she is the ideal candidate for the position', or in the sense of impossibility, as in 'his altruistic society could only function in an ideal world'. It also can be used as a standard of evaluation 'the ideal of gender equality in academic hiring is not fully realised in the UK'. For Wiens, "a political ideal is a set of directive principles that best reflects an ideal balance of basic moral and social values, that is, the balance of basic values realised at a fully just state of affairs" (Wiens 2015: 435-436). We may distinguish between two roles for political ideals, an epistemic and a practical role. The role of ideal theory is, epistemically, to make sense of the values and the ideals that should be promoted in society. And, practically, its role is to identify what actions would be required to promote and realise the values in question. The interpretation of this practical role is central to this thesis.

Another important aspect of the debate is the value we choose to focus on. This thesis focuses mainly on the notion of justice; the ideal/non-ideal theory debate has indeed emerged around discussions about justice. Discussions about justice can potentially overlap with debates about legitimacy or democracy. And the discussion about ideal theory could very well be general enough to encompass other values. I prefer to focus on the value of justice and leave potential ramifications for another time. I will discuss nonetheless how the philosopher theorises about values that are entrenched in public discussions of the civil society. Also, and relatedly, I will mention the relation between discussions about justice and discussions about democracy in Chapter 4, especially with regards of the inclusion of democratic processes in contributing to normative work about justice. The scope of this thesis is nonetheless primarily limited to discussions about justice. Finally, although I am especially concerned with problems that are global in scope, I do not believe the non-ideal theory methodology I put forward needs to be limited to problems of global justice.

A characterisation of ideal and non-ideal theory

The characterisation of ideal theory has to be carefully undertaken. To understand ideal theory as theory that makes whatever kind of idealisations it wishes would be a mistake. No ideal theory assumes a world of abundance. There is no such thing as an ideal theory that abstracts completely from the world by dismissing the inherent friction of political life.[3] As mentioned, ideal and non-ideal theories alike aim to reform some aspect of the world. There must be problems, ongoing injustices, for ideal theories to have a role in the first place.

Ideal theory refers to political theories that aim at the formulation of principles of justice under idealised assumptions, notably full-compliance. It raises questions about the justification of principles and assesses problems arising from the existence of a multiplicity of values, such as trade-offs between values and measurability. Two crucial clarifications are warranted. Firstly, ideal theories formulate principles and seek balance between moral values by abstracting from or idealising about features of current institutional settings or political regimes. Ideal theories need not take into account the specific features of the different practices that contribute to perpetuating ongoing injustices. Therefore, they are not constrained by such features. Secondly, I will argue that it is only natural that ideal theories fail to guide action, in the sense I will specify below. That is because guiding action falls outside of their scope. It is not the function of ideal theory to guide action.

Non-ideal theories have primarily a practical goal: to bring the world closer to a normatively justified state of affairs. They focus on guiding action so as to move society towards more desirable social settings. In that sense, non-ideal theory does not presuppose ideal theory. It needs values, but does not require a complex blueprint of what a perfectly just society looks like or a precise account of trade-offs between values. Values matter for ideal and non-ideal theory. What differs is the role of values across

---

[3] I take that even the most idealised ideal-theories, such as G.A. Cohen's 'fact-insensitive' theories, do not make whatever kind of idealisations they wish. We will see that what makes these theories particularly idealised is their rejection of the notion that theories of justice should be action-guiding.

these two theories, which will be clarified with greater care below (notably, Chapter 4 explains what this difference is and introduces examples that will be discussed in Part II). The contrast between ideal and non-ideal theory becomes apparent in that the latter carries out a different task, one about guiding action taking into account the current state of political regimes and institutional frameworks. Non-ideal theory responds to the critique about action guidance directed to ideal theories in a more thorough and convincing way, for it is conceptually equipped to address it. In short, non-ideal theories are part of the answer to the question 'what actions are required to make the world here and now more just?' Secondly, non-ideal theories proceed from ongoing practices and the potential injustices they perpetrate. They focus on guiding action by addressing political regimes and institutional arrangements that are themselves sources of injustices. They are concerned with addressing the practices that are the causes of injustices in the first place. This entails, for the conception of non-ideal theory that I put forward, that the normative resources employed to address a given topic should focus on an existing political regime or institutional arrangement. I summarize this aspect of non-ideal theories by the idea of adopting a 'bottom-up' approach, which I detail below. Thirdly, non-ideal theories are informed by social scientific data: non-ideal theory defines a conceptual space that is more closely related to the findings of social sciences, natural sciences, psychology, and economics. They do not abstract from the states of affairs and processes that are themselves sources of injustice, for these states of affairs contain important aspects of the necessary information to formulate action-guiding principles. Non-ideal theories aim at regulating specific practices and institutions and at addressing the non-ideal circumstances, and empirical facts of the world, that explain, among other things, non-compliance and path-dependence.

My objective is less to suggest that there is a strong opposition between the two types of theory, than to develop aspects of political theorising that I believe to have fallen off the radar and to provide a conceptual cartography of what I call non-ideal theory. This working distinction, which will be examined further in Chapter 2, links ideal theory to a specific theoretical endeavour with a defined function, where its various features (such as full-compliance, perfect justice, favourable conditions) can be observed to a

greater or lesser extent in contemporary works of political philosophy. Non-ideal theory involves a kind of normative work that is more specifically concerned with action-guidance, in a way that connects political philosophy with the work done in social sciences, economics and natural sciences – a kind of theorising that deserves more attention.

It is worth mentioning here that my own definition will only be examined further in Chapter 2. In Chapter 1, I examine central features of Rawls' definition of the two notions. There are two reasons for this. Firstly, the contribution of my definition will be best understood after the central problem of ideal theory has been exposed. Chapter 1 and the first sections of Chapter 2 are charged with the task of exploring the contributions and limits of ideal theory as understood in the literature. Secondly, my definition aims to open the door to a thorough exploration of what non-ideal theory is. It is thus helpful to have examined the notion of ideal theory, as commonly understood, before turning to the in-depth examination of non-ideal theory.

This thesis is not a radical critique of ideal theory. It is an exploration of an action-guiding non-ideal theory. This research exposes what I believe to be the contributions and limitations of ideal theorising. By doing so, it opens the door to the study of a non-ideal theory that carries on the task that is beyond the scope of ideal theorising, namely guiding action. Non-ideal theories are capable of guiding action in a sense ideal theories cannot, for they focus on the internal functioning of unjust practices and pose a normative judgement based on the features of the practice that explain, among other things, non-compliance and path-dependence.

I want to provide a roadmap that explains what it means not to abstract from the states of affairs and processes that cause injustice and how action-guiding principles can be formulated in this context. It is crucial for political theorising to understand whether and how we can bridge the gap between theory and real world circumstances, and what the role is for non-ideal theory in this dynamic between theory and practice. Getting a better grasp of the relation between ideal and non-ideal theory is no simple task.

This thesis examines several aspects of this debate. Some elements are more philosophical in nature; some are more closely connected to social sciences, economics

and climate science. Whether the political philosopher has a duty to have an impact on real world politics is a question to some extent beyond the scope of this research. Nonetheless, this thesis wishes provide insights about how the political philosopher could do this better, building on works that already do it and insisting on potential work that could do it.

The argumentative structure of this thesis

I will argue for two central theses. The first I call the 'Reflective Integration Thesis'. It states that: *if we wish to formulate principles of justice that can guide action in non-ideal circumstances, we need to integrate ideal and non-ideal theory, and the way to integrate ideal and non-ideal theory is by seeking reflective equilibrium between these levels.* In other words, a theory of justice can only become action guiding if it integrates the appropriate elements from ideal and non-ideal theory.[4]

The second is what I call the 'Non-Ideal Theory Thesis'. This second thesis is composed of four claims: (i) ideal theory is not analytically prior to non-ideal theory, (ii) theories of justice – notably theories of global justice and climate justice – are best advanced as pluralist theories composed of bottom-up contributions, meaning that different sets of principles regulate different practices, (iii) determining our duties of justice in a non-ideal world is not a question of non-compliance[5] with ideal principles, but rather an inquiry about compliance with non-ideal principles, and (iv) whether there is path-dependence between ideal principles and non-ideal reforms can only be determined through non-ideal theorising. Claims (i) and (ii) relate to the *methodology* of non-ideal theory. Claim (iii) relates to the *content* of non-ideal theory. Claim (iv) concerns the relation between ideal and non-ideal theory. I will explore these four claims by distinguishing between four components of action-guidance in a non-ideal theory of

---

[4] I provide an interpretation of the notion of reflective equilibrium in Chapters 2 and 4 (especially in section 2.6 and 4.4), which I will label reflective integration.
[5] I use non-compliance and partial compliance interchangeably.

institutional design: compliance, feasibility, fact-sensitivity and path-dependence[6]. It is notably through these components that I argue for the four claims that compose the 'Non-Ideal Theory Thesis'.

In short, the contribution of this thesis is to demonstrate how non-ideal theory responds to the paradox of ideal theory, by responding to more demanding action guidance considerations.[7] As I will show, this question is not simply about the supplementation of ideal theory by facts and or social sciences, but about the normative incompleteness of ideal theory. The latter aspect is one core aspect of the ideal and non-ideal theory debate, which has been underexplored. This normative completeness and the supplementation of political theory by facts and social science is accomplished at the level non-ideal theory. This thesis aims to highlight the contribution of a non-ideal approach to our theorising about politics and to explore the action guiding potential of non-ideal theory by identifying its defining features.

More precisely, the contributions of this thesis are: to show how principles of justice are formulated in light of non-ideal circumstances and empirical features of the world; to demonstrate that a non-ideal theory methodology is well-suited to address problems of justice; to provide an account of the elements that should structure the content of a non-ideal theory; and to exemplify how they help formulating an action-guiding theory in the contexts of tax competition, climate justice and carbon pricing. This inquiry should allow the clarification of issues that we wish a non-ideal theory to be explicit about, such as whether a non-ideal theory enables the assessment of conflicts between values, whether it points to how the non-ideal circumstances that influence the compliance of agents inform the formulation of regulations, whether it clarifies which feasibility constraints might relax in the future in order to know how this affects the appropriate timeframes for regulation and the path-dependence of future action. The action-guiding capacity of a non-ideal conception of justice can be expressed by these desiderata.

---

[6] I provide a thorough definition of the notion of path-dependence in section 1.6 and a more detailed analysis in Chapters 3 and 4.

[7] I provide a definition of the notion of action-guidance in sections 1.6, 2.3 and 2.7.

Finally, in order to understand the formidable importance of the ideal/non-ideal theory debate, we should see below how it captures a number of fundamental disputes in political theory, notably about the methodology of theories of justice (should we proceed from high principles or ongoing social problems?); the unitary aspect of theories of justice (can bottom-up approaches contribute to debate about justice?); the function of political theorising (should political theories be action guiding and, if so, how?); and the division of labour in humanities and social sciences (with philosophers in charge of formulating principles and identifying values, and social scientist in charge of implementing them). This research is motivated by the need to seek a better understanding of the relation between political action and the judgements political philosophers should be in a position to make.

The above and the following questions will be given particular attention. How can we distinguish between ideal and non-ideal theory? Does ideal theory have analytical priority over non-ideal theory? What is the value of non-ideal theorising? Should a political theory be sensitive to facts? Is there path-dependence between ideal and non-ideal theory? What does it mean to bridge the gap between ideal and non-ideal theory? What are the characteristics of an action-guiding theory of justice?

Overview of the chapters

This thesis is divided in two parts. Part I – composed of Chapters 1, 2, 3 and 4 – focuses on the meta-theoretical aspects of the debate, notably around the notions of reflective integration and action-guidance. Part II – Chapters 5, 6 and 7 – uses three case studies (one related to tax competition two related to climate justice) to provide further support to the meta-theoretical claims of Part I and open to research about justice in practice, meaning bottom-up approaches to justice focussing on targeted practices. That is, the structural coherence of this thesis is partly given by the relation between Parts I and II. The primary objective of the 'Reflective Integration Thesis' and the 'Non-Ideal Theory Thesis' presented in Part I is meta-theoretical: it is about elucidating the relationship between ideal and non-ideal theory, by focusing on the pivotal role of action

guidance. Yet, this meta-theoretical objective is peculiar in one important respect: an empirical analysis is required to derive the conclusion that ideal principles are affected by non-ideal circumstances and empirical features of the world. The ultimate demonstration of the two theses will only become evident in Part II of the thesis.

After providing a historical summary of the debate, Chapter 1 examines questions about the value and limits of ideal theory. It will explore the central role of idealisations in ideal theorising. It will also give particular attention to two alleged roles of ideal theory, that is: to provide the target we should aim at when doing non-ideal theory and to provide the measure according to which we can know to what extent a given situation deviates from what it should be. I call these the functions of 'target' and 'measure'. This chapter will expose some of the central contributions of ideal theory to political theorising and will underscore its limits with regard to action guidance.

Chapter 2 focuses more on the contrasts and relation between ideal and non-ideal theory. It puts forward one way to overcome the 'paradox' of ideal theory. It begins by exploring the 'paradox', according to which the capacity to provide guidance (or 'capacity of guidance' for short) is an important feature of most normative theories, but ideal principles of justice are not well suited to guide action in non-ideal circumstances. It delineates the extent to which ideal theory can respond to the paradox. It then opens the door to how a non-ideal theorist could respond to it. The first, and more general, element of this response will be given by what I called above the 'Reflective Integration Thesis', thereby suggesting that non-ideal theory is key to overcoming the paradox.

Chapter 3 focuses on the methodology and content of a non-ideal theory of global justice. I will argue in this chapter for the four claims that compose my second main thesis – the 'Non-Ideal Theory Thesis' – claims (i), (ii), (iii) and (iv) above. The *methodology* of this theory posits that principles of justice are formulated according to the practice they are intended to regulate. With regard to the *content* of principles, I reject the notion that non-ideal theory is applied ideal theory. I provide instead an account of the main features of a conception of justice for a non-ideal world around the notions of: fact-sensitivity, political feasibility, path-dependence, and compliance. The examination of these notions will support my argument for the four claims.

Chapter 4 expands on the idea of action-guidance, thereby providing further support for the 'reflective integration thesis' and the 'non-ideal theory thesis'. It begins by offering a discussion of theories of second-best, as they relate to the ideal non-ideal theory debate. One of the central ways this chapter will expand on the idea of action guidance will be by exploring in detail the notions of path-dependence and of dynamic duties. It will also provide a more detailed reflection about what the role for the political philosopher at the gates of social sciences is.

Opening part II, Chapter 5 targets tax competition. It aims at showing how tax competition poses a problem to the theories of global justice today and how non-ideal theorising helps us to respond to it. It will distinguish between the various forms of tax competition and identify the principles that could regulate them. This chapter stresses how global justice can be fostered by theories that proceed from ongoing unjust global practices and formulate principles that are sensitive to the internal structure of these practices. This should provide further insights about how to think the notion of 'path-dependence' between an ideal theory of justice and a non-ideal theory of institutional design.

Chapter 6 turns to climate justice. My contribution is distinct from the main contributions to the topic in the literature. It is not to show whether principles of climate justice are justified morally but rather to show, should we choose to include them, what kind of real world challenges we face when making sense of them. The objective of this analysis is to know whether inputs from facts and other disciplines (climate science, economics, social sciences) provide normative supplementations to philosophical inquiry. I will argue that the empirical examination will provide useful inputs for the normative work. I will insist here on how the 'reflective integration thesis' is an appropriate methodology for formulating action-guiding principles.

Chapter 7 focuses on the distributive aspects of market-based approaches to climate change mitigation. This is the most empirical chapter of the thesis. It raises the following question: how can market-based instruments for climate change mitigation (MBIs) respond to requirements of justice? This chapter aims to explore how principles of climate justice are realised in practice. Its contribution is twofold. Firstly, it is to

provide a sound normative foundation for carbon pricing mechanisms around the notions of the 'duty not to harm' and a 'right to energy'. Secondly, it is to identify the normative elements from theories of climate justice that should constrain the design of MBIs so that these become instruments of justice.

I conclude the thesis by explaining how Part II takes forward and completes the analysis of Part I: the case studies on climate justice, tax competition and carbon pricing provide support to the 'Reflective Integration Thesis' and the 'Non-Ideal Theory Thesis'. I return to the idea that the role for the political philosopher who engages with non-ideal theory is to find not only non-normative supplementations, but normative inputs to political theory.

PART I

Under conditions of abundance and generosity, "the cautious, jealous virtue of justice would never once have been dreamed of... Justice, in that case, being totally useless, would be an idle ceremonial, and could never possibly have place in the catalogue of virtues... it seems evident, that the use of justice would, in this case, be suspended by such an extensive benevolence, nor would the divisions and barriers of property and obligation have ever been thought of."

<div align="right">Hume, A Treatise of Human Nature, III.1.</div>

1. ON THE ROLES AND LIMITS OF IDEAL THEORY

1.1 From Plato's Republic

In Book 5 of the *Republic*, Glaucon replies to Socrates:

Take it that I agree that there would be all these things and countless others if this regime that should come into being, and don't talk anymore about it; rather, let's now only try to persuade ourselves that it is possible and how it is possible, dismissing all the rest. (471e)

Socrates presented his account of justice earlier in the dialogue, although without fully revealing it. Glaucon and Socrates "got to this point while seeking what justice and injustice are like" (Plato: 472b). They had so far relied on the ideas of compliance of citizens and of motivation to comply with the law (471d). Glaucon was ready to agree with Socrates, for that was not the real issue Socrates should be addressing. The biggest and most difficult 'wave' assaulting Socrates' argument was to know whether his account of just city is possible, and how. Glaucon notices the sharp contrast between a hypothetical acceptance of an ideal account of justice and the questions of feasibility that arise when seeking to transform a society such that it conforms with this account. At first, it seems that Socrates was claiming that an ideal would remain valid even if it could not be realised. He indeed claims that a role of his ideal city is to judge actual cities on how well they approximate the ideal (472b-473b). In this passage of the *Republic*, Socrates – believing that his ideal city could be realised – rather shifts the focus of his argument in order to show that a theory wishing to provide guidance ought to compare ideal and non-ideal cities (Kraut 1992, Ypi 2010: 538). He first concentrated on the role of an ideal of justice when thinking about the virtues of the polity. At one point he began to think about the agents charged to carry out these duties and on the feasibility constraints the theory should consider. Whether Socrates' arguments in favour of the

practical role of his account of ideal justice are persuasive largely depends on the role of motivation as it relates to the knowledge of justice, which is an eminently complex problem in Plato's *Republic* (Kraut 1992, Brown 2011). Nonetheless, it seems clear that for Plato the philosopher's justice should be taken as a useful paradigm that can be approached by the non-philosophers.

In a similar vein, Rousseau opens Book I of *The Social Contract* with the famous quote with which I opened this thesis: "I mean to inquire if, in the civil order, there can be any sure and legitimate rule of administration, men being taken as they are and laws as they might be" (Rousseau 1762). For Rousseau, the reform of any social institutions should begin with people as they are and laws as they could be. Our goals in society should reflect our best accounts of what an ideally just society should look like. Yet our strategies for achieving it must take into account how society is, with its non-ideal agents and non-ideal political structures.

Following Plato and Rousseau, Lea Ypi observes that a theory of justice has two stages: the formulation of ideal principles that will shape the institutions under the assumption of full compliance, and the assignment of responsibilities to agents that will realise the project. She claims that contemporary theories, chiefly Rawlsian theories, follow a similar pattern. One stage of these theories is devoted to defining the standard of social interaction under favourable conditions and full compliance. Then, they leave this ideal level of reasoning to deal with problems of partial compliance, lack of motivation and historical failures. For Plato, everyone would endorse just principles after an appropriate assessment of the circumstances of justice (*Republic*, 472), (this perhaps contributes to explaining why full compliance seems to be an inextricable assumption of ideal theory). This would support the methodology according to which first we need to know what is justice by assuming that everyone would be motivated to promote it, only after which can we start to think about the non-ideal relations that could favour its realisation.

Rawls is even more adamant than Plato with regard to the analytical priority of 'ideal theory'. He says:

The reason for beginning with ideal theory is that it provides, I believe, the only basis for a systematic grasp of these more pressing problems [i.e. of partial compliance] (Rawls 1999b: 8-9).

In other words, not only there are two stages of reasoning, but the formulation of regulative principles necessary to guide the transformation of political institutions comes first, their implementation comes after (*Ibid.* 216).

Even if we admit that the ideal principles are the standards by which we will evaluate and change non-ideal circumstances, we will still need to know by whom and how these changes will take place. Concerning agents, Socrates would have appointed philosophers, Rousseau the legislators and Kant the moral politicians to rule. The realisation of an ideal needs not to assume that everyone will effectively comply. Socrates replies to Glaucon that if we discovered how a city might be governed nearly as we proposed, we will have discovered the possibility sought by Glaucon (472).

Assuming for the moment that this structuring of political theories in two stages is plausible, we may note one other commonality in the views of Rawls and Plato. It concerns the stage of the theory at which facts intervene and gain normative relevance. In the construction of their theories, we observe that Plato and Rawls make assumptions, or idealisations. Both assume full compliance. Plato also assumed motivation to obey the laws[8] while Rawls assumes favourable circumstances. But that does not mean that they rule out the possibility that facts could have a role to play in political theorising. As Ypi points out, specifying the stage at which facts come into play, about institutions for instance, is a weaker thesis than one that denies the role of factual assumptions or asserts the validity of fact-insensitive principles (Ypi 2010: 539). For instance, Cohen denies the role of facts in the formulation of first-order, fundamental, principles. Yet, he admits a role for facts in second-order principles he names 'rules of regulation'(Cohen 2003).

The division of our theorising about politics in a more ideal dimension and a more practical dimension is thus not uncommon. It has sources in prominent figures of ancient, Enlightenment and contemporary political philosophy. It also echoes, as

---

[8] With regard to motivation, Rawls also provides argument about what he calls the 'strains of commitment'. I return to this below.

mentioned in the introduction, important dichotomies in our common vocabulary addressing social and political issues, such as realism and idealism, the moral and the political, the descriptive and the normative, the theoretical and the practical. This debate is very much alive today. Explicit defences and critiques of ideal theory are numerous in contemporary political theory.[9] Implicit endorsements of one or other methodology or approach are even more numerous. I will talk in terms of ideal and non-ideal theory, as these terms will become clearer in the course of this thesis, in a debate that appears at times to become increasingly murkier.

We should see that, depending on how we understand this division of theories of justice into two stages, we face considerable problems. I will explore these problems in the next sections and chapters. I mention two here. Firstly, it is not clear that the distinction between one ideal level of formulation and one non-ideal level of implementation is defensible. Secondly, it is unclear that there is a clear-cut sequence between these levels. I challenge the view according to which ideal theory has priority.

Rawls was the first to formulate the debate in terms of 'ideal' and 'non-ideal' theory. He saw this as a good way to characterise the relation between political philosophical theory and political practice. And much of the debate proceeds from (and questions) Rawls' definitions, assumptions and questions. It is thus important to review his characterisation of the problem.

1.2 Rawls and Ideal Theory

John Rawls described his theory of justice as an ideal theory. From Rawls, in *A Theory of Justice*, we get a detailed exposition of the function and features of ideal theory. Many of these functions and features reflect the intuitions of the past about ideal theorising. And these different aspects of ideal theory prompted many criticisms. The defences and objections to ideal theory as coined by Rawls provide a fruitful ground to

---

[9] I use political philosophy and political theory interchangeably.

examine different important aspects of the question. The coverage of this rich literature should open the door to the examination of underexplored aspects of non-ideal theorising. And, more importantly, a wider view of this debate also reveals that a new way of approaching political theorising has been neglected in the literature, or at least has not received a systematic treatment. It is paramount to underline the difference in the nature of these various components of ideal theory, i.e. its features, its functions, and its spirit.

We should begin to review Rawls' treatment of ideal theory by examining two idealised assumptions that are central to the notion. First, ideal theory defines the principles for an idealised society where almost everyone complies with the principles (Rawls 1999b: 7-8, 2001: 13). In contrast, non-ideal theory assumes that not everyone complies with principles and deals with this by addressing issues such as civil disobedience (1999b: 8). The extent of non-compliance can vary depending as much on the problem as on the responses to it. The second distinction based on a feature of ideal theory stems from the ideal theory assumption of a well-ordered society under favourable circumstances (1999b: 216). Favourable circumstances indicate the conditions that make a constitutional democracy possible (Rawls 2001: 101). A well-ordered society is a constitutional democracy. The favourable conditions include the social and economic means that are necessary to sustain this regime, such as education and a sufficient level of economic development. It is worth noting that while Rawls' idea of favourable circumstances can be observed in a good number of contemporary societies, full compliance is a feature of almost no society (Stemplowska and Swift 2012). They remain the two central *features* of ideal theories for Rawls.

Another feature of Rawls' ideal theory is the notion that ideal theory is a theory of 'perfect justice'. For Rawls, the full-compliance assumption is necessary for us to find out what perfect justice requires. That is, the features of full-compliance and favourable conditions are meant to lead us not to any theory of justice, but to a state of perfect justice. The exact definition of the idea of perfect justice seems in fact less important than the role it accomplishes (I will return to the definition, and to the full-compliance assumption, in section 1.6 below). The argument begins with the idea that we cannot

know what perfect justice is if we have to think about whether agents will or will not actually comply with its principles. If we factor in non-compliance, the principles will not hold. But the real question is: why do we have to know what perfect justice requires (Stemplowska and Swift 2012: 376)?

The notion of perfect justice is in fact associated with a central *function* of ideal theory: perfect justice provides the only basis to systematically grasp the problems that non-ideal theory seeks to investigate (Rawls 1999b). This claim about the 'analytic primacy of ideal theory' is a central object of study of this thesis. Ideal theory must be undertaken prior to non-ideal theory, so the argument goes, for it is the only way to achieve the basis from which to systematically grasp the problems of non-ideal theory. In fact, it is paramount to distinguish between two functions of ideal theory which constitute the claim for the analytic primacy of ideal theory. Setting the basis to systematically grasp the problems of ideal theory is done in two ways. First, ideal theory provides a target for non-ideal theory: "until the ideal is identified ... non-ideal theory lacks an objective, an aim, by reference to which its queries can be answered" (1999b: 90). The function of the target is explicit in Rawls' treatment of non-ideal theory in the *Law of Peoples*: "Non-ideal theory asks how this long-term goal [of ideal theory] might be achieved, or worked toward, usually in gradual steps. It looks for courses of action that are morally permissible and politically possible as well as likely to be effective" (1999a: 89). Call this ideal theory 'as target'.

Second, ideal theory provides the measure to assess injustices in the real world "identified by the extent of the deviation from perfect justice" (1999b: 216). The function of measure is explicit in Rawls' explanation of the connection of the two theories:

> Viewing the theory of justice as a whole, the ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason (1999b: 246).

It provides a measure in that we can only know what situation is unjust, and how unjust it is, by determining how it departs from the pre-identified ideal (although Rawls admits

that part of this exercise is left to intuition).[10] This is an explicit claim to the effect that ideal theory has an action guiding function. Call this ideal theory 'as measure'.

Finally, besides its *features* and the *functions*, there is something particular to Rawls' understanding of ideal theory, which qualifies the *spirit* in which his ideal theory is conceived. For Rawls, ideal theory must remain a 'realistic utopia' (1999b: 11–12). It must remain within the realm of what is realistically practicable (2001: 13). This can be clarified further on the basis of an account of what is expected of agents, which comes from Rawls' take on moral psychology (Rawls 1999b: 126): theories that are too difficult to expect agents to comply with should be ruled out. Agents should be able to comply with principles given merely their sense of justice. That is, no agent is required to be a saint or a moral hero (1999b: 419). In a similar vein, favourable circumstances do not imply a society of abundance, but rather 'moderate scarcity'.

In a nutshell, for Rawls, ideal theory defines principles of perfect justice for an idealised society characterised by full compliance and favourable conditions, but must remain a realistic utopia. This exercise has to be undertaken prior to non-ideal theorising in order to provide the latter with its target and the measure through which injustices can be assessed. Non-ideal theory deals with more urgent and pressing questions and asks what justice demands of us in non-ideal circumstances. For Rawls, ideal and non-ideal theories aim at identifying the principles that should regulate societies' basic institutions. These are the main functions and characteristics of ideal theory according to Rawls and many commentators. Objections regarding the relevance of these features and functions have been raised. These will be discussed in the rest of this chapter and in the next chapters.

Two minor clarifications are warranted before concluding this section. We may rightly ask whether it is Rawls' principles that are ideal, or the method he uses to obtain these principles which is idealised (Stemplowska and Swift 2012). It seems clear that a method that consists in utilising idealisations is idealised. The assumptions of full

---

[10] Stemplowska and Swift (2012: 477) note an interesting point to which I will return: "Importantly, Rawls is clear that the requirements of ideal theory are never to be followed without regard for the actual circumstances—as opposed to the conditions assumed by ideal theory—and that the requirements should still inform— while not rigidly prescribing—what is to be done in those circumstances".

compliance and favourable conditions are idealisations. This is even more clear in that the principles are chosen from the conditions set forth by the 'original position' under the 'veil of ignorance'. Claiming that he formulates principles of perfect justice also denotes that the principles as such are ideal 'as target' and 'as measure'.

More importantly, the relation between the epistemic and the practical nature of ideal theory will have to be examined. We may agree that Rawls' ideal theory is a theory that aims at providing the values and the corresponding duties that apply to the political sphere. But again, should this be understood as if the goal of ideal theory is to inform us about the necessary actions to realise these values or rather as if this is only a question about knowledge of the values in question? This is a central question that will structure some of the discussion in the first chapters of this thesis.

1.3 Idealisations

Prominent discussions in political philosophy have been centred around questions such as "what is the right theory of justice?", "what is the right 'currency', 'metric' or 'focal variable' of justice?", and "what would a perfectly just society look like?" (Nozick 1974, Dworkin 1981, Arneson 1989, Rawls 1999b are examples). These theories are ideal theories, at least with respect to their outputs: they put forward one framework that presents what should be distributed and what an ideal distribution should look like.[11] Ideal theories are very present amongst contemporary liberal egalitarian theories. But these theories are also ideal because of the idealisations that they make when justifying these principles and arriving at the description of this just society. Robert Nozick's conditions for the basic appropriation and distribution of libertarian rights or Ronald Dworkin's auction system for the distribution of resources are developed under idealized conditions, which should allow to model an ideal distribution. Whatever technical limitations in society to effect this distribution or lack of motivation of the citizens of

---

[11] Although he insists his account is not patterned, Nozick's account fits in this model for he presents an account of the ideal distribution of liberties.

that society to follow the chosen principles are assumed away at this level of theorising (Robeyns 2008, Valentini 2009). That is, ideal theories rely on idealisations.

It is not disputed that ideal theories rely on idealisations to obtain and explore their perfectly just society. Although, features such as full compliance, or even the very notion of perfect justice, have been subject to virulent attacks, most attacks do not challenge the fact that these should be two components of ideal theory. They rather concede that these are features of ideal theory and oppose ideal theorising in general precisely because it is based on these terms. They question whether normative theories should rely on idealisations in the first place.

Besides questioning whether political theories should rely on idealisations, and assuming that to a certain extent that they should, one may rightly inquire about the relation between the use of idealisations and their role in the quest for the perfectly just society. Of course, if idealisations are a constitutive element of ideal theory, they will play a role in the definition of the just society. But what is this role and how should it be performed? This is also of central interest. These questions are the objects of the next sections.

In order to answer these questions, it is sensible to define the notion of idealisation. Onora O'Neill makes a distinction between idealisation and abstraction (O'Neill 1996: 38-44). Abstraction means bracketing the complexities of some problems without making false assumptions. Idealisation means making false assumptions. For O'Neill, there is a difference between bracketing and denying predicates that are true of the subject under discussion. For instance, consider the case in which a theory wishes to make people responsible for their individual choices. If the theory makes people responsible for their choices because they can choose wisely, it is making an idealisation in O'Neill's definition. This assumption alters reality. People do not always choose wisely. Contrastingly, making people responsible for their choices because this can have good outcomes for society is an abstraction. It only brackets the fact that some people might not choose wisely. The difference between the two cases is that in the latter, abstracting from a predicate does not depend on the predicate holding. If some people do not choose wisely in the second case, it does not undermine the thought that the

outcome of making people responsible for their choices can be positive. It could still potentially be demonstrated that making people responsible for their choices has positive outcomes. The former claim alters reality and, for O'Neill, this prevents it from being applicable to the human case (1996: 41).

O'Neill's distinction contributes to the discussion about different kinds of idealisations in the ideal and non-ideal theory debate. In this literature, a similar distinction is made in terms of good and bad idealisations. Although I will not use O'Neill's terminology, I will show that it is indeed important to be clear on the relation between idealisations and the subject they idealise about. Ideal theories, those that are of concern here, rely on idealisations, but not all idealisations have the same implications.

To make an idealisation is to make an assumption about a feature of a theory that we know to be different from reality. This is done with the objective of simplifying a feature of human life or society for the sake of theory. The role of idealisations is to make theorising about a subject manageable. We reduce the number of parameters to solve one particular problem; just like in mathematics we can only solve a set of equations if the number of unknown variables is equal or inferior to the number of equations. In this sense, the Rawlsian ideal theory is designed under idealised (false) assumptions: full-compliance cannot be observed in society, but we may want to assume that if we formulate just principles, the population will comply with them in order to set the standard. We will thus obtain a standard of justice with which we can reasonably expect compliance. Another example is to assume, like some economic theories do, that all agents are fully rational when ranking their individual preferences. Both assumptions are idealisations. They seem to turn people into better persons than they actually are (Robeyns 2008: 353).

To formulate a theory under idealised assumptions has the consequence of making social reality appear 'simpler and better than it is'. This, then, is a broad definition of ideal theory: a theory constructed under idealized assumptions that makes society simpler and better than it is (Holtman 1999, Valentini 2009: 332). As seen, Rawls defines ideal theory in part by means of two specific idealisations: full compliance and favourable circumstances (Rawls 1999b: 4-8). We also know Rawls simplifies his

theoretical framework even further by regarding that societies as self-contained. Dworkin also develops his principle of equality of resources in 'the *ideal ideal* word' (Dworkin 2000: 172). At this level, he puts aside all feasibility and motivation problems.

Also, Rawls' hypothetical agreement is an idealisation (1999b: 118-137). The conditions under which the principles are derived (e.g. ignorance of our position in society) are not real. This is also the case for Dworkin's 'ideal ideal' world of fantasy, of auctions and insurances (2000: 67-68, 172). Dworkin's ideal theory leaves aside a great number of particularities about the actual state of political institutions and characteristics of human behaviour. However, and this is crucial, the fact he assumes away inequalities (gender and race based for instance) does not mean his theory cannot condemn them as unjust.[12] This is a very important nuance in general and one that shields ideal theory from unjustified criticisms. This point raises the question about good and bad ideal theories, which are the topics of the next two sections. So far, we have only shown that some of the best political theories we know have indeed a central ideal component.

1.4 The Value of Ideal Theory

The characterisation of ideal theory offered in the introduction and the previous section is sufficient for us to understand some of its contributions to political theorising.[13] It is important to be clear on some of its contributions before starting to test

---

[12] It is worth noticing nonetheless that some kinds of systemic injustices are downplayed by Dwor克's account, e.g. theorising disability as a negative endowment that needs to be compensated overlooks the social construction of disability.

[13] The definition of ideal and non-ideal provided so far, taken mainly from Rawls, is sufficient for us to understand its general contributions and limits. I will provide my own characterisation of these notions in Chapter 2. As mentioned in the introduction, I prefer to articulate my own definition after exposing the contributions and limits of ideal theory in Chapter 1, as commonly understood in the literature. That is because, firstly, the contribution of my definition will be best understood once we have a better notion of what the central problem of ideal theory is. Secondly, considering that the distinctive contribution of ideal theory is not my central focus and that my definition incorporates these contributions, it is only helpful to have a sufficient understanding of what the problem with ideal theory is. My focus is on how non-ideal

its limits. There are contributions of ideal theory that are independent of it being a response to a practical problem, while others relate to how ideal theory can respond to practical problems. The first category is very important, although the second appears more central to the purpose of this thesis (for I do not exclude that contributions of the first kind could play a role in the dynamics of reflective integration between ideal and non-ideal theory, which will be presented in the next chapter). Contributions (or limitations) of the second kind will allow us to structure the debate about action-guidance. This section focuses on contributions of the first kind.[14] Those of the second kind below are more clearly in line with previous discussions in this chapter so far, but will be discussed more carefully in the followings sections and chapters.

One apparent contribution of ideal theory bears on the claim that institutions are powerful means to 'shape the social ethos' (Weinstock 2006, Valentini 2009, Weinstock 2009). Note, for instance, that institutions ensure women equal standing in their capacity as citizens, through different context sensitive measures. Changes in social ethos contribute to changes in private ethos.[15] We may shape the social ethos by contributing to the implementation of this theory of justice, if we have an ideal theory underpinning the design of social institutions. Valentini points to Susan Moller Okin's argument, who while arguing that theories of justice fail to see problems of justice that apply to the family, claims that there are sufficient resources in Rawls' theory of justice to deal with gender issues (Okin 1989: 89-109). Okin claimed that the social expectation that women should assume more responsibility in the care of a child is inconsistent with real equality of opportunity. In *Justice as Fairness, A Restatement*, Rawls revised his theory of justice such that family becomes part of the basic structure. Social institutions

---

theory responds to the problem of ideal theory. It will become clear how my definition helps me exposing a conception of non-ideal theory which responds to the 'paradox of ideal theory'.

[14] These advantages need not to be unique to ideal theory. We can perhaps achieve these purposes through other means, such as narratives or non-ideal theory. I do not mean to argue that ideal is indispensable as I will clarify in the end of this section.

[15] A clarification here is that, in the example of the family, the problem of ideal theory was less one of action guidance than one of scope. The theory is useful, but of too restricted scope, problems of the family were beyond the reach of ideal liberal theories. This is a challenge to liberal egalitarians. The family example was indeed one about scope, between public and private domains. But this is not what I am focussing on. I am focussing on public questions that have particular complexities.

can help foster a culture of respect between genders (Valentini 2009). And ideal theories can help design social institutions. The Rawlsian principles act as constraints on the forms institutions may take. Valentini (2009: 344) points out that, if ways of arranging the family would infringe a citizen's equal rights and opportunities, corrections would be mandated by the theory. The ideal theory would make recommendations for the social institutions to shape the social ethos so as to correct an injustice.

Take the dramatic example of marital rape. In England and Wales, marital rape only became a crime in 1991. In the US it was not until 1993 that marital rape was criminalised in all 50 states. And in many countries it is still not, or it was made illegal but it is still tolerated. This change in the rule of law was necessary to lessen injustices in the private sphere. And such a change was made possible by decades, and even centuries, of arguments in courts and the public sphere against this noxious practice.[16] Think for instance of John Stuart Mill and Harriet Taylor's arguments against the subordination of woman and for an equality between genders around 1870 (Mill 1869).[17]

To say that ideal theory can influence the social ethos is not to say that it provides a blueprint of what actions should be undertaken and it is not to say that what the basic structure 'should do' about this type of inequality can be determined without reference to the special social context (Cohen 1997, Pogge 2000 debate this latter question). I will return to these two questions in detail. We can nonetheless deduce from this reasoning that one of the strengths of ideal theory is its capacity to shape the social ethos, understood as a distinct matter to guiding action (in the more robust sense, discussed below). Although other factors may have a powerful impact on the social ethos (Rorty 1998: 167-185), it does not prevent ideal theories from also being counted as a factor. Note that nothing has been said so far about *how* ideal or non-ideal theories come to influence institutions.

A second potential contribution of ideal theory is that ideal theories allow us to question the *status quo*. I do not think that this contribution is grounded on the argument presented above about the priority of ideal theory, according to which a good

---

[16] Some of these arguments are ideal theory and some of the arguments are just everyday moral reasoning.
[17] I am not claiming that argument is pre-political, in Bernard Williams' terms.

reason to appeal to idealisations is that we want to formulate the principles and institutions that will regulate a just society; and for that we need to idealise features of human life and society in order to describe and model this just society. Or, in short, we need to know what this just society would look like when we theorise about politics. There is a difference between opening a critical space to criticise the status quo and an epistemic inquiry about what justice is. The specific description of what justice is does not follow necessarily from the critical assessment of the ongoing problems. Idealisations are important for providing theories with the critical capacity to keep the necessary distance to assess the status quo (Valentini 2009: 339), but this does not entail that we need to model a perfectly just society.

That is, we might wish to know what is wrong about domination, oppression and coercion, and ideal theory could provide us with tools to understand these problems. Ideal theory can open the space between these problems and the moral assessment we can make of them, regardless of whether it provides a theory of perfect justice. In other words, there is perhaps a more important contribution made by ideal theorising to political theory than simply setting out what justice is. This is compatible with theories assuming full-compliance. Conceptions of justice that wish to formulate principles while assuming full-compliance should of course still do it. These principles will cast light on the space existing between what people are doing and what they ought to do, according to this view. This shows that a given theory can open a critical space to assess the status quo while assuming full compliance with the principles it formulates. To utilise ideal theory in order to define a theory of perfect justice is to undertake a different task than to use ideal theory as a device to provide a critical space to assess the status quo.

One clarification, regarding the keeping of a critical distance from the status quo, relates to the inclusion of facts. When making comparisons regarding different settings and principles we often appeal to facts. Yet, the choice of which facts to consider is value-laden (Chung 2013). That is, we will need in some cases a critical evaluation of which facts to consider. This point is connected to the value of ideal theory as it relates to preserving a critical distance from the status quo, which is also true for specific facts. Ideal level theorising thus contributes to avoiding some sort of realist reductionism. To

provide normative tools and to populate this critical space both is a less rigid and perhaps more fruitful contribution than the demanding task of stating what perfect justice is.

Opening the critical space to criticise the status quo is a characteristic of ideal theory that does not relate to its capacity to respond to practical problems. It relates only to its capacity to formulate judgements about justice. In sum, there are two contributions of ideal theory to political theorising which are not related to it having a capacity to guide action: its (*prima facie*) capacity to shape the social ethos and its capacity to provide us with normative tools to assess and depart from the status quo.

These two contributions are not exclusive to ideal theory. I have mentioned (in footnote 14) that these purposes can be achieved through other means, such as narratives or non-ideal theory. Ideal theories are not the only way through which individuals or groups attempt to shape the social ethos or open up spaces to criticise the status quo). Think of Martin Luther King's *I Have a Dream*. Shaping the social ethos is often done by the stories a community tells about itself. M. L. King's speech had a transformatory force. In a similar fashion to what is done by ideal theories, *I Have a Dream* refers to an idea of equality that was already found in public culture.

This brings to a point Rawls expresses in *Political Liberalism*. Ideal theories often aim to interpret and articulate ideals that are already part of the public culture, notably in liberal democracies (Rawls 1996). Ideal theories are not necessarily the source of moral insights but a refinement of them.

This raises an important question as to whether ideal theory require sophisticated theoretical articulation. If so, then it is clear that ideal theory is the province of philosophy, but that ideal theory is not the only way through which we can shape the social ethos or criticize the status quo. In this case, ideal theory would not be necessary to achieve more ideal settings or provide guidance. If ideal theory does not require sophisticated theoretical articulation, then we would be stretching the meaning of ideal theory, as understood by Rawls and as I will define it in the next chapter. For the purposes of this thesis, I take ideal theory to require a philosophical articulation. This implies, as it should now clear, that ideal theories often express ideals that are already

part of the public culture and that ideal theories are not the only way through which we can achieve the two contributions exposed in this section. I will say more about the relation between ideals and ideal theory in Chapter 2, especially section 2.5, and Chapter 4, especially in sections 4.5 and 4.6.

1.5 On the Use of Idealisations: Good and Bad Theories

We may now ask whether there are contributions of ideal theory that relate to its capacity to respond to practical problems. This is a general question. One of its central elements concerns the appropriate use of idealisations, which is the subject of this section. This will provide us with insights about how to distinguish between good and bad ideal theories and set the basis to address chief objections to ideal theory in the next three sections. Consider again one idealisation (to which I return in detail in the next section): the full-compliance assumption. If theories did not assume full-compliance, so the argument goes, they would not generate principles of justice. This line of argument has received support in the literature (Robeyns 2008, Stemplowska 2008, Swift 2008, Valentini 2009). Others criticise ideal theories precisely because they make idealisations. It is because a theory involves idealisation that ideal theories are impotent and cannot illuminate normative problems (Mills 2005, Farrelly 2007).

For Rawls, the use of idealisations and the bracketing of certain issues when formulating principles is not a problem *per se*. What is problematic for theories of justice is to fail to articulate the values that are necessary to deal with practical problems (Rawls 2001: 66). This echoes the point regarding *A Theory of Justice* and gender issues: it appears that Rawlsian ideal theory can take into account gender injustices because these could be assessed in terms of rights, liberties, opportunities and socioeconomic advantages, which are protected by Rawls's principles. Some theories can account for the non-ideal problems they wish to deal with. Some cannot. This capacity to take injustices into account we observe in *A Theory of Justice* is missing in the *Law of Peoples*. We may follow Valentini who, in her assessment of the *Law of Peoples*, points out to its incapacity

to deal with the problems it is meant to address (Valentini 2009). The *Law of Peoples* must be evaluated based on its capacity to assess the conduct of liberal societies and the ways they exercise power in the international realm. Many say Rawls neglects issues of secession, environment and trade. But as indicated above, leaving issues aside is not necessarily a problem. The question is whether it has the theoretical tools to deal with them. The Rawlsian theory of domestic justice had the theoretical tools, when applied to the appropriate social context, to deal with a number practical problems. We should thus ask whether the *Law of Peoples* could regulate the actions of liberal societies in the international arena.

For instance, principle 8 is devoted to assistance: "Peoples have a duty to assist other peoples living under unfavourable conditions that prevent their having a just or decent political and social regime" (Rawls 1999a: 37). As it has been noted in the global justice debate, the principle neglects borrowing privileges, resources distribution and unequal borrowing power to negotiate terms of cooperation (Pogge 2008: chap 4). As with the question about whether the *Theory of Justice* can address gender injustices in the domestic context, we may rightly ask whether Rawls' *Law of Peoples* is fit to deal with these problems when put in the context of globalised relations. I suggested above that the problem is not with idealisations as such. The problem concerns in fact idealisations that mischaracterise the subject they intend to regulate.

These idealisations in the *Law of Peoples* are problematic precisely for this reason. The problem with the *Law of Peoples* is that it distorts reality. If we follow Rawls in the context of international relations, we would fail to address the very rules that are at the source of various injustices. Rawls formulates a principle of assistance when the problem at hand concerns rules that are at the root of systemic problems that a principle of assistance could never address. The *Law of Peoples*, by formulating a principle of assistance in a context of where systemic injustices are perpetrated, misunderstands the nature of the problem it is supposed to address: there are problems of global justice which involve coordination, others that involve governance, none of which can be addressed by a principle of assistance. The Rawlsian principles of global justice become useless in non-ideal circumstances. Such international practices that go on unimpeded

make the possibility of development for poorer countries a potentially insurmountable burden. Whether Rawls formulated an inappropriate principle or has a different notion of what kind of action is required at the global level, there is the same outcome: his theory misunderstands the nature of the problems of global justice.[18]

There is also an ideological component in the *Law of Peoples* that exacerbates its flaws (Valentini 2009: 349). Assuming that liberal societies are well-ordered obscures the fact that they are not. That is because they are responsible for many of the on-going injustices in the world today. The principle of assistance regulates relationships it fails to understand. Moving from ideal to non-ideal theory generates a misleading recommendation. We may add to this picture the fact Rawls believed that 'burdened societies' (those that struggle with social and economic conditions, and are incapable of maintaining either liberal or decent institutions) are responsible for their own plight: they are 'burdened' as a consequence of their internal institutional failures. This indicates that Rawls not only mischaracterises the social relations, but also the agents to which principles of justice should apply to. As Valentini explains, including idealisations in the design of principles do not condemn a theory to failure, as long as the objects the theory is supposed to address are not assumed as a condition of implementation (*Ibid*. 353). Rawls' ideal theory of domestic justice and Dworkin's 'ideal ideal' world make idealised assumptions about the conditions under which principles are chosen, but try to remain relatively realistic about the subjects or agents to which their theory should apply.[19] These theories would fail to guide action if fulfilling their principles would require citizens to be rational, altruistic or unbiased; assumptions neither Rawls nor Dworkin make.

In other words, it seems that the assumptions in the *Law of Peoples* are such that they make it impossible to address the problem it means to solve, while *A Theory of*

---

[18] In terms of this research, I will argue that the *Law of Peoples* cannot guide action. We may claim that Rawls and Pogge have a disagreement regarding what action is required. But misunderstanding what action is required, or having a principle that fails to capture the problem it is supposed to address, is just as bad. In either case, this theory will fail to guide action. Moreover, Rawls is committed to a theory about wealth and development at the international level. Yet, even if this theory is shown to be correct empirically, he would be wrong not to focus on the causes of injustice at the global level.
[19] Note that, for Dworkin, the point of political philosophy is a practical one.

Justice is capable of integrating some important non-ideal circumstances. The means by which Rawls and Dworkin can address injustices are the basic structure and the legal system respectively. The subject of justice is not idealised in their case. By contrast, in the *Law of Peoples*, Rawls asks how almost self-contained political communities ought to behave towards one another, with the assumption of a fair background of interaction. But these societies are not well ordered and the fair background does not exist. This is problematic precisely because this is why we need a theory of global justice in the first place (Caney 2002, Pogge 2004, Ronzoni 2009, Valentini 2009).

In Rawls' ideal theory of international relations, trading relations and agreements are fair. Valentini says this is 'too far' from the status quo, in the sense that the idealisations it makes leave too much of a gap between the actual state of institutions and the state of interaction between agents at the level the theory is written. A transition between ideal theory and non-ideal theory would require abandoning the idealisations of self-sufficiency and fair background, which unfortunately are fundamental for Rawls. But changing these assumptions would be to construct another theory. In short, we may say that *A Theory of Justice* and *Sovereign Virtue* tell us what the background ought to look like while the *Law of Peoples* assumes that a fair background is in place.[20]

We may conclude here that there is nothing wrong with ideal theory as such, but we must distinguish good ideal theories from bad ones. There are idealisations built into the formulation of theories which result in a false account of the phenomena the theory wants to address. These theories will fail to guide action.[21] Bad ideal theories are incapable of reintroducing or addressing, in real world circumstances, the facts that they abstracted from at the level of ideal theory. Not all ideal theories are necessarily unsound because they fail to guide action. The problem is less the making of idealisations as such, than the failure to address them in non-ideal theory or to make idealisations that can

---

[20] We may ask whether this is the case also with *A Theory of Justice*. Are full compliance and favourable circumstances less problematic? I return to this question in the next section and in chapter 3, where I examine the notion of compliance in non-ideal theory.

[21] For Valentini, idealisations make the world seem better and simpler than it is, but the difference between good and bad idealisations is that good ones do so temporarily and bad ones do so permanently. It is thus paramount for ideal theories to know whether or not the facts that are denied in ideal theory can be taken into account in non-ideal 'at the level of its application'.

potentially be addressed in non-ideal theory. We saw that some theories can do it, and some theories cannot. Those that cannot are those that idealise the subject they are supposed to address and idealise about background conditions, which in fact they were supposed to be defining. As Robeyns puts it, "the problem is not that idealisations are not acceptable, but rather that we need to know how to deal with the idealisation when moving to the non-ideal level" (Robeyns 2008: 357). We would be right to criticise Rawls for assuming full compliance, not for assuming it in ideal theory (which is not a bad idealisation), but rather if he would not be capable of addressing non-compliance in non-ideal theory. This concludes the point on idealisations.

So far, we have seen that it is possible to do ideal theory on its own, and that such ideal theory contributes to political theorising in important ways, notably in shaping the social ethos and in offering a critical space to assess the *status quo*. Also, we saw that we can distinguish between good and bad idealisations in at least one important respect. Good idealisations do not prevent different ideal theories from addressing the problems they are meant to tackle.

This discussion of ideal theory has not led us yet to an account of the nature of an action-guiding theory of justice. What has been said so far does not yet support the claim that ideal theories can guide action and how ideal theory responds to practical problems; what we have seen is that some theories can be sound and these may have an action-guidance potential. We have not shown either that ideal theories must guide action.

1.6 The Priority of Ideal Theory: Full Compliance and Path-dependence

In order to frame the discussion about action-guidance in the next chapters, the last three sections of this chapter will discuss some of the functions attributed to ideal theory. One central concern about action-guidance stems from the assertion that ideal theorising should be undertaken prior to non-ideal theorising. I distinguish between two central claims that are implicit in this general statement, which will be discussed

throughout this thesis. There is a general methodological claim, which I labelled above as the 'analytic primacy of ideal theory'.[22] And there is a claim about the connection between these two theories, according to which the path to be undertaken in non-ideal theory depends on the target that will be defined in ideal theory. I will call this the 'path-dependence argument'.

The case for this methodological claim is normally made on epistemic grounds, as mentioned above. We cannot know what non-ideal theory has to accomplish if we do not have an ideal theory at our disposal. This methodological commitment is in fact better explained by the functions attributed to ideal theory. The two main functions are: a 'target function' and a 'measurement function', already anticipated in the section on Rawls. From the former we see that ideal theory sets the target that non-ideal theory should be aiming at. From the latter we see that ideal theory is the standard by which we are capable of assessing how far a non-ideal situation deviates from what it should be, thereby exposing how urgent the situation is.

These two functions are central for our analysis and will set the basis for the examination of the relation between ideal and non-ideal theory based on the notion of action guidance in the chapters to come. I will quote Swift in his definition of action guidance. This definition will be sufficient for the purposes of this chapter. Guiding action means 'knowing what options are feasible, over what time scale, with what probabilities, given where we are now' (Swift 2008: 374). This will be further qualified in the next two chapters.

Like the argument for the analytical primacy of ideal theory, the path-dependence argument also maintains that more-ideal theory is a prerequisite for non-ideal theory. Yet, the path-dependence argument focusses on decision mechanisms that connect the two theories, which the previous argument did not. The argument goes as follows:

> If we conceive of less-ideal theory as aimed at identifying short-term reforms that take seriously the feasibility constraints that bind here and now, while conceiving of more-

---

[22] I understand the claim about the analytic primacy as having two components. First, there is a component of priority: that principles of ideal theory must be known *before* non-ideal theory can be attempted. Second, there is a component of prescription: that the function of non-ideal theory is to implement the precepts of the ideal theory. This second claim is related to the two functions of idea theory presented below. I will argue against the two claims.

ideal theory as aimed at identifying long-term reforms that become relevant if feasibility constraints relax, then it might seem that we could view more-ideal theory as identifying a destination that our short-term reforms should keep in view. (Hamlin and Stemplowska 2012: 59).[23]

This implies that if short term-reforms, which appear desirable from the standpoint of non-ideal theory, set a path that is inconsistent with long-term more ideal goals, the short-term reforms should be avoided. The path-dependence argument aims to provide a decision-making mechanism, on the basis of ideal theory, to inform what options to choose between a given set of available options in non-ideal circumstances. This formulation of the argument is different from the treatment of the notion in economics and social sciences. Path dependence theory was originally developed to explain technological adoption processes and industry evolution in economics (Nelson and Winter 1982: 70). In economics, the idea of path-dependence is generally meant to explain that where we can go next, i.e. the social choices that are available for us in the future, depend not only on where we are now, but also on where we have been (Liebowitz and Margolis 2000: 981). I will show that the path-dependence argument formulated above is not convincing. I will show that the notion of path-dependence that best informs the formulation of action-guiding principles of justice is closer to the one we find in economics and social sciences than to the path-dependence argument. In the terms of this thesis, I will show that the identification of the target will be given through non-ideal theorising.

The path-dependence argument raises several questions, notably: what to do in cases that we observe path-dependence? And, is it ideal or non-ideal theory which informs how to make the best decision between different paths? This Chapter, Chapter 2, and sections of Chapter 3, will focus more on the general claim about analytic primacy. Other sections of Chapter 3 and Chapter 4 will concentrate on the path-dependence argument.

---

[23] Although Hamlin and Stemplowska reject this argument, their formulation here is quite helpful, for this is an argument that we find in ideal theories, such as Rawls' and Simmons'.

In addition to the discussion on the notion of idealisations above, two preliminary remarks are warranted before undertaking the discussion of these two functions of ideal theory. One point concerns the general aim of 'perfect justice' and the other concerns the specific idealisation known as 'the full-compliance assumption', which are related to the analytic primacy claim.

First, ideal theory is said to describe a theory of perfect justice, but we may well raise the question 'how ideal should ideal theory be?'. In our discussion of the two functions of ideal theory, as 'target' and as 'measure', and in our discussion of action-guidance in the chapters to come, it will be paramount to keep in mind that theories are not all 'perfect' at the same level. I return in the next section and the next chapter to the idea of perfect justice as an adequate way to qualify ideal theory. We will see that many objections to ideal theory are in fact objections to the narrow criteria of perfect justice.

Secondly, I will pursue the reasoning regarding the full compliance assumption in more detail. The full compliance assumption is one idealisation in particular and it will receive special examination in this section. For Rawls, full compliance theory is ideal theory (Rawls 1999b: 8-9). Non-compliance theory is the theory that tells us how to deal with injustice. The reason to begin with full compliance is that, as mentioned, it is necessary to have the basis for a systematic grasp of the more pressing problems of non-ideal theory. He claims that this is the fundamental part of a theory of justice.

In order to compare theories of justice and choose one, the choosers in the hypothetical original position would need to compare full compliance theories. For their job is to understand what justice demands. Non-ideal issues – of crime, past injustices, restrictions on liberty – would be judged based on how they depart from ideal theory. Non-compliance with the demands of justice is what must be addressed in non-ideal theory. Non-ideal considerations must not be included in the definition of a conception of justice in the first place. Objections that ideal theory is too demanding, too hard and that people will not follow it cannot be rightly addressed to ideal theory.

For A.J. Simmons full-compliance is the only unrealistic assumption that Rawls required in order to arrive at this realistic utopia (Simmons 2010). This is the only way in which Rawls violates Rousseau's requirement of taking 'men as they are, laws as they

can be'. A theory of justice must be assessed independently of whether people will follow its precepts. Whether people will jeopardize the stability of the ideal system cannot be a concern at this stage. A theory should not be corrupted by making compromises for ill-willed people and wrongdoers. That is why ideal theory has to be undertaken prior to non-ideal theory. For Rawls (1999b), Dworkin (2000: 165), Cohen (2006) and Valentini (2009: 345), a distribution has to be designed against a fair baseline. This involves assuming away non-ideal circumstances, such as discrimination and prejudices, and their causes. If we factored into the formulation of principles of justice existing problems such as discrimination and prejudice, their design would not be one of a just distribution. In other words, including non-ideal circumstances at the level of formulation of principles would corrupt ideal theory.[24] This charge against including real world problems in the formulation of principles is important. It implies that including the real world circumstances such as prejudices in the conceptualisation of a just distribution would prevent us from knowing what a truly just distribution is.[25] We must not take into account the fact that people have prejudices in society when we formulate our principles.

Moreover, we should mention at this point that Rawls' arguments to support his principles of justice depend on his conception of a well-ordered society. The idea that the choice in the original position is an agreement involves what Rawls coined as 'strains of commitment': the parties must choose principles they believe they will be able to accept and comply with (1999: 153). People must accept and live with the principles they choose, and these principles must speak to their sense of justice. It would not be rational

---

[24] Valentini concludes that this argument is not about whether these ideal theories are sound but rather whether they are doomed to be unsound. Apparently, they are not and this could relax the paradox of ideal theory. Dworkin says in *Sovereign Virtue* that when egalitarian principles have to be applied at the level of policy, both deontological concerns for fairness and consequentialist considerations of effectiveness will be factored in (2000 chaps 11 and 12). Particular measures (subsidies for education of formerly discriminated minorities, anti-racial discrimination campaigns) cannot be derived from ideal theory itself.

[25] Even though Rawls is concerned by the capture of the democratic processes by wealthy interests, he is not doing non-ideal theory, considering the idealisations involved in the formulation of his principles and the general target of the basic structure of society. As we will see in Chapter 2, not all ideal theories are idealised at the same level and about the same objects. Moreover, wanting to change one feature of the world that needs reformation is proper to all political theories, ideal and non-ideal alike, as pointed out in the introduction.

for parties to accept principles they would later disagree with and not comply with. It appears only rational that parties would be averse to risk when it comes to the choice of principles of justice for the basic structure.[26] All this supports the notion that without the assumption of full compliance we logically cannot arrive at a conception of justice.

By contrast, for Levy, quoting Madison, 'if men were angels, no government would be necessary'. What is at stake is to ask if it is right for a theory of justice to assume that humans will comply with whatever a theory demands of them. To do that would be to assume away the crime and injustices that justify the state (Levy 2014: 5).

Estlund and Simmons argue in response that distribution cannot be plagued by the fact that someone might not do what he or she ought to. Consider a group of people together in the restaurant. We cannot think of what would be the right way to split a bill among these people in the restaurant if we factor in that people might not pay. We need to know who owes what before considering problems of trespass and non-conformity. Observance of justice has to be taken as prior to the non-compliance with this background (Estlund 2014).

But this in fact takes the question of compliance to another level of debate. For Levy, this might be right within one domain of rules, such as in the restaurant case. Yet, he argues that the question 'why should we have rules?' can only be answered if we know not everyone will obey whatever rules there will be. Although Levy fails to see that we could have rules to coordinate the action of purely altruistic people, he is right to point out to the following question (*Ibid.* 6-7): can the justification of a rule based on strict compliance in one particular rule be generalised to the level of choosing the system of all social cases?

Cohen, in the camping trip analogy, does the opposite (Cohen 2009). The camping trip analogy is meant to illustrate what a social order organized around the principles of equality and community would look like. The camping trip analogy uses a small size social practice with emotional ties to describe the desirability of the social order. By doing so, Cohen scales down from the social to the individual. He imports the ethos of

---

[26] I thank Aaron James for pointing this to me.

camping trip into the individual level. Rawls' argument goes the other way around. He generalises the individual demand of fairness to the general choice about organising society.

Is there anything that should prevent us from drawing such analogies? In Levy's words: "partial compliance at the micro-level might be the result of bad will, but partial compliance with political justice is a central feature of political life" (Levy 2014: 9-10). Morally good agents would still disagree amongst themselves. Jeremy Waldron believes this to be part of 'the circumstances of politics' (Waldron 1999: 102). This is crucial. If justice is the first virtue of social institutions as claims Rawls, but justice is considered in a state of non-friction, then justice would be a solution to a state where there is no friction, according to Levy. 'If we could stipulate full compliance with moral rules however demanding, then there is no reason not to stipulate better virtues than justice and a morally good enough humanity not to need a coercive state at all' (Levy 2014: 8). This is also Philp's view: if morality could give us all the right answers, there would be no need for politics (Philp 2008: 635).

Perhaps Rawls' position is better constructed on the grounds of the role of the basic structure of society, his concern for stability and his argument about the strains of commitment, which limit the relevance of the full compliance assumption. Perhaps Rawls offered a theory that is more clear about the limits of the analogies between the social and the individual than some commentators seem to care. We could admit that Rawls has a valid ideal theory, but the ideal element of his theory cannot be based on the full-compliance assumption. If we adopt this line of reasoning, we would also lose the argument about the primacy of ideal theory, for we could proceed from, say, an account of the role of the institutions that constitute the basic structure.

But is it the case that Rawls really thinks about justice in a situation with no friction? On the one hand, any ideal account of justice based on the full-compliance assumption alone would be vulnerable to Levy's objection. On the other hand, it is worth noting that Rawls' account is also built on various other features, such as the 'strains of commitment', which are supposed to constrain the resulting principles to ones that could plausibly be internalised. Moreover, the problem with Levy's argument is that he

does not see any value in thinking justice in two different stages, with a first stage that proceeds from idealised (and valid) assumptions. For him, there is no value in this exercise, what we already shown there is in 1.4 with the contributions of ideal theory to our theorising about politics.

In my view, the problem is to not distinguish between levels of reasoning that deal with fundamentally distinct problems. We must make room for a level of reasoning that makes such idealisations. But we need to know what type of reasoning best allows us to deal with institutional failures, political disagreement and lack of compliance. The value of the ideal kind of reasoning has already been noted above, and it is substantial. The mistake would be to believe that it covers the whole field of debates about justice. Much of this debate is concerned with looking for ideal theory responses to questions that fall way outside of its scope. In other words, we should not wait for ideal theory to guide action in the real world and explain to us how to deal with institutional failures, political disagreement and lack of compliance. [27]

Knowing what justice is, as Rawls, Estlund and Simmons put it, means in fact knowing possible avenues and tools that could indeed be used to assess situations in the non-ideal world. But this does not mean knowing the principles that we should 'implement' in the real world. This nuance is important for it is at the core of institutional action. Using principles of justice to assess situations does not mean that we know *how* to change these situations. The epistemic access to insights about justice does not entail an epistemic access to solutions to injustices. If our task is to have a better notion of how to change situations, we do not need full compliance. The full compliance assumption is thus only useful in a very limited context, i.e. in some cases about the epistemic access to justice. It is not necessary to ideal theory for ideal theories can be constructed based on various other elements. Therefore, for the rest of this thesis, I will not assume that the soundness of the roles of 'target' and 'measure' of ideal theory depend on the validity of the full-compliance assumption.

---

[27] One may note that this thesis is motivated by some political realist objections to political theory. It must be clear in my view that falling into the traps of political realism would be just as problematic.

To conclude, one of the central problems surrounding the discussion of the full-compliance assumption is that we also assume that theorising about justice is really about effecting a just distribution of a given 'currency'. So far, we have not questioned that the function of justice is to search for what this distribution is. If we relax the assumption that distributive justice is more than a question of distribution of a given currency, as we should and will in the next chapters, we should see that the function of justice, especially in non-ideal theory, is not limited to effecting a distribution of a given set of goods. This should come to the fore in the next chapters in the discussion of the question of compliance in non-ideal theory. The question of compliance in non-ideal theory is not about thinking partial compliance with ideal principles, it is about dealing with partial compliance and seeking compliance with non-ideal principles. Ideal principles formulated under full compliance assumption do not help to understand the role of non-ideal theory. They bear no relation to the work to be accomplished in non-ideal theory (and to this extent I agree with Levy).

Moreover, and centrally, I will argue that including injustices at the level of formulation of principles is a crucial step in our theorising about justice (Chapter 3). It follows that there are ways to say something about justice without knowing what perfect justice is, as this section have shown and as I will argue in the rest of this thesis.

1.7 Ideal Theory as Target

In order to pursue our exploration of the contributions and limits of ideal theory, the two last sections of this chapter will focus on two apparent contributions of ideal theory to political thinking. These are the two functions of ideal theory according to Rawls, which I mentioned above: ideal theory 'as target' and ideal theory 'as measure'. This section begins with the first.

For Rawls, we need ideal theory to give non-ideal theory an objective, even if this objective can only be achieved in the long run. Ideal theory tests "the limits of

practicable political possibility" (Rawls 2001: 4). If we want to correct what is wrong, we must know what is right. This function has an epistemic and a normative component. Epistemically, we cannot know what direction to take, if we do not know the ideal we should aim for. Normatively, the path we should take will be determined by the ideal.

This epistemic element of ideal theory is summarised by the Everest analogy in the literature, given by Sen in *The Idea of Justice*. According to Sen, knowing that Mount Everest is the tallest mountain on the planet is not necessary, nor very helpful, when comparing the height of two mountains or when figuring out which way is uphill? Do we need to know the location of the Everest to know what uphill means? For Sen we do not need to know the location of the Everest to move uphill (Sen 2009: 101).

There are two main critiques of the idea of ideal theory as target. First, for Sen, it is basically that we do not need ideal theory. Secondly, it is that ideal theory distracts us, and thus is dangerous. I will discuss these in turn.

The first question is less about the danger of ideal theory, or its unfeasibility, but its superfluousness in political judgment. Sen considers the hypothesis that the identification of the ideal, even if not sufficient, would be necessary to rank any two alternatives. For Sen, there is no obvious reason why when comparing two alternatives, X and Y, it is necessary to identify a third alternative Z (2009: 101). Stemplowska and Swift summarize Sen's analogy with regard to the 'target function':

> To climb the highest mountain within range, we do not need to know that Everest is the highest mountain in the world but which mountains are within range and how to compare them to each other. Similarly, if justice not mountaineering is our aim, we can easily judge, for example, that the elimination of extreme poverty would constitute progress toward justice without knowing what principles of justice would be recommended by ideal theory. (Stemplowska and Swift 2012: 377)

In this sense, ideal theory is useless as a target. This could potentially contrast with the arguments that ideal theory provides us with the critical space to assess the status quo and that ideal theory can influence positively the social ethos, the two contributions of ideal theory mentioned in 1.4 above. But this is a different case. Ideal theory can very well help with identifying values we wish to promote, open the conceptual space to allow institutions to progress and still fall short of helping us choose between two alternatives.

However, ideal theorists like Simmons will reply that we need ideal theory to make these choices, because we cannot start climbing any hill we happen to find (Simmons 2010). We need ideal theory to know the possible path to higher hills, and for us not to be trapped in second best equilibria. He argues: "The requirement that non-ideal policies be 'likely to be successful' requires that we know how to measure success; and that measure makes essential reference to the ultimate target, the ideal of perfect justice" (Simmons 2010: 25). In other words, ideal theory informs us about the path we should be taking. The contrast between Sen's and Simmons's positions suggests that the answer to the 'target' question is central in order for one to support or to reject the importance of ideal theory.

In reply to Simmons, while it is true that we cannot start climbing any hill we start to find, we know that for whatever climb people need ropes, food and equipment. And we should make sure there are no social structures that prevent people from getting access to these goods. With the many ideals that can be pursued in society, we should make sure people can start their climb unimpeded and they could climb whatever hill they want. We do not need to know what is the highest mountain to know that there are people preventing other people from having access to their ropes, food and equipment.

Take the example of poverty. We may have in hand a theory of human rights, one that was carefully designed through history,[28] and that expresses the aspirations of many people around the world. This theory helps us to better understand atrocities committed around the planet and failures on the part of many countries to do what is necessary in order to help people live a decent life. But to eradicate poverty, do we need a theory of human rights, or rather measures targeting borrowing privileges, tax competition, and unfair intellectual property rights on essential medicines? In choosing between two systems of intellectual property rights we need to know which one will have the best impact in terms of providing incentives to produce the medicines that address the problems of the poor and provide incentives to make these medicines available. It is less the theory of human rights but an empirical assessment of the impacts of two

---

[28] One idea to which I will return in chapter 4 is that ideal theories have evolved substantially in recent history. We have a much better notion now of what is required for people to live decent and fulfilling lives.

institutional frameworks. The exact definition of this ideal seems to be less pressing than to address the structures that prevent them from being realised. Perhaps the notion of 'target' needs to be clarified. The target 'fulfilling human rights' is not what allows us to decide between two real world policies in this case. Knowing about values is important, but its practical relevance seems limited.

This leads us to assess a second concern powerfully voiced by Amartya Sen and by Charles Mills. For Sen and Mills, by focussing on perfect justice, we might lose sight of real world concerns, which should be our focus. Note that Sen uses the term 'transcendental theory' for 'ideal theory':

> The answers that a transcendental approach to justice gives—or can give—are quite distinct and different from the type of concerns that engage people in discussions on injustice and justice in the world, for example, iniquities of hunger, illiteracy, torture, arbitrary incarceration, or medical exclusion as particular social features that need remedying. The focus of these engagements tends to be on the ways and means of advancing justice— or reducing injustice—in the world by remedying these inequities, rather than on looking only for the simultaneous fulfilment of the entire cluster of perfectly just societal arrangements demanded by a particular transcendental theory. (Sen 2006: 218)

This point regarding how and whether ideal theory connects with people's concerns is significant. For Mills, ideal theory distracts our attention from existing injustices. The idealisations particular to the methodology of ideal theory necessarily result in a lack of concern for real world problems (Mills 2005). Mills and Sen are perhaps right to claim that, in some cases, ideal theories can lose sight of the pressing issues that are of vital concern for people around the world. Sen's main charges (2009: 10-15) on this point concern redundancy and theoretical feasibility. Regarding the 'redundancy' charge, Sen states that exercises of practical reason that involve actual choices should focus on choosing between feasible alternatives and not on identifying a perfect just society that could not be transcended (*Ibid*., 9). Whether this is the case for ideal theories in general remain to be seen. But if this aspect of Sen's theory allows us to identify pressing problems, we must perhaps look elsewhere to know what guiding action means. I will examine below the relation between the ideas of selecting a relevant

problem, pointing to the relevant courses of action that answer the problem (comparative theory), and guiding action in a particular case, in order to see whether that the three works are logically separable.[29] Guiding action, in the sense defined by Swift, i.e. 'knowing what options are feasible, over what time scale, with what probabilities, given where we are now' (Swift 2008: 374), requires further investigation.

What is of paramount importance in this debate is that one should not conflate the epistemic role and the normative role of political theory (Levy 2014). One thing is to be able to identify the best and another is to know whether the best is what allows us to move towards it. I will show in the following chapters why knowing what the best social setting is and knowing how to move in the right direction towards more desirable settings are two completely different things and the former does not tell us anything about how to do the latter. Even if we need to identify the best there are no guarantees that we will know how to get there. The epistemic advantages of ideal theory are thus not easily transferable to non-ideal level thinking.

Yet, thinking about long-distance goals is not necessarily a distraction from important paths we should be thinking about. Ideal theories can be ambitious and they contribute to our thinking about politics. The contributions of ideal theory to political theorising mentioned in 1.4 avoid the danger of only thinking about where we can get based on where we are now (which in some cases means a very unfavourable scenario). If we only accept paths that are obvious, we give too much weight to people who oppose inspiring normative aspirations, and who are wrong. People unfortunately have the power to effectively stop right things from being done. But this cannot (ever) mean that, because it will be opposed, what we demand in terms of justice is wrong. Ambition is not the problem. It is more its status in political theorising that is. The value of ambitious theory says nothing about its priority and says nothing about action-guidance.

---

[29] This relates to the important critiques that positions – such Anderson's, Wolff's and Scheffler's – address to distributive theories.

In sum, despite its contributions to political theorising, ideal theory does not provide the target that should be used at the level of non-ideal theory. Ideal theory provides elements that can in fact open new spaces of reasoning, help understand situations and frame institutions for progress. But it does not provide a target for what to do in non-ideal theory, and even less a notion about how to do it. We have seen that the idea of action guidance raises an inquiry of a distinct and complex nature.

One last clarification has to be made. Ideal theories, when they shape institutions, provide powerful tools for action in the real world. For instance, when we have at our disposal a theory of rights that becomes legislation, we have meaningful tools to help prevent rights violations. Rights to free speech, rights of association, right to security and due process rights might once have been ideal theories. That is, one may claim that, in time, ideal theories become part of institutional design, and then become action guiding. This idea will be explored in Chapter 4.

1.8 Ideal Theory as Measure

For Rawls, ideal theory "presents a conception of justice that we are to achieve if we can. Existing institutions are to be judged in the light of this conception" (Rawls 1999b: 246). This second presumed function of ideal theory is that it provides the measure from which injustices can be assessed in the real world. This claim contributes to the demonstration of 'analytic primacy of ideal theory' argument, according to which ideal theory should be known prior to non-ideal theorising and that ideal theory prescribes what non-ideal theory should implement. The claim is that we need to measure improvements within the sphere of justice by means of ideal principles. Rawls and Swift believe this to be the case. This is of recurring importance, considering that the debate about the measure (or, more narrowly, the currency) of justice is one of the most prominent debates of the past decades in political philosophy. In fact, promoting candidates for the right currency of justice constitute some of the most important ideal theories in the literature.

A central objection to ideal theory as measure comes from Amartya Sen in *The Idea of Justice*. As mentioned, for Sen a transcendental theory is neither sufficient nor necessary for comparative theory (Sen 2009: 96-105).[30] A judgement about a comparative theory is not based on how distant it is, or how it departs, from a given transcendental theory. Some criticise idealisations in ideal theory because their lack of connection with the real world would never allow for ideal principles to be applicable, i.e. it would not allow for any transition from ideal to non-ideal theory (Phillips 1985). Sen argues that descriptive similarity does not imply valuation similarity (Sen 2009: 18). The fact that we possess a second-best option that is more similar to the first-best does not imply that it is the option we should choose.[31] There are a variety of examples that can be used to illustrate this idea. One of Sen's famous examples is that although we prefer red wine to white wine, we would not choose a mixture of red and white over a white wine, if red wine was not available. We would still prefer a glass of white wine.

I will first argue that Sen's critique of ideal theory 'as measure' is based on a too demanding conception of what ideal theory is. I will then compare a less demanding notion of ideal theory with an example of measurement in non-ideal theory.

Amartya Sen addresses an important critique to the epistemic potential of ideal theory (Sen 2009: 95-101), which brings us to the plurality of ideals that can be used as measurements. Regarding theoretical feasibility, he rejects ideal theory on the basis that there is little agreement about the formulation of the principles that will regulate society, or on the hierarchy between such principles. In this sense, we must recognise the epistemic limitation of our inquiry. Rawls would be wrong in assuming that a consensus about the principles will emerge. People will formulate a plurality of principles. Sen claims we have more chances of obtaining agreement on the progress of justice in specific contexts without aiming to achieve perfect justice (Sen 2009: ix-x).

---

[30] It seems strange that Adam Swift agrees with that. He says that to do comparative theory we need to know relative values of different values that will enter trade-offs. In fact, one missing element in Swift's argument is that ideal theory does not need to be a transcendental theory. Then he claims that although principles of transcendental theory are not always applicable, the reasons that underpin transcendental theory are necessary for comparative theory.

[31] This idea will also be discussed in Chapter 4. This section focusses more on the notion of ideal theory 'as measure' whereas chapter 4 will concentrate on decision-making in non-ideal circumstances.

Part of this charge stems from Sen's characterisation of ideal theory, which identifies ideal theory with perfect justice. Valentini suggests that, intuitively, a theory that is not ideal in this sense is not a normative theory (Valentini 2009: 337-338). I suggested in the last section that we should not equate 'ideal theory' with 'perfect justice theory' (i.e. a theory of defining the principles for a perfectly just society, above which no principles other are required). I argue here, and in the next chapter, that this can be interpreted in a way that yields a too demanding conception of ideal theory. For instance, Sen's characterisation of transcendental theories depicts a very demanding version of ideal theory. Sen's challenge is based on the implication that disagreements about the formulations of justice require ideal theories to define the perfect hierarchy between principles. By doing so, perhaps Sen misses the fact that, as Gilabert points out, there is a difference between finding political demands that are superior to alternatives and claiming to have shown that the alternative proposed is clearly superior to any other (Gilabert 2012a: 45). A theory can be superior, without being the best. Sen does not show that the apparent quest for 'unanimous choice of a unique set' set by proponents of ideal theories seriously implies that they believe their theory to be superior to any conceivable alternative.

Yet, one may claim that Sen has not missed the difference above, for that is precisely how he distinguishes comparative theory from transcendental theory. A theory that identifies a superior theory that is not the transcendental best, is simply a comparative theory. Perhaps this is true and Gilabert's critique of Sen is unclear. But the conclusion still holds, because even if this were true, what is at stake is precisely Sen's distinction between the transcendental and the comparative, based on a too demanding conception of what ideal theory is. Indeed, there are epistemic limitations in human understanding and most theories of justice, including Rawls' and Dworkin's, theories which do not pretend to be superior to any conceivable alternative. The implication above perhaps makes Sen's account of transcendental theory implausible for his characterisation attributes to ideal theory a too demanding role.

We need not deny that there would be disagreements, but we also need not define the ultimate supreme alternative. We can propose a robust and ambitious theory of

justice, which we may use in comparative assessments of ideals, even though this theory would not always permit the conclusion that it is superior to all others. However, it is important here to distinguish between disagreement between different theories and indeterminacy within the structure of a given theory. Sen's argues that partial ordering (that is, ordering between options in a given circumstance) is much less epistemically demanding than complete ordering (which ranks any possible options against one another). We must then observe that partial ordering will often be possible even in cases of indeterminacies within theories and disagreements between theories.

Also, we must acknowledge (as Sen does) that when transitioning from one social setting to another (more desirable setting) we do not always consider every conceivable feasible alternative before choosing one (Gilabert 2012: 45). Ideal theorising does not require us to do that. The epistemic limitation of the challenge must indeed be taken into account, and that implies not attributing an impossible goal to ideal theory. A theory can prove to be superior to some alternatives, but does not need to bring closure to the debate in this matter and can always be challenged by a better one.

Therefore, the challenge for ideal theory as measure is not to complete the epistemic task of arriving at a single and ultimate ideal theory. It is rather to see how an even partial and bottom-up ideal theory could contribute to measuring injustices in the real world.

Now, consider again the example of poverty. Do we need an ideal theory to adequately measure the problems associated with poverty (malnutrition, lack of shelter, deaths from curable diseases, morbidity)? Again, let us not take a very demanding theory of global justice. Let us only consider human rights. Perhaps a theory of human rights could tell us by how much a situation is far from the ideal and therefore guide action in telling us which problems are more urgent.

The tools used in the real world to account for poverty are more and more sophisticated. There is real progress made in the development of analysis indicators that

map how progress in fighting poverty is achieved (Bessell 2015, Wisor 2015).[32] Is it the theory of human rights that allows us to measure it? And is this urgency measure a tool that should tell us where to start? Surely, the latter case cannot be made. It would be too condescending and would disregard all the work that is accomplished every day in poverty eradication. No one really argues that workers in the field should wait for philosophers to agree upon an ideal theory.

We must again distinguish between the epistemic role and the normative role of ideal theory. The epistemic role allows us to provide a precise characterisation of social situations, and allows us to see whether progress is made given this characterisation. But this should be contrasted with practical measurement devices. To eradicate poverty, the instruments needed are more likely to be very applied notions about monetary policy, taxation and intellectual property rights. It is unclear how non-ideal principles targeting these issues will be modelled on a perfectly just background, or a theory of human rights, or any perfectly just distribution. What is at stake is knowing whether valuation similarity, an appropriate measure, allows us to use principles from one situation (ideal) in another situation (less ideal). So far, it seems that the epistemic role of ideal theory 'as measure' is as limited as the epistemic role of ideal theory 'as target'.

We should nuance this point. I will do this in two ways. Firstly, I would like to mention that human rights allow us to judge situations that are particularly unjust. Cases of violation of human rights show that human rights can be particularly strong tools to identify a situation as being unjust. If this is what we should expect from an ideal theory, then its epistemic contribution is not trivial. Ideal theory might not give us practical devices to prevent an injustice but could provide epistemic devices to know that a situation is unjust. I would thus prefer to relate this to the contribution of ideal theory as 'opening a space to criticise the status quo', rather than argue for the practical contribution of ideal theory 'as measure'. In sum, I would say that ideal theory as providing the measure to judge a situation as unjust has epistemic value, especially if we do not equate ideal theory with theory of perfect justice, and this role can be understood

---

[32] Sharon Bessell's and Scott Wisor's research on up-to-date effective poverty measurements was presented in a conference I co-organised at Yale University from October 30 to November 1st 2015.

in a similar fashion as the contribution of ideal theory as a tool to open the space to criticise the status quo. The epistemic role of ideal theory 'as measure' is thus perhaps more useful than the epistemic role of ideal theory as target. However, its practical role 'as measure' is as limited.

Secondly, when we compare institutional settings, we do it based on their realisation of certain values. This implies that we want to know what these values are (Gilabert 2012a). Ideal theory might not be conceptually suited to comparing feasible alternatives. But it is also true that institutional designs must be more explicit about the values they realise for us to know what settings best accomplish certain values. And for that we need to know these values.[33] This brings us to a fundamental point. What the charge against ideal theory 'as measure' does, like some of the previous charges, is call for a better integration between ideal and non-ideal theory. Ideal theory may fall short of having practical relevance, but its epistemic contributions could be useful. Whether practical measurements connect with more ideal measurement is the topic of the next chapter.

1.9 Conclusion

This chapter has taught us valuable lessons. First, we have seen that the debate about distinguishing between two levels of theorising, an ideal and a non-ideal level, has roots in the beginning of political philosophy in Ancient Greece and is still very much alive today. Secondly, we saw that ideal theories make important contributions to

---

[33] For Swift, we do not need transcendental theory to do comparative theory but we still need ideal theory because we need to know about values, we need to assess and weight different aspects of justice and maybe other values too. Furthermore, we need to know how to choose between two options in non-ideal theory (Robeyns example 2008: 345). Swift gives the example of education to show that we need ideal theory to judge situations in non-ideal theory. For Swift, we must know that we should not allow private schools to know that private schools should be allowed in cases of persistent discrimination. This again relates to the unproven point about the extent to which we can do ideal theory without non-ideal theory. Also, I do not believe that in this example Swift rules out that private schools could be acceptable and be helpful in fighting discrimination.

political theorising. I focussed on two: ideal theories open the conceptual space within which we can criticise the status quo and ideal theories can provide tools to positively shape the social ethos. Thirdly, we saw that although ideal theories rely on idealisations, they are not condemned to failure because of that. Rather, there are idealisations that do not prevent ideal theories from addressing problems in non-ideal theory, while there are other theories that mischaracterise the subject they are intended to address because of certain idealisations. The latter theories are impotent. They are bad theories. Fourthly, I have distinguished between the features of ideal theory, such as idealisations (full compliance, favourable circumstances) and perfect justice, and the functions of ideal theory, 'as target' and 'as measure'. These two functions have both an epistemic and a normative component. We have seen that these functions are limited to an epistemic role, and the epistemic role of ideal theory 'as measure' is more useful than the epistemic role of ideal theory 'as target'. However, none of the two accomplish their normative assignment well. In short, there is very little material to suggest that ideal theory is action guiding.

This chapter has pointed to questions that were left answered by contemporary theories of justice. The general problem in this debate is to provide a systematic account of an action-guiding theory. We saw that this question involved sub-questions that most ideal theories did not or needed not address. What does it mean to guide action in order to advance justice in the non-ideal world? For instance, to what extent does a non-ideal theory need to favour compliance? Also, what determines the path to be followed between more ideal and less-ideal social settings? These are some of the questions that motivate the rest of the Part I. The critique of ideal theory will be pursued in the next chapter, while the following chapters will focus more specifically on non-ideal theory. The objective of Part I is to provide a systematic account of an action-guiding non-ideal theory of justice.

2. OVERCOMING THE 'PARADOX' OF IDEAL THEORY

2.1 Introduction

In his *Theory of Moral Sentiments*, Adam Smith says that moral theories should help practice, but would have little usefulness if they presented only the general structure of right and wrong (Smith 1759: 315). More helpful would be a theory of the errors we are prone to making in the process of formulating our moral judgements. From Smith, one may then infer that we can very well start from the particular wrong and, perhaps, that when we encounter wrongs, we generalise about what makes those wrongs wrong, moving back and forth between particular wrongs and a general principle about what is right (Levy 2014). The right does not have logical priority.

Smith's moral theory has been criticised for being too particular to sustain a theory of justice (Griswold 1999), but assessing Smith's moral theory is beyond the scope of this thesis.[34] Nonetheless there is something in Smith's intuition about the formulation of moral principles that I wish to systematise in this chapter and assess how it contributes to our theorising about justice.

From what we saw in the last chapter, one could argue that ideal theory indeed faces a paradox. The 'capacity of guidance' is an important feature of most normative theories, but ideal principles of justice are not well suited to guide action in non-ideal circumstances. Ideal theories aim to guide political action, yet in themselves do not offer solutions that would realise these principles. Valentini and Swift among others argue that ideal theory can overcome this 'paradox' (Swift 2008, Valentini 2009). This is not

---

[34] Also, it is worth noting that insofar as Smith is putting forward a moral theory that could sustain a theory of justice, his theory of justice would be an account of equality that targets social relations, not a distributional account.

entirely right. Ideal theory can escape the paradox, meaning the paradox does not undermine the value of ideal theory. But it does not overcome it, in the sense that it does not guide action. I believe this distinction between escaping and overcoming the paradox to be important. I argue that to affirm that ideal theory has value, as it has been rightly pointed out in the literature, still falls short of showing that ideal theory guides action and 'overcomes' the paradox. This charge presses us to seek plausible avenues to connect ideal values with the non-ideal realisation of justice.

The objective of this chapter is to introduce an analytical framework in support of what I call the 'reflective integration thesis'. This two-step thesis states that: *if we wish to formulate principles of justice that can guide action in non-ideal circumstances, we need to integrate ideal and non-ideal theory*, and *the way to integrate ideal and non-ideal theory is by seeking reflective equilibrium between these levels.* This chapter introduces a theoretical framework that favours the construction of conceptions of justice capable of overcoming the paradox of ideal theory (Gajevic Sayegh 2016 also develops this question). It also examines structural elements underpinning the ideas of reflective equilibrium, action guidance and non-ideal theory.

This chapter will open the door to the exploration of the features of a non-ideal theory of justice in the chapters to come. The argument of this chapter is limited to the claim that, in order to guide action, our principles need to be formulated in the light of real-world considerations, which we only obtain by integrating the relevant empirical work on the matter. Secondly, this chapter suggests that ideal theory is not analytically prior to non-ideal theory, thereby providing grounds to reject the dominant model according to which the role of non-ideal theory is to implement ideal theory.[35] This claim will be confirmed in Chapter 3.

---

[35] In the climate justice literature, Simon Caney also has a position in favour of integration. He argues in favour of integration between *different spheres* of justice, namely global justice and climate justice. See Caney, S., 2012. Just emissions. *Philosophy and Public Affairs,* 40 (4), 255-300. Caney and I use this term in radically different ways. Integration between ideal and non-ideal theory, in the way I argue for it, requires fact-sensitivity to the non-ideal circumstances of a particular sphere of justice in the formulation of the principles of justice that will regulate this sphere. The realisations of justice in different spheres (global health, tax fairness, climate justice) are, to a large extent and in the short to medium terms, independent. That is notably because fact-sensitivity to the non-ideal circumstances that influence the realisation of justice is particular to each sphere.

The primary objective of this chapter is meta-theoretical. It is about elucidating the relationship between ideal and non-ideal theory, by focusing on the pivotal role of action guidance. Yet, this meta-theoretical claim is peculiar in one important respect: a solid empirical analysis is required to derive the conclusion that ideal principles are affected by non-ideal circumstances and empirical features of the world. Indisputably, non-ideal theory influences the implementation of principles. My claim is that it also influences the content of principles. I draw from the methodology of reflective equilibrium, which consists in working back and forth from our considered observations about particular cases and the theoretical principles we believe should be regulating a particular case.

That being said, the ultimate demonstration of the reflective integration claim will only become evident in Part II of the thesis. The three case study presented in chapters 5, 6 and 7 will provide support for this argument in the context of climate justice and global tax justice. If the reflective integration thesis obtains for these spheres of justice, it might also obtain for other spheres.

2.2 The Paradox of Ideal Theory

Despite its contributions, there are limits to ideal theory, specifically in what concerns action-guidance, to which we should now turn. Put broadly, the main fear is that we might not be able to bridge the gap between ideal theories of justice and what should be done in non-ideal circumstances (Valentini 2009: 333).[36] Ideal theories formulate principles, but these principles do not allow agents to make decisions in the real world. There are two formulations of the conundrum of ideal theory in the literature.

---

[36] Both Valentini (2009) and Swift (2008) present versions of the paradox in order ultimately to relax it. They conclude the paradox is to a certain extent only apparent and does not undermine the value of ideal theory.

First, there is what Laura Valentini (2009: 333) calls the apparent 'paradox of ideal theory':

P1) Any sound theory of justice is action-guiding.

P2) Any sound theory of justice is ideal (models a just society).

C) Any ideal theory of justice fails to be action-guiding.

Second, there is Adam Swift's question about whether there is a real separation between two claims (Swift 2008):

1)   Fundamental principles are disconnected from questions of feasibility and facts about institutions and human nature.

2)   Normative theorising should integrate empirical realities of a society.

I will start by focussing on Swift's formulation of the debate. Swift argues that positions (1) and (2) can be held jointly. For Swift, when formulating principles, we can 'invoke highly abstract hypotheticals' and 'take seriously non-ideal considerations' (Swift 2008). In my view, we must disentangle two questions here. One thing is to know whether we can hold the two together and another is to know at what stage of the argument empirical features of the world and non-ideal circumstances intervene.[37] Let us begin with the latter question.

What does it mean to hold (1)? When holding (1), we claim that the formulation of the principles will not be affected by empirical considerations. That is, either we hold (1) and accept that fundamental principles are disconnected from feasibility and facts, or we do not accept this disconnection. Swift's claim (1) focuses strictly on the

---

[37] I distinguish here between non-ideal circumstances and empirical features of the world more generally, which include non-ideal circumstances. Ideal theory is concerned with the two. Non-ideal circumstances are problems related to the compliance of agents. Non ideal circumstances include for instance the context of tax competition where the ongoing system allows for global scale tax avoidance which could diminish compliance with domestic rules. Empirical features of the world are questions about tendencies that can be observed and which can have undesirable consequences, such as preventing the progression of justice. I give in section 6.7 the examples of wrongheaded industrialisation and the problem of carbon leakage, which are consequences that should be taken into account by the formulation of principles of justice (which could however become a non-ideal circumstance related to compliance if, say, some countries would on purpose relax climate legislation to attract investors).

considerations of feasibility and facts. Although they can be related, other characteristics of ideal theory are bracketed in his analysis – e.g. ideal theory as a theoretical stance about the use of idealisations (a broad or a narrow conception, the latter specifying what idealisations we are talking about) or as a stance about the state ideal theory defines (the state of perfect justice, or another). Claim (1) defines a position about the disconnection from feasibility concerns and the use of facts.

But more precisely, holding (1) in the sense Swift first phrases it means that when we are formulating the principles of justice, in their design, we are not required to include facts or feasibility considerations. This is a radical view about the level of theorising concerned with the modelling of principles. But it does not entail a radical view about political theorising in general. We can take seriously non-ideal circumstances, but only *after* we formulate our principles.

Now, regarding the former question, holding (1) and (2) together implies a net separation in the two levels of theorising about the moment facts intervene. We would thus be qualifying ideal theory as a fact-insensitive theory. This view seems aligned with Cohen's positions, one of the few who objected to Rawls for not being ideal enough in the formulations of principles. As we will see shortly, for Cohen, facts intervened only at the level of 'regulation' (Cohen 2003, Cohen 2008).

As we have seen, it is possible for theories to make idealisations and still not be doomed to failure, as long as they do not idealise the subject they intend to regulate, for this leads to many inconsistencies such as mischaracterizing ongoing injustices. However, claiming that ideal theories ought to be fact-insensitive in the formulation of principles necessarily implies making a series of idealisations about the subjects of justice. Holding (1) does not mean doing ideal theory on its own, but doing a very specific kind of ideal theory on its own, one that is highly fact-insensitive. As the previous chapter already suggested, this position is unsound. The section will follow on this demonstration.

For the moment, I should put aside this radical view about (1), but still ask the question about whether we can do ideal theory on its own. I wish to see how the other positions in the literature regard (1). Let us distinguish between the following positions.

a)   We can do ideal theory on its own and accept that our principles of justice can be formulated strictly in ideal theory (Cohen and Rawls should both agree with this claim, and so should Dworkin, Swift, and Valentini);

b)   We can do ideal theory on its own, but we accept that we must revisit our principles of justice in non-ideal theory (this is a much weaker claim about the value of ideal theory; it is the position I defend)

c)   We cannot do ideal theory on its own (Farrelly argues for the logical impossibility of having a separated ideal level, while Sen argues less hardly that ideal theory is not of much use at the level of non-ideal theory);

d)   We should not do ideal theory (Dunn 1990, Dunn, Mills and Farrelly point out to the undesirability and danger of theorising at the ideal level).

Let me say a word about positions of type (b), which will be developed in this thesis. What this position means is that we need to engage in non-ideal theorising in order to get the concepts of ideal theory right. This is a position that claims that far too little attention has been given in ideal theories to how structural facts about unjust and hostile social structures impact on the formulation of principles. It also means that we get to better understand *what* justice is if we ask the relevant questions with regard to *how* to move forward about justice. We must engage in a reflection about how to guide action, by means of including empirical features of the world and non-ideal circumstances, in order to get our concepts of ideal theory right. Too little attention has been given to the content of non-ideal theory.

To sum up, the answer to the question 'can we do ideal theory on its own?' is that although we *can* do it, it is an exercise of limited practical, and to a lesser extent theoretical, value. But this is not the most important question in my view. More importantly is that the interest of this debate is not only about whether we can do ideal theory on its own, but what the contribution of non-ideal theory is. Focussing on

structural elements of non-ideal theory will cast light on the importance of the action-guidance elements of theories of justice.

Positions of type (c) and (d) have also been addressed. These positions point out various ways we should be careful when doing ideal theory. For instance, one argument raised in the previous chapter is that relying too much on ideal theory might distract us from the problems that are the causes of injustice in the first place. Nonetheless, these positions were not conclusive in showing that ideal theory does not contribute to political theorising, or in other words that ideal theories fail *qua* ideal theories. I will conclude the examination of positions of type (c) and (d) in the next section, when exploring Valentini's formulation of the paradox

Regarding positions of type (a), there is one specific view within this group that is not moved by the requirements of (2) – the claim according to which normative theorising should integrate empirical realities of a society – and not moved by the paradox as formulated by Valentini. Indeed, some theories reject the requirement for theories of justice to be action guiding. Swift, Valentini and Stemplowska call these 'fact-insensitive theories', in the sense defined by G.A. Cohen. Facts here mainly concern human nature and social arrangements. I suppose that totally fact-insensitive positions reject the action-guidance requirement because such a requirement implies the knowledge of facts, such as facts about human behaviour and likely compliance, and such facts should not corrupt the formulation of principles. There is thus one kind of view which is fact-insensitive and rejects the idea that ideal theory faces a paradox, simply because ideal theories are not required to guide action.[38] Ideal theories make idealisations in order to define justice in its pure form, where principles of justice need not consider whether people are motivated by them or whether people will comply with them, among other things. Holding a view of type (a), Cohen claims that justice is not about what we should do, but about what we should think even if what we should think

---

[38] For Valentini fact-sensitive theories arise only in the presence of a certain social fact: the existence of power-relations channelled through institutions. Principles are meant to regulate these institutions. Thus, these factual considerations must be taken into account.

has no practical difference (Cohen 2003: 243). Cohen distinguishes between, on the one hand, the claim that ideal theories ask what justice is as such, and on the other hand, non-ideal theory which is concerned about principles of regulation (Cohen 2003, Cohen 2008: 275).

This is a position that defines one *feature* of ideal theory and provides a more precise account of its *function.* The feature is fact-insensitivity and the function is to only be concerned about what justice is, what it requires. Knowing whether principles guide action is not part of its function.[39] The function of such theories is simply to understand something about justice. It is not to move action towards it. Adam Swift claims that the goal of this kind of theories (not necessarily his) is simply epistemological, like theories of mathematicians (Swift 2008: 368). That is, we can be truth seekers without being guides of action. Such a theory could very well leave all action considerations to non-ideal theories that wish to follow its commands. In other words, we need not criticize for not guiding action a theory of justice that does not offer guidance and only seeks truths.[40] These views would not be troubled by the conclusion of last chapter according to which ideal theories only accomplished their epistemic function, but not the practical role that was given to them.

Such theories do not guide action, for they do not have the ambition to do it. If we accept that the role of ideal theory is strictly epistemological, in the sense Swift uses it, we need only say that guiding action is not a concern for theories of justice and conclude therefore that the paradox does not arise. There is a trade-off for any one endorsing a theory of this kind: enjoy that your theory does not face a paradox but accept that your theory will not be action guiding. The trade-off for thinkers like Cohen will not make them sleepless. It is easy for them to accept that their theory will not be action guiding.

---

[39] Valentini and Ronzoni offer a more thorough account of Cohen's view.

[40] Swift claims that the role of theories of justice is to evaluate what are the best feasible options in society. The role of social sciences is to tell us what are the feasible options. Once we know these two things, we know what to do. It should become clear in the following chapters that I challenge this division of the intellectual labour between philosophy and social sciences. For Swift, political philosophy appears to be less important in non-ideal circumstances, because disagreements in political philosophy are less about what should be done (e.g. eradicate poverty) but about the why (e.g. utilitarianism, Rawlsian egalitarianism, sufficientarianism).

For anyone interested in bridging the gap between political theory and political change in the world, ideal 'epistemological' theories are unsatisfactory. Of course, we may point that these theories at least are concerned about justice; and, that perhaps the truths they seek bear on actions.

For all theories that wish to have an impact on the real word, this is not an acceptable position. As mentioned in the introduction, I think most political theories are motivated by actual change and wish to reform features of the world. As I will argue, ideal theories should not be totally fact insensitive if they want to have practical relevance at all, but Cohen is right in saying that guiding action, as such, is not part of ideal theory's mandate.

Contrastingly, Rawls's theory of justice is also a position of type (a) – which also claims we can do ideal theory on its own – although it is fact sensitive in two respects. Firstly, it is important to mention there are facts that are considered in the construction of the original position, known as the 'circumstances of justice', which fall under two categories. Objective circumstances include facts such as moderate scarcity (Rawls 1999b: 109). Subjective facts include among others limited altruism of persons in society (Rawls 1999b: 127). Secondly, his theory is also fact sensitive because he is concerned about the conception of justice that would make citizens legitimately exercise coercive power over one another (Rawls 2001: 41). Rawls takes into account what can be reasonably accepted by citizens.

We know Rawls explicitly asserts he is talking about ideal theory while claiming his theory informs the distribution of resources in society and the structure of citizen's power-relations. This shows that a theory can be both ideal and fact-sensitive,[41] while having concerns about guiding action. Dworkin also 'understands justice in connection with the justification of political power' (Valentini 2009: 337). Justice defines the principles according to which governments should treat their citizens (Dworkin 2000: 1,

---

[41] In *Political Liberalism*, Rawls (1996: 44-46) argues that deferring to idealised conceptions of society and the person 'is a way of continuing public discussion when shared understandings of lesser generality have broken down', and the task of political philosophy is to 'reduce (hopefully resolve) such disagreements'.

65). Dworkin's position has one particularity to which I promised to return. Besides claiming that justice is understood in connection with the justification of political power, for Dworkin, the point of political philosophy is a practical one. It is to respond to politics. Abstract political philosophy should begin in real life because only then it will have the right shape, not only to guide action, but for us to make sure the problems we will bring to the abstract level are genuine (*Ibid*. 4). For Dworkin, we can still do ideal theory on its own, and determine just distributions in abstraction, but this exercise will be informed by certain facts.

The objective of making these references to Rawls's and Dworkin's theories has been to point out, as Valentini underlines, that for them the point of ideal theory is to guide action in non-ideal circumstances. Rawls works mostly in ideal theory, and guiding action seems to be what he believes the point of ideal theory to be. They think that a theory of justice should be action-guiding, 'in the sense of offering a framework of thought, within which to assess, criticize and reform the way power is exercised within society' (Valentini 2009: 337). Swift supports this claim and says that whatever else it might do, theories of justice should guide action (Swift 2008).

There are bad ideal theories, as we have seen in the previous chapter, which make bad idealisations and become useless. And there are theories that are not aimed at guiding action as we have seen in this section. The latter positions should not concern us any more in this thesis, because I am interested in action guidance and I leave aside theories that do not purport to guide action. I will only use them briefly in section 2.4 to make a distinction. The former will be discussed in the following sections. More central attention will be given to ideal theories that make good use of idealisations and have action guidance as an important aim.

In sum, those who advance good ideal theories have said only very little about how their theories can be of practical significance. Abandoning ideal theorising altogether because they do not give explicit commands is not necessary. Yet, for all those who believe or care about the fact that theories should say something about how they guide action, there is something missing in ideal political theorising. Ideal theories may be

sound but unsatisfactory for social scientists or political philosophers interested in empirical problems.

2.3 Exploring the Paradox

The previous section set the basis to explore whether ideal theory truly faces a paradox. Following Valentini's formulation of the paradox, presumably, principles that are designed in ideal conditions are incapable of guiding action in non-ideal circumstances. Given that the capacity of guidance is considered to be an important feature of any normative theory, the 'guidance critique' poses a serious threat to liberal egalitarian theories (Farrelly 2007, Sen 2009 are among those who believe this objection undermines ideal theory).

I have used Swift's definition of action guidance until now: 'knowing what options are feasible, over what time scale, with what probabilities, given where we are now' (Swift 2008: 374). This definition implies that a theory capable of guiding action is a theory capable of understanding and, to a certain extent, addressing the relevant non-ideal circumstances that affect the realisation of justice, for these non-ideal circumstances are what allow determining what is feasible, and to what extent it is feasible. It tells us what to do to progress towards more just social settings.

This brings us to a central point. It is important to distinguish between normative and non-normative supplementations that could be required to make a theory of justice action-guiding. Non-ideal theorising will provide elements that count as normative thought and as social scientific input. Swift's description seems to describe it as it would be a purely social scientific work. In the following chapters, I will provide an account of action guidance that clearly integrates normative and non-normative components.

In this section, I will distinguish between different ways to interpret the guidance critique of the ideal theory paradox.[42] There are different ways to understand the

---

[42] This expands from Valentini (2009).

impossibility of 'bridging the gap' between ideal and non-ideal theory. I will label them G1, G2 and G3.

Firstly, one may claim that failing to bridge the gap means being incapable of implementing principles in the real world. In other words, – this is what I will refer to as G1 – ideal principles would not be applicable today. We will see that in this first interpretation, the objection cannot be rightly levelled against ideal theory.

This interpretation of the ideal theory paradox is that ideal theory cannot induce compliance: a theory of justice cannot be considered action guiding if people do not follow it (Geuss 2005, Geuss 2008). Ideal theory lacks the capacity of guidance because it fails to motivate agents. In practice, it does not work. Yet, one may claim that a theory of justice is meant to provide a conceptual framework to judge people who do not comply. If it is reasonable to expect compliance, actual non-compliance does not say anything about the theory (Goodin 1995a, Goodin 1995b, Valentini 2009: 340). Indeed, G1 models an objection that demands too much of ideal theory. It is not reasonable to require of ideal theory to provide a blueprint of potential actions for people to follow its precepts. Ideal theory can very well provide valid reasons for action and people may still not follow it. That is because motivating action depends on non-ideal structures that influence individual capacity to comply with a reason one may have been initially motivated to follow.[43] That is, ideal theory does not need to motivate action. A non-ideal theory associated with this ideal theory could do it. Therefore, this formulation of the objection is not compelling.

A second way (G2) to understand the critique concerns feasibility: it is to say that ideal theory lacks the capacity of guidance because is not immediately applicable in political decisions here and now. Although ideal theory leaves question unanswered, Valentini says this is not an objection because it puts unreasonable expectations on what ideal theory can offer.[44] She quotes Kant: 'No matter how complete a theory may be, a middle ground term is required between theory and practice, providing a link and a

---

[43] This response I offered is slightly different from Valentini's (2009).

[44] As I will show below, the question is not only about the supplementation of ideal theory by facts or social sciences, but about the normative incompleteness of ideal theory.

transition to one another' (Kant 1970: 61). We must not expect from ideal theory to give a manual for reforms that we should undertake here and now. In other words, this formulation of the critique misunderstands the role of ideal theory. By understanding 'not guiding action' as 'not being applicable', the paradox would not really challenge ideal theory.

Instead, we could formulate the feasibility critique such that not including certain facts could render the theory unfeasible, not only here and now, but always. We would be targeting the more specific idea of fact-sensitivity and, by extension, general idea of feasibility. This brings back the idea that the appropriate choice of facts is crucial to the practical value of ideal theories. The abstraction from certain facts can limit a theory's practical significance. But it does not necessarily do so, as discussed above with the contrast between the considerations of facts in Rawls' *A Theory of Justice* and in his *Law of Peoples*. We have seen that some kinds of abstractions are valid and some are not. Again, this objection is not an insurmountable challenge for ideal theory.

A third interpretation of the paradox of ideal theory has already been discussed briefly, in the critique of ideal theory 'as target'. In sum, it says ideal theory can be morally destructive. This is what I will refer to as G3. We may formulate this objection in various ways. A first formulation is that we must be careful when applying ideal principles in non-ideal circumstances not to obtain morally counterintuitive results (Gajevic Sayegh 2016). A second is that ideal theory distracts us from addressing real-world problems (Phillips 1985, Mills 2005). A third formulation is that there is a gap between ideal and non-ideal theory, and there is no way to bridge it (Feinberg 1973). A fourth one is that one might argue that ideal theories can mislead us in thinking about second bests (Goodin 1995b).

I will say a brief word about the last formulation of G3, for I will return to this in Chapter 4. This last formulation of the third interpretation of the paradox says: all the criteria needed for a first best option do not necessarily inform the criteria needed to select a second best (Goodin 1995b: 53). Think of Goodin's example of the new silver Rolls Royce. If you want a new silver Rolls, although the attributes of a new silver Ford approximate better your new silver Royce than a one-year-old black Mercedes, your

second best option would still be the black Mercedes. There is a difference between surface attributes and underlying motives (*Ibid.* 45). That is because the reasons why you would choose the first best are more important than those characteristics. This explains why, for Goodin, political ideals are superficial.

The first formulation of G3 of the paradox is that the implications of ideal theory can be counterintuitive and harmful when the time comes to use them in the real world. This claim is not that ideal theory is not sufficient. This would only reiterate G2. The claim is that ideal theory is misleading. The real world interpretation of principles can have unintended consequences. For instance, consider the case of a principle for a strictly egalitarian distribution of GHG emissions per capita, with sensitivity for all past emissions (this example will receive careful attention in Chapter 6). If applied strictly to the distribution of GHG emissions, industrialized countries would be allocated very little emission rights. If that would truly set the emissions reduction target of all industrialized countries, the political 'infeasibility' (which should probably be understood as a lack of political will) of achieving such a demanding objective might discourage their climate change mitigation effort altogether. This could be considered as a negative unintended consequence of an ideal theory.

The second formulation of G3 is that ideal theory is part of the causes that prevent us from addressing crucial problems, such as real scarcity, when theorising about justice (Farrelly 2007). For example, it has been argued that ideal theory masks the background structures of power and institutional failures that must be tackled if justice is to be promoted. Another example is to say that ideal theory is counterproductive, in the sense of becoming ideological (Mills 2005). The focus on a fully just social order obscures power-relations, such as those of gender and race. They become blind to past and present injustices and thus contribute to their perpetuation.

The third formulation is unfortunately too vague. The supposed impossibility of bridging the gap is what motivates this thesis. The conceptual space that constitutes this gap has been underexplored in political philosophy. To occupy this space is the central objective of the next chapters. We should see that the first and second formulation of

this third objection will provide insights about how to occupy the space identified by this third formulation.

Versions of G2 and G3 have some authority. This gives us a sense in which ideal theory faces a paradox. G2 has been examined in Chapter 1. It is not a critique directed to all ideal theories but only to some, namely those that are fact-insensitive in a bad way. G3 stressed that normative theories want to guide action, and that normative theories must have an ideal component, but these ideal components risk rendering the theory blind to current injustice or risk leading to important counterintuitive results. This critique will be explored in further details and exemplified in the rest of this thesis. It enlightens us about the limits of ideal theory and the tasks that should be carried on by non-ideal theory.

2.4 Exploring the Distinction

Ideal theory's responses to the paradox are limited: ideal theory is not doomed to failure but it does not overcome the paradox. It does not guide action per se. The best ideal theories provide elements that could be put to use in the formulation of principles in 'reflective integration', a thesis I will present shortly. In other words, ideal theory does not respond to the requirements of a robust account of action guidance, but it provides valuable theoretical inputs into the exercise of reflective equilibrium. In order to have a good understanding of the reflective integration thesis we need a more detailed conceptual framework about what constitutes ideal and non-ideal theory.

This section expands on two important works, by Hamlin and Stemplowska and by Valentini.[45] They initiated the task of reviewing different definitions of ideal and non-

---

[45] Hamlin and Stemplowska discuss some points found in sections 2.4 and 2.5 of this chapter, see Hamlin, A.P. & Stemplowska, Z., 2012. Theory, ideal theory and the theory of ideals. *Political Studies Review,* 10, 48-62. Our two analyses are different in several respects worth mentioning. The four formulations of the debate we distinguish are not the same. The authors do not point out the two possible interpretations of Sen's account. Importantly, I do not reject the distinction between ideal and non-ideal theory based on characteristics one, two, and three for the same reasons. Also, they do not believe these three characteristics apply to both ideal theory and institutional design. Valentini, like Hamlin and

ideal theory (Hamlin and Stemplowska 2012, Valentini 2012). While I build on their work, I depart from it in order to define a theoretical framework that does not presuppose the analytical primacy of ideal theory. I will need to review some elements already addressed so far as they relate to the frame of this definitional work.

Fact sensitivity and fact insensitivity. One way to distinguish between ideal and non-ideal theory is to separate fact-insensitive from fact-sensitive theories. Fact-insensitive theories are those that hold that the formulation of principles of justice should not refer to any facts. G.A. Cohen, for example, claims that facts about human nature such as the relative altruism of agents in society should not influence the formulation of our principles of justice (Cohen 2003). More generally, Cohen claims that every fact sensitive principle depends upon a fact-insensitive principle (Cohen 2003: 218). Contrastingly, theories that refer to facts when formulating principles of justice are called fact sensitive theories. John Rawls considers limited scarcity and limited altruism in the justification of his principles (what he calls the circumstances of justice). Facts play a role in the process leading to the statement of principles. Yet, in order to make this distinction clear, let us call 'fact-sensitive' the positions according to which facts influence the 'formulation' of principles and indeed their 'content'.[46] This would imply that facts influence not only the applicability but also the content of principles.

Full compliance and partial compliance. A second distinction relates to the question of compliance. The full compliance assumption is the supposition according to which (almost all) people comply with (almost all) the demands of the principles of justice that regulate society (Simmons 2010: 8-9). As mentioned, it is often considered to be a central characteristic of ideal theory since Rawls identified strict compliance with

---

Stemplowska, rightly points out the heterogeneity of the debate. This section wishes to build on this cartography. However, it ultimately offers a conceptual framework that does not presuppose the analytical priority of ideal theory. It goes beyond these views in that it argues that a comprehensive non-ideal work must be undertaken in order for principles of justice to overcome the critique addressed to John Rawls, according to which his principles of justice cannot 'have an impact in real-world politics' as Valentini puts it.

[46] Thanks to an anonymous reviewer at CRISPP for drawing my attention to this point.

ideal theory: "Thus I consider primarily what I call strict compliance theory as opposed to partial compliance theory [...] The reason for beginning with ideal theory is that it provides, I believe, the only basis for a systematic grasp of these more pressing problems [i.e. of partial compliance]" (Rawls 1999b: 8). Conversely, a non-ideal theory would be one that does not assume full compliance. According to Rawls, problems of non-compliance, punishment, and disobedience, among others, should be judged in light of the ideal principles once these have been worked out (Rawls 1999b: 215-216).

Maximalist theories and minimalist theories. Amartya Sen distinguishes between transcendental theories and comparative theories: 'The two exercises for identifying perfectly just arrangements, and for determining whether a particular social change would enhance justice, do have motivational links but they are nonetheless analytically disjointed' (Sen 2009: iv). Although Sen does not map his distinction onto the contemporary ideal/non-ideal theory discussions, his intuition describes an important understanding of ideal theory (Robeyns 2008). However, there are two plausible interpretations from Sen's framing of the question.[47]

On the first, we should oppose maximalist theories to minimalist theories. The maximalist/minimalist distinction has two components, one about the scope and one about the demandingness of a conception of justice: scope in the sense of how it occupies the domain of justice (how perfect are its principles), and demandingness about how robust the principles should be in the sense of how much they ask of individuals or the state. A maximalist theory is sometimes called a theory of perfect justice. The characteristic of seeking the perfectly just principles that will regulate society relates solely to the component of scope: a theory of perfect justice defines principles above which no considerations of justice arise. It cannot be surpassed in terms of justice (Sen 2009: 6). A minimalist theory with regard to scope is one that identifies a particular

---

[47] Robeyns (2008) as well as Hamlin and Stemplowska (2012) fail to see this. They seem to envisage, rightly so, that transcendental theories need not to be maximalist. Yet, they do not examine the implications from the two possible interpretations.

threshold of justice, under which injustices occur, but above which other principles could apply.

Rawls's theory of justice is a maximalist theory with regard to scope and demandingness. The scope of its principles is meant to define the 'perfectly just society' – it does not require other principles – and the principles are demanding in that they require very extensive equal rights and liberties, equality of opportunity and fairness in the distribution of socio-economic advantages. Although linked and often associated, the characteristics of scope and demandingness are separable.[48]

Transcendental theory and comparative theory. There is a second and quite different interpretation of Sen's distinction between the transcendental and the comparative. As argued, identifying a 'transcendental' value, aimed towards the evaluation of social and political issues, does not entail defining the perfectly just society. A theory is transcendental when it provides us with the values and tools to assess different social settings, and in this sense transcends *the social settings*. A theory of this kind needs not commit itself to describing the best or the perfectly just institutional setting. Contrastingly, a comparative theory focuses on comparing two institutional settings in order to know which is more just, without following the precepts of a transcendental principle. On Sen's definition, a comparative theory need not appeal to transcendental values, let alone the perfectly just institutions, like the ones of a maximalist position.[49]

There is something particularly important in the intuition put forward by Sen: an exercise of comparative theory can be undertaken independently of the commands of ideal theory. This is the premise this chapter wishes to provide grounds for: non-ideal theory is not applied ideal theory. In other words, these two theories have distinct

---

[48] We could divide them in two distinctions and make five distinctions in total. I use the notion of scope in different ways in this chapter and in the next. Scope in the sense of occupying the domain of justice and scope in a geographical sense are two different meanings of the notion.

[49] Sen would admit that if there was a feasible transcendental theory that allowed us to construct a complete ordering of all possible state of affairs, then a transcendental view would be sufficient for comparative theory. The problem for Sen is that there is no such view and that a complete ordering is not necessary to effect a partial ordering.

functions, and when they connect, their connection is not simply that the latter is an attempt to implement the former. This premise is necessary to the validation of the reflective integration thesis.

Moreover, to distinguish ideal and non-ideal theory simply by appealing to a single dichotomy between two features is to fail to appreciate the many dimensions on which theories can vary. We should not presuppose that the content of non-ideal theory is determined by the content of ideal theory. And we should not assume that their relation is the one of non-ideal theory seeking to implement ideal theory.

That is because the implementation of justice requires a complex analysis of political structures and factors that are particular to non-ideal circumstances. A theory of action guidance involves providing analytical tools for problems that are of very different nature than the problems of ideal theory. If we do not assign different functions to ideal and non-ideal theory, neither will ever overcome the paradox of ideal theory, because neither has the necessary resources to finding ways to guide action.

The framework introduced in the next sub-section will emphasize the need to assign to each theory its own function. These functions will reflect the nature of the constraints faced by each theory. Therefore, it will highlight the importance of non-ideal theory in finding ways to guide action in the world today.

2.5 The Functions of Ideal and Non-Ideal Theory

By the 'functionalist account', I mean the conceptual and analytical framework that distinguishes ideal and non-ideal theory based on their functions. This framework provides tools that will be used to overcome the paradox of ideal theory. It highlights that these functions reflect the nature of the problems and constraints faced by each theory. The account also provides insights about the complex relation existing between ideal and non-ideal theory.

The functionalist account distinguishes three theoretical levels with distinct

objectives, according to the function they serve when theorising about justice.[50] Although they are generally not developed within the same conception of justice, they could be different dimensions of the same conception. Ideal theory is the first level. Non-ideal theory is the second level. Political processes correspond to the third level.

As put forward in the introduction, the first level, ideal theory, seeks to clarify the nature of the values and formulate the principles we wish to pursue in society. Ideal theory raises questions about the justification of principles and assesses problems arising from the existence of a multiplicity of values, such as trade-offs between values and measurability (Hamlin and Stemplowska 2012).[51] The formulation of ideal principles is a quest for balance of moral values which are realised at a just state of affairs (Wiens 2015: 435-436 offers a slightly different definition). Recall the two clarifications made in the introduction. Firstly, ideal theories formulate principles by abstracting from or idealising about features of political regimes. Ideal theories need not take into account the specific features of the different practices. They are not constrained by such features. Secondly, ideal theories fail to guide action, because guiding action falls outside of their scope. Guiding action is not the function of ideal theory.

Whether ideal theories are constrained by the functioning of actual institutions is an important question. There is no need to limit the epistemic role of ideal theory in such a way. Ideal theories will be sensitive to institutional mechanisms in different ways. Yet, it is important to note, in order to contrast ideal and non-ideal theory, that the

---

[50] Other important contributions have attempted a similar threefold distinction employing a similar terminology, but have not focused on the structural elements defining the relation between these theories. See for instance Gilabert, P., 2012a. Comparative assessments of justice, political feasibility, and ideal theory. *Ethical Theory and Moral Practice,* 15 (1), 39-56. and Gilabert, P., 2012b. *From global poverty to global equality : A philosophical exploration,* 1st ed. Oxford: Oxford University Press, Hamlin, A.P. & Stemplowska, Z., 2012. Theory, ideal theory and the theory of ideals. *Political Studies Review,* 10, 48-62. It is important to stress here that my view is different from Gilabert's (2012b) in some respects. Gilabert takes institutional design to be constituent of ideal theory, while for me it is the central part of non-ideal theory, for it is at the level of institutional design that we can appreciate how non-ideal considerations influence the formulation of principles. This opens to a second difference. Ideal theory still has analytical primacy in Gilabert's account. Non-ideal theory focuses mainly on the implementation of ideal theory in his view. The way I explore this theoretical framework opens the door for the exploration of theory of non-ideal justice that does not depend on the prior agreement upon an ideal theory of distributive justice.
[51] As mentioned in note 44, my definition to some extent resembles Hamlin and Stemplowska's definition of a theory of ideals.

nature of the problems of ideal theory are not limited by the functioning of actual institutions. These values and principles need not be aimed at the regulation of practices.

We have distinguished between two roles for political ideals, an epistemic and a practical role. The epistemic role is to make sense of the values and principles that should be promoted in society. The practical role is to identify what actions would be required to promote and realise the values in question. We saw, however, that this practical role is of limited value in ideal theory. The first sections of this chapter and the last sections of the previous chapter have demonstrated that it would be a mistake to attribute an action-guiding role to ideal theory. But as we should see, this does not prevent ideal theorising from trying to provide inputs for practical reasoning, from the outputs of its epistemic function. I will argue that it is not because we see the lighthouse that we know how to manoeuvre the ship. Ideal theory might give us an idea about where to go, but it does not tell us how to get there.[52]

As mentioned in Chapter 1, the core of Rawls' theory of justice is an ideal theory. Rawls is concerned with the formulation of principles that will regulate a just society, the ideal part of his theory presents a conception of a just society that we are to achieve if we can (1999b: 246). Another example of ideal theory in the sense I understand it is part of Simon Caney's cosmopolitan political theory (Caney 2005). Caney takes very seriously factual and explanatory accounts of the global order and he envisages institutional applications of his principles. Nonetheless, the portion of his theory concerned with analogies between the domestic and the global level and dedicated to the formulation of principles is situated primarily at the level of ideal theory as I understand it. The methodology and the content of his theory would be considered 'ideal' in the sense of this thesis. Briefly, Caney's argument in support of the extension of domestic principles to the global realm is based on the idea that, if these principles are justified domestically, they should be justified globally as well. Since we are entitled to rights and other goods

---

[52] Section 1.4 already mentioned the distinction between ideal theory and other ways that individuals have to achieve the objectives of associated with ideal theory. In Chapter 4, sections 4.4 and 4.5, I explore the relation between ideal theory and values which are already entrenched in civil society. This will allow me to make a nuance: those are the cases when institutions have sufficiently evolved (although not as a result of ideal theorising) so that ideal theories can be of greater practical use.

in virtue of a common humanity that all humans possess, there are no morally relevant differences between the national and the global realm for principles not to apply to both. This principle and the reasoning supporting it are appropriate to the level of ideal theory. Note that ideal theories can be less idealised to a greater or a lesser extent.

Contrasting with ideal theory, this thesis wants to provide a more thorough account of the methods and content of a non-ideal theory. I will provide a definition of non-ideal theory that departs from the dominant view according to which non-ideal theory is defined in relation to ideal theory. I argued up to this point that it would be a mistake to believe that ideal theory provides the target or the measure to non-ideal theorising and that the practical role of ideal theory is limited. My definition of non-ideal theory is one that allows for a theory that accomplishes this practical role.

As defined in the introduction, the primary goal of non-ideal theory is a practical one: to bring the world closer to a normatively justified state of affairs. Non-ideal theories focus on guiding action in order for society to move towards more desirable states of affairs. This is why I argue that non-ideal theory does not presuppose ideal theory: non-ideal theory carries out a different task, one about guiding action given the current state of political regimes and institutional frameworks. Non-ideal theory responds to the guidance critique directed to ideal theories in a more thorough and convincing way. That is because it is conceptually equipped to address it. Non-ideal theories are thus part of the answer to questions about 'what actions are required to make the world more just'. Although non-ideal theories need values, they do not require a complex blueprint of what a perfectly just society looks like or a precise account of trade-offs between values. Values matter for ideal and non-ideal theory, but the role of values differs across the two theories. I will pursue the discussion about the role of values in non-ideal theory in the section on reflection integration below, in Chapter 4 and in Part II of this thesis.

Secondly, non-ideal theories follow what I labelled a bottom-up approach. They proceed from ongoing practices and the potential injustices they perpetrate. They focus on guiding action by tackling political regimes that are themselves sources of injustices. This entails that the normative resources employed to address a given topic focus on an

existing political regime or institutional arrangement. A bottom-up approach means that the philosophical analysis would begin with the examination of a political framework, such as the climate regime to address climate change, international agreements such as the TRIPS agreement on intellectual property, the regime (or the lack of regulation) on monetary policy worldwide or the current system of tax competition at the global level. Domestically, it can mean beginning the examination of a given healthcare system, or even a national emissions trading system. A bottom-up approach may very well begin the philosophical analysis of a regime and raise questions about what principles should regulate it, but needs not exclude that considerations about other regimes or spheres are relevant. Globally, this could mean that the principles of climate justice could draw on considerations of global distributive justice in a broader sense. Or, domestically, it could mean that the principles of justice that should regulate a given healthcare system could be informed by fiscal considerations of taxation more generally. The objective of a bottom-up approach is to provide political theorising with an informed understanding of institutional set-ups and political frameworks. Crucially, as it will become evident in the next chapter, this approach casts light on what features of different political regimes or institutional frameworks influence the compliance of agents and determine the path-dependence for future action.

Thirdly, non-ideal theories are deeply informed by social scientific data: non-ideal theory defines a conceptual space that works in close proximity with the findings of social sciences, natural sciences, psychology, and economics. They do not to abstract from the states of affairs that are themselves sources of injustice, for these states of affairs contain important aspects of the information required to formulate action-guiding principles. Non-ideal theories aim at addressing the non-ideal circumstances and empirical features of the world that explain, among other things, non-compliance and path-dependence.

Non-ideal theories focus on what we could call 'institutional design'. Theories of institutional design raise questions about rational choice, responding to deviant behaviour, and assessing the real world impact of institutional changes, among other things (Goodin 1998). By definition, they want to guide action. They have an action-

guiding function. The action-guidance component and the attention to real world practices are the two most important characteristics of a non-ideal theory of institutional design. Known questions of political theory at this level take particular form given the action-guidance component. For instance, non-compliance is approached in a way that aims at finding the mechanisms that will favour compliance, through incentives and sanctions.

A further question we may address relates to the connection between principles of institutional design and the more abstract principles from ideal theory. As opposed to the implementation-based models, such as the Rawlsian model for which non-ideal theory consists in applying ideal theory, in an integration-based model, the principles of institutional design will be given by engaging in reflective integration between the ideal principles and non-ideal circumstances. On the Rawlsian view, even if we consider relevant empirical matters at this level, ideal theory would still have analytical primacy over non-ideal theory. Therefore, institutional alternatives would always be assessed based on how they fare in the light of an ideal standard. Facts would only be considered with regard to the implementation of principles – the latter already been formulated and not subject to revision.[53] By contrast, the integration model is open to adjusting principles in the light of institutional facts. Consider a theory of institutional design targeting climate policy in an integration perspective. This theory will focus on institutional settings, their ruling principles and the design of treaties that tackle climate change. It will take notice of the functioning and the challenges faced by institutional frameworks such as the United Nations Framework Convention on Climate Change (UNFCCC 2014) and the annual Conference of the Parties (COP). Also, it will integrate facts that can be inferred from the behaviour of agents or that need to be anticipated

---

[53] Thanks to an anonymous reviewer at CRISPP for pressing me to clarify this point. I would add that Rawls does allow his principles to be adjusted in reflective equilibrium, but his version takes into account our considered judgments, not non-ideal circumstances. For instance, Rawls describes the four-stage sequence that clarifies how principles for institutions are to be applied, which model a system for the constitutional powers of government and the basic rights of the citizens "subject to the constraints of the principles already chosen" (Rawls 1999: 172). That is, this constitution already "satisfies the principles of justice" (*Ibid.*, 173). In order words, even though the procedure to derive the just constitution might one of imperfect procedural justice, the principles as such are not subject to revision. In framing the constitution, the principles of justice will define an independent standard.

given the current state of the relevant legislation. For instance, in the design of treaties around the COP, it has been observed that allowing countries flexibility in their choice of policy to meet climate targets is more likely to induce agreement. The US, Canada and the EU seem to be more favourable to market-based solutions as part of the strategy for emissions reduction, and more precisely cap-and-trade measures over carbon taxes (Krugman 2010, Dessler 2012). A theory of institutional design in an integration-based model will balance ideal principles of climate justice against institutional considerations of flexibility.

We may distinguish here a third level, which I call the level of political processes. It addresses questions about political forces and agents' motivation. For principles of institutional design to be implemented, there are always political actions to be undertaken. Political processes include questions such as: what needs to change in the political culture for politicians to be sensitive to a particular issue? In the sense adopted here, the third level involves processes such as campaigns, demonstrations, lobbying and mediatisation of different issues. Proposals targeting political processes tackle the current state and the development of institutional contexts and strategies that are likely to induce change. Proposals can take into account facts such as the influential power of climate sceptics over the US Senate in a way that impacts on the likelihood of domestic climate policy implementation, which has repercussions on retaining agreement of countries that wait upon the US to seriously commit to fighting climate change. This level of reasoning focusses more on the political actions than on the institutional modelling of the responses to injustices. Political actions are not my primary focus. Yet, in a similar way to the level of institutional design, the theorising about political processes is also influenced by inputs from other disciplines, such as social sciences and economics. I return to this relation, especially with regards to non-ideal theories of institutional design, in Chapter 4.

The aim of this section is thus to shift our attention from the distinction between ideal and non-ideal theory based on their inputs (such as the full compliance assumption) to a distinction based on their respective functions. Sen (2009) and

Stemplowska (2008) have also suggested a definition of the two theories based on their function. I am distancing myself from Sen's definition of the debate, mainly because his distinction between the transcendental and the comparative does not capture adequately the contributions of ideal theory or non-ideal theory. Sen fails to consider the possibility that accounts of ideal theory, which are less demanding than what he labels transcendental theory, can provide valuable inputs for the work in non-ideal theory. And his conception of comparative theory seems to limit the contributions of non-ideal theory to the comparative function. As mentioned and as I will show, non-ideal theory has also a reparative and a transitional component. Ideal and non-ideal theorising are more complex exercises than Sen's concepts seem to suggest.

Outputs, in Stemplowska's terminology means a theory's set of principles. For Stemplowska, the different functions of the theories are given by the fact that some theories aim to formulate recommendations that are both achievable and desirable and some do not (Stemplowska 2008: 325). Ideal theories are those that do not wish to formulate recommendations that are both achievable and desirable in the near future, whereas non-ideal theories are those that aim precisely to do that. Stemplowska's intuition is thus similar to mine. The function of ideal theories is not to issue recommendations that are achievable and desirable. The central differences between our accounts is, firstly, that Stemplowska's goal is to inquire about the value of theories that do not aim to issue such recommendations, whereas my goal is precisely to understand the structure of theories that aim to issue them. Secondly, I prefer the term of 'action guiding' instead of 'achievable and desirable'. I will be explicit in arguing that, by exploring whether the implications of principles are desirable through non-ideal level reasoning, we are more likely to obtain action-guiding principles. And I have deconstructed the notion of achievable into its various components. The mechanics of how principles can be considered desirable and achievable in non-ideal theory is not given from the start. It is an intricate process, which reveals the theory's desirability and feasibility, and also its relation with more ideal values and principles. Thus, I maintain the term of 'action-guidance', which reflects more accurately the dynamic aspect of the endeavour.

I will clarify one last idea in this section. I have mentioned that non-ideal theory has components which are reparative, comparative and transitional in nature. I do not wish to suggests that accounts of non-ideal theory should focus on each one of these dimensions individually. I believe many accounts of non-ideal theory can have components of all three kinds. I will not insist on this distinction at this point. The following chapters will provide explanations and examples about the content of non-ideal theory which will clarify the role of these three components of non-ideal theorising.

2.6 Reflective Equilibrium and Integration

The previous section raised important theoretical issues in support of the reflective integration thesis. The thesis states that: *if we wish to formulate principles of justice that can guide action in non-ideal circumstances, we need to integrate ideal and non-ideal theory*, and *the way to integrate ideal and non-ideal theory is by seeking reflective equilibrium between these levels*. It shows that, in order to guide action, our principles of justice need to be formulated in light of real world considerations.

Central to the reflective integration thesis is my argument against what I called 'the analytical primacy of ideal theory': the claim according to which ideal must be known prior to undertaking non-ideal theory, and non-ideal theory is applied ideal theory. The next chapter returns in detail to my critique of this argument. I would only mention here for the purposes of the reflective integration thesis that, indeed, there is no need to assume that ideal theory has analytical priority over non-ideal theory or settle on a hierarchy between ideal and non-ideal theory. My claim is not that non-ideal theory has analytical primacy, but rather that ideal theory does not have analytical primacy. In fact, the reflective integration methodology supplants the question of primacy. It is not necessary to know whether ideal justice is pre-requisite to advance justice in the non-ideal world. Seeking reflective integration means we may very well begin with the analysis of an ongoing injustice, an approach seen in the works of Jonathan Wolff and

Avner De-Shalit, Elizabeth Anderson or Iris Marion Young (Young 1997, Wolff and De-Shalit 2007, Anderson 2010), or begin at the more abstract formulation of ideal principles. The question of primacy does not arise.

Secondly, the methodology of reflective integration is a response to the failure of ideal theory as measure, as mentioned above and championed notably by Rawls: "Existing institutions are to be judged in the light of this conception [of ideal justice] and held to be unjust to the extent that they depart from it without sufficient reason" (Rawls 1999b: 216). To oppose this view is to claim existing institutions need not be judged on the extent to which they depart from ideal theory, and more generally, non-ideal theory should not be seen as the application of ideal theory. The reflective integration thesis is a positive response that complements the critique to ideal theory 'as measure' presented in Chapter 1.

We should approach the dynamics between ideal and non-ideal theory in terms of reflective integration, meaning we can revise our beliefs about ideal principles once we seek coherence between them in non-ideal circumstances and weigh them against non-ideal factors in a particular field of inquiry.[54] In order words, I put forward a method for the formulation of principles that consists in applying Rawls' notion of reflective equilibrium (Rawls 1999b: 18-19, 42-25) to any conception of justice that wishes to formulate action-guiding principles by means of taking into account ideal and non-ideal considerations in a particular field of inquiry. In Rawls' words:

> 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. (Rawls 1999b: 18)

Suppose we consider place of birth as morally arbitrary in the determination of

---

[54] As it will become clear in this section, I understand reflective equilibrium as a method of justification of principles, and not as a descriptive method. On the former interpretation, reflective equilibrium is a method that aims to show why a set of principles is or is not justified and under which conditions more judgements are likely to be correct. On the latter interpretation, it is a method that aims at characterising the conception of justice held by a certain person or group (see Scanlon 2003 for an in depth examination of the notion). On Scanlon's view, reflective equilibrium as a method of justification (the deliberative interpretation as he calls it) has to be undertaken prior to the descriptive exercise.

our right to emit greenhouse gases (GHG). One implication of the moral arbitrariness of the place of birth is the demand for a principle of equality, among which we may find a principle for an equal right to emit (Neumayer 2000, Athanasiou and Baer 2002, Singer 2004, Gosseries 2005). However, we may realise that some people need to emit more in order to fulfil their basic needs (think of those in northern Canada) or build the necessary infrastructure to develop a basic economy (think of power plants in places that lack electricity for their basic industry). People in these two countries would thus need a different right to emit to those living in a country that already enjoys a considerable level of material comfort due to previous industrialisation and does not need that much energy for the basic functioning of its society. In other words, a principle for an equal right to emit can lead to potential injustices. Country of residence might be morally arbitrary, but that does not entail equality in the distribution of the right to emit. These real-world circumstances could lead us to revise our principle of an equal right to emit. This example is clearly too simplistic, but it shows what it means to accept that a real-world circumstance can affect an ideal principle of equality. Not all non-ideal circumstances and empirical features of the world should influence the formulation of principles, of course. The exercise to know what circumstances should be factored in, and to what extent, is part of the work in non-ideal theory.

I believe this interpretation of reflective equilibrium expands on, without being in conflict with, Rawls's own definition of the notion. It not only accounts for challenges between principles and considered judgements (from different moral perspectives sometimes) but also for challenges from real-world considerations that confront the ideal formulation of principles. The relation between ideal and non-ideal theory seems to be more complex than a relation based on implementation.

As with reflective equilibrium, one may ask whether a definitive state of reflective integration is likely to be reached or whether it is more likely that this exercise of going back and forth between principles and non-ideal circumstances will continue indefinitely (Scanlon 2003). For the purposes of discussion, I assume that it is a method through which we can formulate action-guiding principles in the non-ideal circumstances we are now. Considering the idea that we can progress towards more

desirable social settings, the principles we obtain can very well be used again in an exercise of reflective integration.

It is important to mention here that in my account of reflective integration, like in reflective equilibrium, principles can come to be abandoned if they are proved to be indefensible. It is not impossible that we come to see that ideal theory principles are wrong *qua principles*. There is room to criticise ideal theory on the grounds of incompatibility with the ethically relevant features of the circumstances we find ourselves in. That is, although some metaphors of guidance – like the high mountains or the lighthouses – might give the impression that ideal theory gives us something fixed, this is not necessarily the case. In the exercise of reflective integration, we should not take insights from ideal theory as fixed points that provides guidance towards them.

In Scanlon's words, reflective equilibrium is "a level playing field of intuitive justification on which principles and judgements of all levels of generality must compete for our allegiance" (Scanlon 2003: 151). Similarly, in reflective integration, nothing is sacred. Reflective integration allows us, methodologically, to start from ideal theory or from non-ideal theory when formulating of principles. And, in reflective integration, we may come to see that we must abandon principles if they are proven to be indefensible. Chapter 4 returns to this definitional task. Nonetheless, as we will see in the next chapters, in reflective integration, some functions can only be performed by one of the two theories. I will focus on the action guidance aspect of a theory, which can only be given in non-ideal theory.

A third important clarification is warranted (this will be developed in the next two chapters). Not only dealing with imperfect or partial compliance is only one feature of a non-ideal theory, but compliance – i.e. seeking ways for agents to comply with climate policy – does not mean finding ways for agents to comply with ideal principles of climate justice. The question of compliance is not directed to ideal principles but aimed at understanding the factors that influence agents' compliance with non-ideal principles, regulations and policy today. Agents do not comply with ideal principles. They comply first with policy, regulation and potentially non-ideal principles. Work in non-ideal theory is thus aimed at finding ways for agents to comply with the regulations, such as

climate agreements, that are feasible in the short to medium term. And the factors influencing compliance with actual regulations will inform the design of policy that could be action guiding in the near future, thereby exposing why the ideal principle cannot serve as a guide in non-ideal circumstances in the way non-ideal theory can.

This leads to a second point about action guidance. The integration thesis casts light on the notion that both ideal and non-ideal theories contribute to the formulation of principles at different levels. It is not self-contradictory to claim that both have different functions, yet both contribute to formulating principles. Ideal and non-ideal theories have a role to play in the design of action-guiding principles in reflective integration. The content of each theory supplies the necessary components to undertake the exercise of reflective integration. However, this exercise is more complex than simply saying that we have, on one side, the identification of the principles and values formulated at different levels of idealisation and, on the other side, the practical consideration of institutional frameworks and real-world implications of the regulations that seek to promote justice. To include the contributions of non-ideal theory in the formulation of our principles of justice means reflecting about the institutional realisation of principles. This means going beyond the formulation of principles under idealised constraints and considering the institutional realisation of justice in order to determine what the practical considerations tells us about our principles. And this does not preclude values from playing a role in non-ideal theory. Values such as autonomy and responsibility will be considered in conjunction with the practical considerations that are relevant for the realisation of principles in a given institutional context. This thesis aspires to be a step in the direction of giving appropriate attention to non-ideal theory and to the transition between ideal and non-ideal theory.[55]

Now, it will be helpful to clarify the relation between the threefold distinction of the functionalist account and the characteristics of fact sensitivity, compliance and maximalism, identified in section 2.4. These characteristics apply to ideal theory and

---

[55] For other attempts see Brock, G., 2008. Taxation and global justice: Closing the gap between theory and practice. *Journal of Social Philosophy,* 39 (2), 161-184. and Gardiner, S., 2011. *A perfect moral storm* Oxford: Oxford University Press.

institutional design. Theories can be idealised to a greater or a lesser extent according to each one of these features. There are idealised theories that score quite high on all three respects: fact-(in)sensitivity, compliance and maximalism. But not all do. This suggests that we can theorise about ideal principles and values in a more idealised or less idealised fashion. At the level of ideal theory (i.e. at the level of formulating principles and identifying values without referring to actual institutions and practices), it is possible to formulate principles that are more sensitive or less sensitive to facts, and more demanding or less demanding. A theory may aim at the formulation of a minimalist theory of basic rights, do so by being sensitive to facts such as disagreement about the justification of such rights, but be an ideal theory nonetheless, for it may not address the non-ideal circumstances and empirical features of the world (institutional failures, ongoing injustices) that prevent the realisation of these rights.

This reasoning also applies to non-ideal theories of institutional design, which can also be idealised to a greater or lesser extent. This should not trouble the reflective integration thesis. One may navigate between ideal theories and more idealised or less idealised theories of institutional design. That is because we can test feasibility constraints by adjusting levels of idealisation. Feasibility constraints may relax in the future, and putting forward idealised versions of institutional design is a way to cast light on the feasibility constraints that we wish to loosen. Non-ideal theories are thus capable of informing us which feasibility constraints need to relax in order for justice to progress and this can help in thinking about the appropriate timing of change. Consider the following case. An institutional design 'A' fixes emission reduction targets that allow for all countries to stay under the 2 °C threshold. Agents would not comply with it in 2015 for it would be too costly. However, in 2015 agents could commit at reasonable cost to a scenario 'B' that would keep temperature increase to 2.7 °C. In scenario B, sustainable energy alternatives (like solar or wind power) are traded at the current market price. Nonetheless, under scenario B, the price of sustainable energy sources will go down after a few years, increasing the chances of scenario A becoming feasible. This is an example of how a feasibility constraint could relax over time.

Moreover, one may notice one potential tension between a method of reflective

integration which includes ideal and non-ideal theory and a non-ideal theory inquiry which has no ideal theory component. One question that should come to mind is whether it is possible to have an action-guiding theory which is strictly non-ideal. Considering my definition of ideal and non-ideal theory, and considering the exploration of their relation, notably in chapter 4, it should be clear that all normative inquiries contain references to values and principles. Although, as I mentioned, not all reference to values requires sophisticated ideal theorising, non-ideal theory always involves justification that appeals to values and, sometimes, justification with reference to full blown ideal principles. That is, reflective integration can use ideal theories that are idealised to a greater or a lesser extent. If ideal theory is defined, as I did above, to include less maximalist accounts of justice, and if such accounts do not require heavy ideal theory apparatus (such as full-compliance assumptions or other idealisations), the formulation of action-guiding principles in non-ideal theory might indeed only require reference to ideal values at a minimal level. In other words, work in non-ideal theory, at some level, will require reflective integration. Moreover, this clarification insures that the two central theses ('reflective integration' and 'non-ideal theory') of this thesis are always compatible. If non-ideal theory could be done without reference to values, the reflective integration thesis would not hold.

Finally, contrasting with the two previous theoretical issues of ideal theory and institutions, political processes are necessarily non-idealised. They are inevitably bound by the real-world constraints faced by proposals aiming at the realisation of justice. Demands of political processes are not subject to play with parameters of idealisations for they need to address non-compliance and be fact sensitive.

2.7 Action Guidance

Besides having a strong moral justification, it is reasonable to hold that having the capacity to guide action should be among the relevant features of a principle of justice (Stemplowska 2008, Swift 2008, Valentini 2009, Valentini 2011a : 24). This is the kind of

normative work I wish to explore. Stemplowska and Valentini pointed out as discussed above, and as I have in Chapter 1, that there is value in theories that do not aim to guide action. But these are not the theories that concern me anymore. I wish from now on explore the potential of theories that do aim to guide action.

Before examining further elements of action guidance in non-ideal theory, we should say a last word about action guidance in ideal theory.[56] Although ideal principles do not inform us about the relevant non-ideal circumstances and the timeframe considerations we need to address, ideal principles of justice provide us with reasons to act.[57] Sound principles provide robust moral reasons for us to follow a particular course of action. Here are some examples. Climate change has the potential to threaten the lives of individuals everywhere, especially in coastal and equatorial parts of the world. If human life has any value, we have reason to act to prevent this threat (an intrinsic value of human-life argument). A second example is that climate change is occurring as a result of human action, which implies that those emitting polluting gases are responsible for harming others (a responsibility agent-based argument). Third, by acting upon climate change we avoid disastrous consequences (a consequentialist argument). Fourth, we can argue that we collectively own the earth and we are all entitled (a collective ownership argument) to some part of the atmosphere (Risse 2012: 126-128). A fifth avenue would be to assign intrinsic or instrumental value to the natural world. For instance, we could motivate action by pointing to the ecosystemic services that are vital to life on earth, e.g. the regulation of atmospheric composition, the regulation of hydrological systems and the absorption of pollution. Providing reasons to act is a third contribution of ideal theory to political theorising, which can be added to the two previously mentioned in Chapter 1 (shaping the social ethos and providing a critical

---

[56] This chapter is aligned with David Wiens view with respect to the importance of integrating the input from social sciences when defining the directive principles we should aim for (see Wiens, D., 2015. Against ideal guidance. *The Journal of Politics,* 77 (2), 433-446.) Wiens is also critical about the action guiding capacity of ideal theory. Our works differ however about the way we see the relation between ideal and non-ideal theory, for I adopt a reflective-integrative approach, and the way we structure non-ideal theory, for I focus on its action-guiding capacity.

[57] I use 'reasons to act' in a broad way, such as for my claim to be compatible with reasons internalism and reasons externalism, in Williams' words, as it relates to moral motivation, see Williams, B., 1981. *Moral luck : Philosophical papers 1973-1980* Cambridge: Cambridge University Press.

space to assess the status quo). Nonetheless, these three contributions do fall short of guiding action in the sense of this thesis. As I said, it is not because we see the lighthouse that we know how to manoeuvre the ship. The lighthouse guides the ship but does not tell us what to do here and now to avoid crashing the ship on the rocks.

Non-ideal theories of institutional design, both more idealised and less idealised versions, are not only capable of adjusting the different parameters of compliance and fact sensitivity, but are capable of addressing non-ideal circumstances that arise in the existing institutional contexts and practices. Non-ideal theorising allows for the investigation of the real-world circumstances that constrain the realisation of justice.

In other words, I take non-ideal theory to be the response to the paradox of ideal theory. A non-ideal theory can respond to more demanding action guidance considerations. In the next chapters, I carry the task of giving depth to Swift's definition (Swift 2008: 374) mentioned above (in 2.3). I will distinguish between four components of action-guidance in a non-ideal theory of institutional design: compliance, feasibility, fact-sensitivity and path-dependence.

It is through these components that I argue for the four claims that compose the *Non-Ideal Theory Thesis*: (i) ideal theory is not analytically prior to non-ideal theory, (ii) theories of justice – notably theories of global justice and climate justice – are best advanced as pluralist theories composed of bottom-up contributions, meaning that different sets of principles regulate different practices and (iii) determining our duties of justice in a non-ideal world is not a question of non-compliance with ideal principles, but rather an inquiry about compliance with non-ideal principles, and (iv) whether there is path-dependence between ideal principles and non-ideal reforms can only be determined through non-ideal theorising.

This inquiry will clarify issues that a non-ideal theory should be explicit about, such as whether a non-ideal theory enables the assessment of conflicts between values and between courses of action; whether non-ideal circumstances that influence the compliance of agents inform the formulation of regulations; whether it clarifies which feasibility constrains might relax in the future; and how this affects the timeframes of regulation and the path-dependence of future action. The action-guiding capacity of a

non-ideal conception of justice is expressed by these desiderata.

This is a demanding account of action guidance. Highlighting the importance of a robust theory of action guidance is central to the project of Part II. It is intended to express the meaning of guiding political action and to reflect the difficulties that arise when we try to promote justice in the world today. The four claims would benefit from further conceptual exploration and concrete exemplification. Chapters 3 and 4 will carry on with the conceptual exploration. Part II is devoted to exemplification and further support. The practical implications of these claims seen in Part II will clarify elements from the theoretical investigation that follows, by highlighting how non-ideal theories of climate justice and tax competition fulfil these three desiderata in practice.

2.8 Conclusion

In Chapters 1 and 2, we saw three ways of overcoming the paradox of ideal theory. The first way is to point out the value of ideal theory in general. We may point to important roles that ideal theory plays in political theorising and claim that these characteristics are central to our capacity to make progress with respect to justice in the real world. Chapter 1 undertook a serious attempt to habilitate ideal theory. It showed that ideal theory contributes to political theorising in important respects, such as by creating a theoretical space to assess and criticise the status quo.[58] But it has been suggested so far that this falls short of constituting an action guiding theory. That is, ideal theory is not necessarily undermined by the paradox, but the paradox still identifies significant limitations of ideal theory. This first response to the paradox was not successful.

---

[58] In Chapter 1, I have tried to habilitate ideal theory before providing my own definition. I proceeded this way in order to assess the value of ideal theory based on its common definition. From Chapter 2 onwards, I provided a new definition of ideal theory that is to a certain extent compatible with its common understanding. The greatest contribution of my new definition concerns the description of what non-ideal theory is, a notion that will be explored in the rest of this thesis.

A second way to respond to the paradox is to say that ideal theory can take into account the non-ideal circumstances it abstracted from in the construction of the theory, when it is implemented in the real world. This is the most robust interpretation of what a good ideal theory can be. We have seen that some theories are conceptually equipped to take into account the relevant non-ideal circumstances and some theories are not. This strategy focuses also on showing that ideal theory is not useless, this time by showing that it can accommodate the non-ideal circumstances it failed to address in the beginning. As opposed to the previous response, this strategy emphasises ideal theory's capacity to address oppressive power relations, and past and present injustices. However, this strategy, like the previous one, still focuses heavily on the value of ideal theory and not enough on a theory which has action-guidance as a central component. Moreover, this strategy runs the risks of presupposing that non-ideal theory is simply applied ideal theory and of presupposing that we cannot do non-ideal theory without ideal theory, which is a much stronger claim that simply saying we can do ideal theory on its own. Although ideal theory can be done on its own, I have suggested that ideal theory does not have analytical priority over non-ideal theory, mainly because the conceptual tools of and function of non-ideal theorising are particular to this level of reasoning.

This leads to a third way to overcome the paradox of idea theory. It implies putting forward a new theoretical framework for the distinction, one that clarifies the dynamic between ideal and non-ideal theory, and also emphasises the theoretical structure that takes the idea of guiding action seriously. This third strategy does not entirely contradict the two previous strategies. Rather, it includes elements of them. Which means that it will also need to take seriously the charges against ideal theory, but will offer a framework to address them. This strategy aims at providing an action-guiding theory. For that, we need to focus on what the requirements of action-guidance are.

I first presented this new theoretical framework contrasting with the main ways ideal and non-ideal theory have been defined in the literature. I presented a functionalist account: ideal and non-ideal theory were distinguished based on the function they are

devised to accomplish. It argued that these functions reflected the nature of the problems and constraints faced by each theory and that the function of guiding action was particular to non-ideal theory. This put us in a better position to revisit the value of ideal and non-ideal theory, as well as the guidance critique.

3. ACTION-GUIDANCE IN A NON-IDEAL WORLD

3.1 Introduction

In the global justice literature, growing attention has been given to problems particular to a globalized economy such as tax competition and illicit financial flows. Political philosophers have started to reflect on how these problems intersect with theories of global justice (Brock and Pogge 2014). This chapter explores the idea that action-guiding principles of global justice can only be formulated at such intersections. This idea guides my account of a non-ideal theory of global justice. The *methodology* of this theory posits that principles of justice are formulated according to the practice they are intended to regulate. Individual practices provide insights about the formulation of principles, for the non-ideal circumstances that prevent the realisation of justice are only revealed through the analysis of the internal features of each practice. With regard to the *content* of principles, I reject the notion that non-ideal theory is applied ideal theory. I offer instead an overview of the main features of a conception of justice for a non-ideal world based in the ideas of incentivizing compliance, fact-sensitivity, envisaging political feasibility and assessing path-dependence.

The objective of this chapter, like of the rest of Part I, is essentially meta-theoretical: it is to explain the contribution that a non-ideal approach can make to our theorising about global politics and to explore the action guiding potential of non-ideal theory by identifying its defining features. The contribution of this chapter is twofold: to show that a non-ideal methodology is well-suited to address global justice problems such as tax competition, and to offer an overview of the elements that should structure the content of a non-ideal theory. These elements will provide a conceptual framework for an action guiding theory of justice in the context of tax competition, climate justice and

carbon pricing (in Part II). I will not develop a full-blown theory that demonstrates how to regulate these issues; I will instead use them as case studies, allowing me to show how non-ideal theorising might contribute to advancing justice in the real world.

More precisely, this chapter will argue for four claims, which have not yet been fully developed: (i) ideal theory is not analytically prior to non-ideal theory, (ii) theories of justice – notably theories of global justice and climate justice – are best advanced as pluralist theories composed of bottom-up contributions, meaning that different sets of principles regulate different practices, (iii) determining our duties of justice in a non-ideal world is not a question of non-compliance with ideal principles, but rather an inquiry about compliance with non-ideal principles, and (iv) whether we can observe path-dependence between ideal principles and non-ideal reforms is something that can only be determined through non-ideal theorising.

Claims (i) and (ii) are argumentative in nature and they relate to the first contribution of this chapter: they aim at showing how the conception of non-ideal theory advocated in this thesis overcomes central problems in the global justice literature and contributes to our theorising about global politics. Global justice will be used here as a starting point. I take that the bottom-up approach to theories of global justice and of climate justice in this thesis can be expanded to other domains of justice. Claims (iii) and (iv) are analytical and relate to the second contribution: they explore the internal structure of an action guiding theory of justice. In other words, after providing arguments that show how non-ideal theory contributes to our theorising about global politics, this chapter offers insights about the conceptual elements that should be developed in non-ideal theorising.

This chapter proceeds as follows. Section 3.2 argues that cosmopolitan theories of global justice have failed to make room for the specificities of social relations in the formulations of principles of justice. The subsequent two sections focus on the methodology of non-ideal theory. Section 3.3 covers the second step of the argument: clarifying that social relations are practice-mediated and therefore that principles should target individual practices. Section 3.4 highlights the contribution that can be made by a theory that takes practices as its starting-point: such a theory stands to give past and

ongoing injustices the normative weight they require. Collectively, the arguments of these sections support claims (i) and (ii). Sections 3.5 and 3.6 explore the key features structuring the content of non-ideal theory and revealing its action guiding potential (claims iii and iv). Section 3.7 concludes by introducing the issues that will be discussed in Part II, suggesting that a non-ideal conception of justice that proceeds from the practice delineates the questions that have to be addressed for global justice to progress today.

3.2 The global justice debate: an overview

The global justice literature today is less centred on the debate between cosmopolitans and statists. Studies concerning the pressing moral problems of a globalized economy have been undertaken recently. In this section, I wish to point out one aspect of the global justice debate that has largely gone unnoticed. In the next two sections, I will argue that this element of the debate casts light on the locus of non-ideal research in the field.

All cosmopolitans hold that individuals are the ultimate unit of moral concern and that this status applies to all human beings (Pogge 2008). They also believe that principles of justice should be extended from the domestic realm to the world at large (Tan 2004, Caney 2005). As Pogge explains, the cosmopolitan position is generally motivated by three premises.

> First, individualism: the ultimate units of concern are human beings, or persons –rather than, say, family lines, tribes, ethnic, cultural, or religious communities, nations, or states […] Second, universality: the status of ultimate unit of concern attaches to every living human being equally […] Third, generality: this special status has global force. (Pogge 1992: 48-49).

For cosmopolitans, these are sufficient reasons for principles of justice to apply globally. Cosmopolitanism is one of the key positions of the global justice debate. This debate was originally structured around one question, the question of extension: Should

principles of global justice be the extension of principles of domestic justice? Cosmopolitans answer it in the affirmative. This answer implies that, if we agreed that one set of principles (Rawls' three principles of justice, say) should regulate the domestic realm, the (non-relational) cosmopolitan would argue that these principles should be extended globally, because of the equal moral status of individuals *qua* persons (Tan 2004, Caney 2005).[59]

Statists answer the question of extension in the negative. They argue that egalitarian principles are of concern only within self-contained political communities. Principles of assistance (supererogatory principles) may apply globally, but not principles of justice. They justify their views in many ways. Statists maintain that egalitarian considerations arise solely in the presence of certain features, such as a coercive force or a political authority, and that these features are only present within state borders (Blake 2002, Nagel 2005). Other statists arrive at the same conclusion, but hold different views about what constrains the boundaries of justice – for example a shared common culture (Walzer 1987), or a shared nation (Miller 1995). For statists, therefore, principles of justice apply only at the domestic level. For Rawls this is even more so the case, considering he believes that the problems of justice that occur in foreign countries are mostly due to the political failures of these countries (Rawls 1999a).

In our analysis of theories of global justice, it will be helpful to distinguish between four theoretical levels of a conception of justice: justification, scope and content of principles (Sangiovanni 2007, Valentini 2010) and also the notion of the function of justice.[60] *Justification* is the process employed to justify principles of justice. Justifications describe processes, such as about acceptability, i.e. the idea that principles of justice are justified when they can reasonably be accepted by all rational agents. Alternatively,

---

[59] This is true only for 'non-relational' cosmopolitans. There is a distinction between relational and non-relational cosmopolitans. Relational cosmopolitans do not believe that moral equality is sufficient for considerations of global justice to arise. For them, principles of justice apply insofar as individuals share a set of political, economic and legal institutions. They argue that such features are present at the global level. That is, relational cosmopolitans believe principles of justice should be extended from the domestic to the global realm because the global arena shares these relevant features with the domestic realm.

[60] This distinction does not map onto the distinction between ideal and non-ideal theory. Elements of justification, scope, content and function can be addressed from an ideal and a non-ideal theory perspective.

justification might involve identifying a morally relevant relation, such as oppression or domination, which triggers the formulation of such principles. The *scope* of justice determines which agents and practices are subject to the relevant principles of justice. For example, some think that principles of justice are of concern only within culturally determined contexts, while others believe they should have global reach. The *content* of justice is the appropriate way to understand the political values that are used to interpret justice, e.g. equality or reciprocity. On the one hand, one may claim equality implies an equal distribution of social goods such as resources or capabilities, or the identification of a sufficiency threshold of goods all individuals should enjoy. On the other hand, the content of principles can consist in the regulation of specific relations within which we would interpret equality. Finally, the *function* of justice is the role that justice should play in the organization of society – e.g. to prevent undue coercion, or to determine how a given set of goods should be distributed.

3.3 Moving beyond cosmopolitanism and statism

Now, consider again the question of extension, around which much of the global justice debate has revolved: should principles of global justice be the extension of principles of domestic justice? As some authors have pointed out (Miller 2007, Wolff 2009), there are two steps in answering this question. First, we may ask whether there are considerations of justice that are global in scope. Second, we can determine whether the principles of justice that apply globally are simply the same principles that apply domestically. To a certain extent, I argue that we can claim that the scope of justice is global because there are relations at the global level that raise justice-related concerns, without having to claim that the same principles will apply at both levels. Yet, given that the two steps above do not require an inquiry about particular regimes and institutions, this reasoning presupposes that the two questions can be answered at the level of ideal theory. A bottom-up approach cannot make this assumption. The task of non-ideal theory is precisely to understand the specificity of regimes and institutions, why they

raise justice-related concerns, and how a conception of justice can guide action in order to address these concerns (James 2012 offers a powerful account about the moral importance of considering the particularities of relations).[61] In this regard, the non-ideal bottom-up approach cuts across cosmopolitism and statism. The first question will be answered in the affirmative only if particular regimes and institutions *at the global level* are the source of considerations of justice. But the second question does not arise. Given that questions of justice are asked only in relation to particular regimes or institutions, the debate over whether principles of justice should or should not be extended from the domestic to the global level is not relevant.

That said, given their inability to distinguish the social relations that are specific to the global arena from those that concern only the domestic arena, cosmopolitans arguably face a major problem. Cosmopolitan principles are not well suited to regulate all the background of global interactions. Because they extend principles from the domestic to the global level, they propose a 'one size fits all' approach to the multiple problems of global injustice. This presupposes that all problems of global justice are of the same nature. However, as I will show in Part II, this is not the case: for instance, the problem of tax competition has the form of an 'asymmetric prisoner's dilemma', whereas the problem of climate change is one of collective action, but one where one agent would be justified in undertaking the collective action based on the individual benefits resulting from this action. Sharing a common humanity does not entail advocating for one single set of principles for the plurality of problems of global injustice. The global practices are what guides us best in the formulation of principles of justice, not the fact that we share a common humanity. The causes of global injustice – e.g. tax competition, colonialism, migration, borrowing privileges, resources privileges, trade agreements on intellectual property rights, climate change – are profoundly heterogeneous.

---

[61] There are aspects of non-ideal theory that are not directly related to relations. For example, parts of the law are justified on the grounds of public policy rather than morality. For the purposes of this research, the scope of this thesis is limited to the domain of political morality.

Statism does not fare much better than cosmopolitanism in the global justice debate. Some argue that the problem with statist positions is that they are not committed to the moral equality of persons. I disagree; I think most statist positions do in fact demand equal consideration of persons (Valentini 2010 shares this view). What makes their position distinctive is that they believe egalitarian principles can only be realised when individuals share a certain bond, such as a shared political authority or a shared culture. Statists are right not to extend principles of justice from the domestic to the global realm, but they are wrong in maintaining that no considerations of justice apply at the global level. Most statists would probably agree that there are some relations at the global level that have a pervasive impact on each individual's opportunities. Because of this, though, statists ought not to constrain the scope of justice to the level of nation-states, since in doing so they fail to recognize a number of morally problematic relations. In a nutshell, statism is unsatisfactory from the standpoint of political morality because it fails to give sufficient consideration to problems of justice in a globalised world. In a world where international institutions implement borrowing privileges that encourage corruption and create poverty, where agreements on intellectual property prevent individuals from having access to basic medicine and where illicit tax flows deprive nations of important revenues that could be used to implement social policies, it is at the very least uncomfortable, morally speaking, to hold that no considerations of justice apply beyond state borders.

This chapter puts forward a conception of justice which supports the cosmopolitan claim that principles of justice – which establish a system of rights and duties (Valentini 2011a: 8-9) – and not simply principles of assistance apply at the global level.[62] However, as it has just been pointed out, different principles apply to the domestic and global realms, because the regimes and institutions that are relevant from the standpoint of justice in both realms are different. We can distinguish between the

---

[62] This says something about the division of labour between philosophy and social sciences. The new theories in political philosophy mentioned in the opening sentence of this chapter initiated a shift from ideal theories (that overly focussed on defining individual entitlements and obligations to implement ideal principles) to philosophical theories that integrate the empirical work of social sciences (in order provide a normative analysis which proceeds from the non-ideal circumstances that constrain the resolving of one particular problem of justice). I will return to this in Chapter 4 and the Conclusion.

morally relevant features of social relations that govern the domestic realm on the one hand and those that govern the global realm on the other. The nature of the problem, notably as it relates to compliance, faced in each case is given by the characteristics of the regime or institution.

3.4 The methodology of non-ideal theory

The second contribution of this chapter is to explore the idea of action-guidance by examining the defining features of non-ideal theory. In order to do this, further scrutiny about the first contribution is warranted. That is, we should try to clarify why a non-ideal methodology is well suited to address global justice problems such as tax competition. It will be easier to explore the content of non-ideal theory if first we define the methodology of non-ideal theory. To this end, I will offer support for claims (i) and (ii) in this section and the next.[63] To recap, claim (i) states that ideal theory does not have analytical primacy over non-ideal theory, and claim (ii) states that theories of justice – notably theories of global justice and climate justice – are best advanced as pluralist theories composed of bottom-up contributions, meaning that different sets of principles regulate different practices.[64] Broadly taken, these two objectives bring us closer to narrowing the gap between the formulation of principles and the practices that constrain the advancement of justice in the world today. More specifically, they claim that it is the understanding of these practices, not the abstract interpretation of the principles, that allow us to bridge this gap. The second claim aims to overcome the

---

[63] This will set the table for exploring claim (iii) in sections 5 and 6.
[64] It is important to note that the bottom-up approach I put forward should not be equated to what has been labelled in the political philosophy literature as a 'practice-dependent approach'. These two approaches differ in one key respect. While I argue that principles of justice are formulated in order to regulate a particular regime or institution, I do not argue that the bottom-up approach is interpretative like practice-dependent approaches are. The bottom-up approach does not depend on an interpretation of the goal or purpose of each practice. In this regard, I am closer to a Rawlsian constructivist approach in that I argue that insights for the formulation of regulative principles are given by an account of the nature of the regime or institution. The nature of the problem and its associated questions such as agency inform the solution to it.

problem of cosmopolitanism, that theories of global justice seem incapable of formulating principles that speak to the causes of injustices in the global realm. This point about global justice will provide a powerful example about the contributions of a bottom-up approach to political theory, which will be used in the three chapters of Part II.

The first claim must be established first, however: I must show that there are good reasons to proceed from the causes of injustice, in order to see that we can learn about what to do here and now by studying these unjust practices. The first step of this demonstration is showing that there is value in proceeding from political regimes and institutional arrangements, which is the question the previous section left us with: how exactly does a theory that proceeds from individual practices contribute to our theorising about justice?

I will offer an answer to this question in the next section. First, though, we must define the notion of *practice* and the role of practices in the formulation of principles of justice.

I have so far employed the ideas of relations and practices quite indistinctly. Henceforth, by *relations*, we should understand things like domination, exploitation, reciprocity and coercion, or, more positively, reciprocity or participation. Practices are more precise. They mediate relations, and can therefore perpetuate unjust relations. To a certain extent, my definition of *practice* follows Sangiovanni's definition of *institutional systems*. Practices should be understood as a set of formal or informal norms, rules and decision-making procedures that regulate a political or social activity (Sangiovanni 2008: 142). Social and political activities regulate access to goods, the division of opportunities and patterns of political authority. Practices can include political organizations or regimes; political organizations are structured around a set of rules, most often formal rules, and act as social actors exercising authority over its members. A political organization exercises *de facto* authority when it claims to impose duties and confer rights, and exercises *de jure* authority when this clam to authority is legitimized by its members. Regimes are sets of formal or informal rules that regulate a sphere of activity.

Regimes can be regulated in part through political organizations, but can also be totally unregulated and still have an impact on individuals' lives.

Examples of political organizations are states or international institutions such as the World Bank, the United Nations or the European Union. Examples of *formal* regimes are the practice of human rights or the Agreement on Trade Related Aspects of Intellectual Property Rights (the TRIPS agreement) under the World Trade Organisation (WTO), which controls among other things, access to medicines. *Informal* regimes are unregulated phenomena such as illicit financial flows, global tax competition or the escalation of trade tariffs.

Relations have different meanings according to the theoretical contexts and levels of analysis where they are used, as we have seen. Social, cultural and institutional practices affect and often shape the relations individuals can bear to one another. There are relations of reciprocity, sharing, participation, domination, and exclusion, all of which are social relations. Approaches to justice may envisage agents as being able to engage fairly in social relations and will set safeguards to prevent the rise and spread of unjust relations. But these relations are not the objects of regulations; they are 'practice-mediated', i.e. they are composed by practices. I want to suggest here that *we do not formulate principles to regulate relations as such in non-ideal theory. We formulate principles to regulate formal or informal practices in order to obtain a society that prevents the rise and perpetuation of unjust relations, and also distribute social goods to enable individuals to engage fairly in social relations.* Most relations are mediated through practices. For instance, discrimination can be the consequence of bad laws. Exclusion can result from inappropriate social safety net measures. Domination can be the consequence of missing working regulations (missing regulations would still be a practice in the sense of this thesis, considered as informal regimes that call to be formalized). Once regulated, informal regimes become at least *de facto* formal regimes. Of course, regulations do not immediately prevent unjust relations or do so on their own. The social or cultural implementation of rules through time and contingencies will condition the possibility of preventing unjust relations. I assume henceforth that it is

plausible for principles to target practices, in the sense defined above, rather than relations.

> In other words, we can say that international practices are the site of global justice.
>
> The site of justice is not the same as its scope: the site of justice refers to the *kinds of objects* (individuals' actions, individuals' character, rules, or institutions, and so on) appropriately governed by principles of justice, that is, to which the principles of justice rightly apply, whereas the scope refers to the *range of persons* who have claims upon and responsibilities to each other arising from considerations of justice. (Abizadeh 2007: 324).

Take Rawls's theory of justice. For Rawls, the site of justice is the basic structure of society, whereas the scope of justice is limited to the national borders of each constitutional democracy (or any decent society). Indeed, Rawls and Scanlon explicitly conceive different principles for different domains of justice (Rawls 1993, Scanlon 2000). In *Justice as Fairness* and in *Political Liberalism*, Rawls states that the principles that are formulated for the basic structure of society are not fit for all subjects. These principles do not apply to universities or churches, and neither do they apply to the law of peoples (Rawls 1996, Rawls and Freeman 1999: 522-523).

I believe it would be a mistake to assume that by identifying a given site of justice we would be determining the scope of justice. According to Rawls, it is the different structure of social frameworks, the role of their various components, and the way they all fit together that explains why principles apply to different subjects (or sites). That is, I am arguing that the *sites of justice are individual practices (a series of particular practices), whereas the scope of justice is global.* A non-ideal bottom-up approach therefore allows us to address the practices that cause the problems of global background injustice.

Recall one of the interesting elements of Rawls's constructivism. Rawls employs the idea of the function of justice to distinguish the domestic realm from the global realm. The role of the basic structure in a constitutional democracy is to create and distribute social primary goods within a scheme of social cooperation. The role of the law of peoples is to uphold peace, national autonomy, and assistance – not justice – between nations and to favour the maintenance of domestic justice (James 2005: 300).

Rawls is right in that agency is different at the two levels, but wrong in thinking that there is no room for justice, beyond assistance, at the global level.

3.5 Proceeding from the practice

As seen in the two previous chapters, for Rawls, the main characteristics of ideal theory are the formulation of a theory of perfect justice, under the idealized assumptions of full compliance and favourable circumstances (Rawls 1999b: 8-9). For Rawls, this inquiry has to be undertaken prior to the development of any non-ideal theory, so that it can provide a systematic grasp of non-ideal issues. To prove (i) – the claim that ideal theory does not have analytical primacy over non-ideal theory – is to contradict Rawls on this point. It is to challenge the notion that non-ideal theory is applied ideal theory.

The validation of (i) will show that principles of non-ideal theory need not be modelled according to an ideal standard. This will open the door to developing non-ideal theory independently of ideal theory and, simultaneously, to think of ideal theory as something less demanding than perfect justice theory. Amartya Sen argued that the feature of perfect justice – understood as a situation above which no principles of justice of a higher order could be formulated – made the two theories analytically disjointed (Sen 2009). My point is that even if we relax this constraint (as Chapter 2 argued we should) and accept that an ideal theory can be less demanding than perfect justice (say a theory of basic rights), non-ideal theory would still not be modelled according to this standard.[65] Claim (i) thus means that we do not need an ideal theory of any kind (as a theory of perfect justice, basic justice or some other form) to undertake the work in non-ideal theory.

To prove (i), it is important to begin by clarifying that, while they proceed from considering practices that cause structural problems of justice, the non-ideal approaches of the kind I am defending might only show concern for equality indirectly, or in time.

---

[65] I show in Chapter 2 how this connection occurs: ideal theory and non-ideal theory connect in reflective equilibrium.

For instance, principles that regulate tax competition will probably target states, financial institutions and corporations. Such principles might not directly reflect normative individualism. Nonetheless, this methodological commitment to aim at regulating practices does not deny the moral equality of persons. In other words, while individuals are the ultimate units of moral concern, principles of justice do not necessarily define entitlements or duties for each individual.

Now, we may rightly ask why do we regard a particular situation as unjust or, more precisely, what justificatory element triggers the formulation of principles of justice. Although we might not have a definitive answer to the question of what global justice requires, we could nonetheless ask another question: what normative element allows us to say that a given practice is *prima facie* unjust and should trigger a non-ideal theoretical analysis? Whilst a non-ideal theory can be pluralist about values and accommodate heterogeneous accounts of why individual practices are unjust, there might nonetheless be a single sufficiently weighty normative element that would press us to consider a regime or political organisation from a justice standpoint.

One possibility is that any institutional system or practice that has (or had) a significant, pervasive or profound negative impact[66] on individual opportunities is subject to considerations of justice. One of Rawls's reasons for choosing the basic structure as the site of justice is that the basic structure has a profound impact on individual opportunities (Cohen 1997, Abizadeh 2007). If we ground this view on the notion of pervasive impact alone, any practice that has a pervasive impact on individuals' opportunities would be a subject of justice, regardless of its function in coordinating social cooperation.

As opposed to a view based on social cooperation, a view based on the idea of pervasive impact is not limited to the participants in a system of cooperation but rather to all those significantly affected by the impacts of an institutional system. That would notably include people that might only feel the negative impact of institutions over which they have no control. It would also include people that are only affected by the

---

[66] For the purposes of this thesis, I will mainly consider practices that have a negative impact on individual opportunities.

externalities of an institutional system without ever being subject to its policies. In all such cases, agents (like states and individuals in the case of tax competition) cannot escape from the impact of these practices. This view is thus a good candidate for providing further normative grounds for a non-ideal approach to global justice (adding to the strengths of a normative theory that already considers the particularities of individual practices).

Without saying anything about the content of principles of justice, a practice that has a pervasive impact on individual opportunities triggers the formulation of principles of justice.[67] In this case, a formal or informal practice that has a profound effect on individuals' lives has to be a *pre-existing* condition for concerns of justice to arise. Thus, *on a bottom-up approach,* pervasive impact is a *sufficient* condition for considerations of justice to arise. That is because although the necessary institutions, political authority or policy tools to address a problem of justice might not be in place, the fact that there is a problem is a sufficient reason to look to establish the adequate structure to address the problem. As Ronzoni argues, we have a duty to create the background of interactions that would allow for the implementation of principles of justice (Ronzoni 2009). Contrastingly, I argue that we have a duty to create the adequate institutional scheme, or reform it if the existing mechanisms are insufficient, to address problems of justice; and that the regulation of the background of interactions does not require a pre-existing conception of justice that will be 'implemented'. The appropriate notions required to formulate principles will be given by the content of non-ideal theory and an account of the relevant practice, as I will show below.

In sum, an analysis based on the notion of pervasive impact is one way to cement the methodology of a non-ideal account of justice. Therefore, if we accept the conclusion of the last section, that international formal and informal practices are the site of global justice and that all practices that have a pervasive impact on individuals' life prospects trigger the formulation of principles of justice, we have established claim (i). Indeed,

---

[67] This is a stronger claim than just saying that these practices trigger concerns of justice. I assume that there is transitivity between these claims, in the sense the weaker claim is implied by the stronger, and the stronger claim can be made based on the weaker. Surely, we can identify injustice without formulating principles. However, in these cases, the formulation of principles would also be justified.

following section 3.3, we can claim that the scope of justice is global, but that different principles apply to the domestic and the global realms. Section 3.4 provided solid reasons to consider practices as the site of justice. And to the extent that these practices have an impact on individuals' life prospects and thereby cause ongoing problems of justice, we have reasons to address them directly and thereby contribute to the realisation of justice. This shows that practices that have a pervasive impact on individual opportunities are sufficiently weighty to be given consideration, for the particular injustices they create must be addressed. This shows that there is no reason not to address these practices individually. No prior conception of ideal theory is necessary to do so. Ideal theory is thus not analytically prior to non-ideal theory. The moral assessment of practices and their regulation based on considerations of justice need not be done in reference to an ideal theory of justice. I stated above that claim (i) aimed more at cementing the methodology of non-ideal theory that its content. One could reply that any theory of this kind will not necessarily be non-ideal in content. Perhaps this is right, and following Chapter 2, the level of idealisation of any theory of this kind can only be given through an exercise of reflective integration. We may very well admit this and the idea would remain: ideal theory does not have analytic primacy over non-ideal theory.

In the remainder of this section, I will reply to one objection that might seem to threaten both (i) and (ii). This will support claim (ii), which provides further grounds to a position that addresses practices individually.

The objection challenges the project of a non-ideal theory of justice in both its content and methodology:

> To assign any greater role to institutions and practices—to allow them, as I have said, to influence the formulation and justification of first principles of justice—is a fundamental mistake: constraining the content of justice by whatever social and political arrangements we happen to share gives undue normative weight to what is, at best, merely the product of arbitrary historical contingency or, at worst, the result of past injustice itself. (Sangiovanni 2008: 137. Sangiovanni rejects the objection, but not for the same reasons I provide here.)

'First principles' of justice are principles that are not themselves derived by applying other more fundamental principles to particular circumstances (Sangiovanni 2008:

137n1). They can be derived from facts about human nature or from moral values (such as fairness or respect), but not from other principles. Non-ideal theories including the one I am defending claim that principles should apply to practices individually. Moreover, the method of reflective integration, which I presented in the last chapter and I will defend in Part II, denies that there are any first principles in the sense defined by Sangiovanni (I will argue below that although we may accept the existence of 'first-order' duties, it does not imply we need to define 'first-order' principles).[68] I argue that the nature of the practice will influence the formulation of the principles. The objection above rejects this methodology, claiming that the nature of practices should not play any role in the formulation of first principles of justice.

The reply to this objection is central to the project of non-ideal theory. I argue that *it is wrong to claim that the consideration of practices at the level of justification implies giving undue normative weight to contingent products of history. It is quite the opposite. By doing so, we give ongoing unjust practices and past injustices the normative weight they require.* Colonialism, access to medicine or tax abuses are all past or ongoing practices that should be subject to considerations of justice.[69] Addressing them as such is to show equal respect to all those that suffer (or have suffered) from the injustices sustained by these practices. Responding to particular injustices and making sure our principles reflect the specificities of these injustices is perhaps to give them the weight they command in order that they may be prevented in the future and for compensation to be given for the past.[70]

---

[68] One idea that could potentially be explored is the following: first principles are formulated in ideal theory and second-order principles are formulated in non-ideal theory. I am ruling out this idea. I wish to explore the plausibility of a theory that is thoroughly non-ideal and for I believe the reflective equilibrium methodology overcomes the need to distinguish between first-order and second-order principles. Although I will distinguish between first-order responsibilities and second-order responsibilities below, I believe it is a mistake to insulate one level of principles from real-world institutional and practical considerations.

[69] Ideal theorists can say that ideal theories do this and that their theories can be adapted to different circumstances. However, by showing (i), I argued that ideal theory is not analytically prior to non-ideal theory and I am now giving reasons for why the non-ideal theory methodology is right in proceeding from ongoing unjust practices.

[70] This provides a better understanding about why non-ideal theory is reparative, comparative and transitional. Chapter 4 will explore in further detail its comparative and transitional aspects.

In light of this, it is odd that the burden of proof about analytic primacy is taken to lie on those who try to address these injustices directly and individually.[71] We may pursue this reasoning and look for other reasons to tackle these practices individually. A good candidate is agency: the different demands made on different agents in different cases. Justice with respect to colonialism requires present agents to attempt to compensate for the wrongs of past agents. Climate justice demands that present agents to attempt to remedy for the wrongs of past and present agents in order not to undermine the life prospects of future agents. Tax competition involves international institutions, states and corporate agents, that may have nothing to do with the two problems just mentioned. These examples support claim (ii): that we should formulate different principles for different practices. Indeed, as the rest of this thesis will confirm, the formulation of action-guiding principles of global justice will more likely, and perhaps will only, be determined by an analysis of these practices individually. This reinforces the argument against the analytical primacy of ideal theory – claim (i) – and surely suggests that the set of principles that will target each practice will be formulated for these practices individually – claim (ii). Although we had sufficient practical reason to address these practices individually, such as to give past injustices sufficient consideration, we also have major conceptual reasons, like the differential demands of agency, to do it. I believe agency is a complex enough case to support claim (ii) on its own, for it involves: the temporal action of agents (past, present, future), the nature of agents (individuals, states, transnational institutions) and the nature of actions (harm reduction, distribution, background regulation, compensation) among other problems. In Part II, and as mentioned above, we should see other reasons in support of (ii) such as that an analysis of the practices might reveal the different nature of the problems, for instance, when we observe that tax competition is best described as an 'asymmetric prisoner's dilemma' (in Chapter 5) case and the problem of climate change another type of collective action problem (in Chapter 6).

---

[71] Moreover, as we will see in the rest of this thesis, I will show that there is value in philosophical projects that are not about looking for simplicity and higher levels of abstraction.

Claims (i) and (ii) consolidate the methodology of non-ideal theory: non-ideal theorising proceeds from existing practices, addresses them individually in a bottom-up approach and does not presuppose an ideal theory.

However, views that limit their scope to particular regimes or institutions (or even relations more generally) must consider a second objection: in the case of a great inequality of resources or wealth, if two agents have no relation with each other, no considerations of justice arise. Consider the following analogy.[72] The society on island A lives in affluence. They share an abundance of natural resources and they have the potential of converting their resources in the necessary goods to live a decent life. The society on island B lives in poverty. They do not lack resources to the point of starvation, but are unable to develop the necessary infrastructure to build a sustainable and fair economy. Now, assume that although each society is aware of the other's existence, they have never interacted with each other. Does society A have duties of justice towards society B? In the view defended in this chapter, it does not. As mentioned above, principles of justice can only apply between agents who stand in a special relation to each other, one mediated by practices that have a profound impact on individual opportunities.

Society A may have duties of assistance towards society B, but not duties of justice. A non-ideal bottom-up account of global justice should be satisfied with this response. There are cases where moral obligations are based on duties of assistance. The example of the two islands illustrates this. It is possible to claim that Society A has moral duties of assistance to give a portion of its resources to Society B in order to enable society B to develop a sustainable and fair economy, if it could do so without imposing a too great sacrifice on its citizens. In this case, principles of assistance apply, but not principles of justice. If this seems counterintuitive, then it is simply a bullet that bottom-up non-ideal theorists will have to bite. In the case of the two islands, only principles of assistance obtain, not principles of justice.

---

[72] I take this analogy to be distinct from one that compares societies by playing with the number of people in each. Derek Parfit has used analogies of this second kind, see Parfit, D., 1984. *Reasons and persons,* Repr. with corrections. ed. Oxford: Clarendon Press. I take my analogy to apply regardless of the number of people involved.

Assume now that these two islands are in the same conditions described above, with the difference that the two islands have interacted for quite some time. During part of this time, institutions in society A made it possible for the wealthiest members of society B to shift their fortunes away from B, in order for them to enjoy vacations and leisure outside island B in exchange for minimal fees, thus causing funds once raised by taxation to be severely diminished. Society B might have been poorer from the beginning, but now we can point out that the interaction with society A has contributed significantly in deepening society B's poverty. In this case, society A would have duties of justice towards society B, because a shared practice is the source of an injustice. In a non-ideal approach, this situation would need to be addressed from the standpoint of justice. Nonetheless, the principles to address this situation will be formulated to address this practice alone. The practice will inform the nature of the principles that will be formulated, as we shall see.

It is now time to turn to the content of non-ideal theory. The aim of the non-ideal theory of the kind I have in mind is to formulate a conception of justice that is better designed to address injustices of the world today and to regulate the interactions of agents in our non-ideal world.

3.6 The content of non-ideal theory and the question of compliance

The recent debate on ideal and non-ideal theory has revealed different ways in which we can interpret the action guidance paradox of ideal theory. Important contributions focussed on how it is possible to respond to it from the standpoint of ideal theory (Robeyns 2008, Swift 2008, Valentini 2009, Hamlin and Stemplowska 2012). Chapter 2 pointed to the limits of these kinds of responses. This section finally opens the door to a more promising approach, by outlining how a non-ideal theory can respond to the paradox. It explores the content of non-ideal theory, a task that will be continued in subsequent chapters.

For Swift, as we saw earlier, action guidance means 'knowing what options are feasible, over what time scale, with what probabilities, given where we are now' (Swift 2008: 374). These are indeed important issues, with each deserving careful examination. In this section and the next I will examine the notions of compliance, feasibility, fact-sensitivity and path-dependence of a given institutional proposal. These are four components of action-guidance in a non-ideal theory of institutional design. These analytical tools allow us to gain a better grasp of the nature of an action-guiding theory. Yet, contrasting with Swift's definition, I wish to show that a proper understanding of action guidance should reveal the normative contributions of non-ideal theory, and not simply non-normative complementary elements from social sciences.

I do not pretend to put forward here a thorough theory of action-guidance; I merely hope to take a step in that direction. By exploring these components of action-guidance, I aim to give more substance to the four claims that together comprise the *Non-Ideal Theory Thesis* regarding: analytical primacy, bottom-up reasoning, path-dependence and compliance in non-ideal theory.

Non-compliance

Non-ideal theorising is commonly understood as a matter of dealing with partial or complete non-compliance with a given set of idealised principles (Murphy 2003, Caney 2015). One of its central questions is whether our duties of justice change in situations of partial compliance (Murphy 2003). The pivotal issue of this debate is to determine whether or not moral principles should increase their demands on agents as expected compliance with the principles by other agents decreases (Murphy 2003: 77). This is an interesting question and I will return to it, but it is not, I want to argue, the question from which we should begin our analysis of compliance in non-ideal theory, considering we should be looking to bridge the gap between ideal and non-ideal theory. One important reason to depart from this interpretation of the question is that, as claim (i) already made clear, we do not need to assume in advance that we know, or that we need, an ideal theory of justice with which we must comply in the real, non-ideal world. I propose therefore an alternative interpretation: one that centres on the questions of

compliance with action-guiding non-ideal principles and of path-dependence. This leads us to claim (iii): action-guiding principles are not given by determining whether our duties change if agents do not comply with ideal principles, but rather by an appropriate understanding of the different ways to seek compliance with non-ideal principles. This radically changes how we should understand the central question of non-ideal theory, and it does so by emphasising that we should not abstract away from the actual causes of injustice (James 2012: 13).[73] That agents fail to comply – or even worse, create and exploit a system of non-compliance as often they do in the case of tax competition – is the heart of the matter. Whether or not agents comply with a principle of ideal distribution is secondary to the questions of what the causes of injustice are and of how agents could comply with a principle that targets this injustice specifically.

Consider the case of tax competition, to which I will return in detail in Chapter 5. In a context of tax competition, some countries have tax jurisdictions that allow for companies and individuals to pay taxes at very low rates, thereby depriving (lawfully or unlawfully) the countries in which they operate of important revenues (Dietsch and Rixen 2014). There are agents involved in this system – corporations and individuals that seek to abuse low-tax jurisdictions, banks that contribute to account secrecy and countries that offer low tax rates – that do not comply with rules already in place in their own tax jurisdictions. Compare the case of global agreements on intellectual property rights (the TRIPS Agreement under the World Trade Organisation, WTO) and the problem of tax competition. The TRIPS Agreement systematically prevents people from having access to life-saving treatments which would have been affordable under a feasible alternative system (Pogge 2008: Chapter 9). We may rightly assume that this practice should be subject to considerations of justice. The case of global tax competition is not analogous to the case of intellectual property rights in one important respect: there is no global authority that regulates tax competition. There is no international governance body. The structure in place, the Organisation of Economic Cooperation

---

[73] In *Fairness in Practice*, Aaron James points out to the dangers that abstracting away from the existing social relations could obscure the question of how distinct fairness responsibilities could emerge from these relationships.

and Development (OECD), is a coordinating body. In the case of agreements on intellectual property, there is an authority, the WTO, that regulates the TRIPS agreement (although some may argue that its rules are unfair or that some rules have unfortunate consequences).

An adequate understanding of the practice of tax competition shows that the laws and regulations already in place are insufficient to tackle the problem (Rixen 2011a, Dietsch and Rixen 2016; I return to this in Chapter 5). International tax rules are insufficient and there is a lack of compliance with the few rules already in place. Furthermore, the problem of tax competition can be interpreted as a collective action problem (Nili 2014, and below, in Chapter 5, I explore in more detail what kind of collective problem this is), in which no agent has a moral duty to act alone. This thesis presents the view according to which finding the mechanisms to ensure compliance, such as with transparency agreements, is a step towards facilitating collective action, notably because multilateral governance will give individual agents reasons to believe that other agents will act appropriately.

Massive tax abuses are made possible through legal loopholes, missing laws and an absence of regulatory bodies. To address this question effectively, a non-ideal theory tackling tax competition will ask what regulations make agents comply with fair norms of tax interactions and what specifically exacerbates the problem of non-compliance. *It is not to ask what agents should do if other agents did not comply with ideal principles of global justice. The question of compliance in non-ideal theory is about what needs to be done for agents to comply with principles.* These principles will be given by an account of the practice that reveals notably which of its features influence compliance. Seeking compliance is an unavoidable aspect of the problem for any transitional non-ideal theory that seeks to move towards more desirable social settings. If the new regulations are not aimed at what can reasonably be agreed upon by agents in an adequate timeframe, we will not have an adequately action-guiding non-ideal theory of justice. The set of rules we come up with in this way will be non-ideal in content because only by addressing non-compliance with regulations in their formulation will they be feasible and desirable. For instance, an account of the practice of tax competition should explain

among other things how the existing laws and legal loopholes facilitate the immoral behaviour of global agents and whether the consequences of tax competition affect states differently depending on their size and wealth. In order to guide action in the real world, a theory of justice tackling tax competition must understand these features of the problem. These are internal constraints of the problem that influence or explain the behaviour of agents and therefore must be understood for the problem to be effectively addressed. In sum, an adequate understanding of the question of compliance in non-ideal theory is not assessed by the extent to which agents fail to comply with ideal principles. This shows that claim (iii) is sound.

Now, let us consider again the standard interpretation of the question of compliance, but apply it this time to non-ideal principles. This means asking whether non-compliance affects the demandingness of the associated non-ideal duties of justice. In other words, if agents fail to comply with the principles of justice, do the complying agents have to compensate for this lack of compliance? Assume that we already have the set of non-ideal principles that will regulate tax competition, that this set is one with which we can reasonably expect compliance, and that a group of agents are not complying with them. Does non-compliance with non-ideal principles affect the demandingness of the duties of justice? I will make two remarks, which should help in answering this question, as well as further elucidating the question of compliance in non-ideal theory.

First, one may argue that it is easier for agents to comply with a specific regulation than with a demanding cosmopolitan conception of distributive justice, such as a demand for global equality of resources or a global difference principle, although this does not diminish the moral obligation towards morally justified principles. Consider, then, an ideal but less demanding conception of global justice, an ideal that does not aim at perfect justice, such as one based on the respect for human rights. Human rights constitute a recognized and powerful set of rules by which we assess international behaviour and which provide reasons to act, for instance by triggering national obligations to reduce the magnitude, pace and impact of climate change, by assessing countries' socio-economic performance or by decrying the political oppression of a

group in a given country. How do human rights fare in comparison to a more specific set of rules targeting tax competition (or intellectual property rights), in terms of inducing compliance? I argue that, where human rights prove to be ineffective, a non-ideal theory should take over. If pointing to the violation of human rights will not make agents cease harmful tax competition, a non-ideal theory should provide principles that will target the real-world circumstances that are at the heart of the problem of non-compliance. Indeed, a principle that will regulate tax competition and illicit financial flows will need to include notably aspects of transparency and information exchange, just as a principle that regulates intellectual property rights and patents on drugs will need to be sensitive, among other things, to the incentives that the pharmaceutical industry requires to pursue research. In the two cases, what will induce compliance and favour the assessment of non-compliance is more specific than the most detailed conception of global justice based on human rights. This suggests that, with regard to many problems of global justice, the problem of non-compliance will be more effectively addressed by a set of non-ideal principles than by a set of ideal principles, again supporting (iii). The interpretation of the question in non-ideal theory is therefore more likely to be useful for a theory of transition.

Also, we must work out whether or not the demandingness of the duties will vary according to the compliance of agents. The duties associated with the real-world problems mentioned so far are duties to implement just practices and just institutions to regulate a certain sphere of activity. Both principles of background regulation and principles of distribution of wealth or resources imply political will and monetary interventions at different levels. The demandingness of the duties to create and uphold just institutions (measured by the strictness of regulations and sanctions) cannot be measured in the same way as the demandingness of the duties to effect a distribution (measured in a given currency of justice). Whether duties involving less distribution and more political traction are more likely to succeed is contingent on time and context and is, at this point, a matter of speculation. However, it is clear that with a fairer background of global interactions there would be less need for redistribution. This strengthens the case for (i), (ii) and (iii) – since formulating duties to regulate practices

individually diminishes the causes of injustice in the first place, the duties of redistributive justice might also become less demanding and therefore draw more political traction.

In sum, determining whether duties of justice change in light of non-compliance is not of primary importance in non-ideal theory. The question of compliance in non-ideal theory is thus one about what to do to ensure compliance with non-ideal level principles and how to appropriately judge and sanction the non-compliant.

3.7 Fact-sensitivity, feasibility and path-dependence

I pursue the examination of the content of non-ideal theory by outlining the notions of fact-sensitivity, feasibility and path-dependence, which should together inform a more thorough conception of action guidance.

Fact-sensitivity

There are two ways to understand fact-sensitivity in the context of the formulation of principles of justice. One way to interpret the notion is to choose to reflect, or to not reflect, choices and circumstances in the formulation of the principles. For instance, consider the example of providing the surfers of Malibu with minimal income in Van Parijs's guaranteed basic income theory (Van Parijs 2004). This theory is not sensitive to the individual contributions in the creation of wealth in society. A theory sensitive to this fact is more likely to make allocations of income contingent upon actual contribution, and so would choose not to compensate the surfers of Malibu. This is one way (a Dworkinian way) to understand the idea of fact-sensitivity: to determine what facts should be reflected in the formulation of principles (Dworkin 2000). Fact-sensitivity in this sense ultimately aims to determine who is entitled to the allocation of goods in society.

A second way to understand the idea of fact-sensitivity is to attempt to model the potential outcomes of a given policy, considering where we are here and now. For

instance, one idea sometimes invoked to address the problem of tax competition is to harmonize tax rates between countries. The harmonization of tax rates diminishes tax competition and allows countries to increase their revenues. However, in the long run, "it tends to deprive them of the capacity to react to specific circumstances in a sufficiently swift and context-sensitive manner" (Ronzoni 2014: 43). It can, for instance, prevent "governments from raising taxes in response to rising spending requirements and from detaxing labour in response to growing unemployment" (Genschel 2002: 245). In this second sense, when designing the principles to tackle the problem of tax competition, the idea of harmonization needs to be understood in the real world context, where governments require a certain degree of flexibility in the design of their fiscal policies.

Being sensitive to facts in this second sense is a central aspect of work in non-ideal theory. *Non-ideal theorists must test principles of justice such as to model whether they will have undesirable consequences in the world today. To do so it has to be sensitive to a set of relevant facts.* To know whether egalitarian principles will have desired or undesired effects, one must try to understand what policies could best model this principle, under defined institutional conditions and in a given political culture. Testing implications of principles in ideal theory is of limited value, is imprecise, and does not easily guide action. If to build our solution to tax competition, we abstract from facts of economic policy such as a context of unemployment, we might not develop the best policy in real world context. Fact-sensitivity in the sense described above is thus central to the construction of an action-guiding theory. The introduction of fact-sensitivity of this kind is likely to be specific to non-ideal theory.

Feasibility

Fact-sensitivity concerns how a theory integrates facts about the real world into the formulation of its principles. Feasibility is a matter of how well a theory, already written, is applicable in a determined context. Fact-sensitivity looks inwards. Feasibility looks outwards. Recent contributions to the ideal/non-ideal theory debate have distinguished between feasibility levels that correspond to more and less ideal theories. It

has been argued, notably by Gilabert and Ypi, that feasibility should not be understood in the same way in ideal and in non-ideal theory (Ypi 2010).

We can, following Pablo Gilabert (2012: 50-51), distinguish three levels of feasibility. The first level is *theoretical feasibility*. One example is what Rawls calls the 'circumstances of justice'. Circumstances of justice consist in moderate material scarcity and the existence of conflicts of interests. Theoretical feasibility may demand that, when formulating principles of justice, we acknowledge that conflicts of interest exist, but does not require us to address particular conflicts. The second level we may call *institutional feasibility*. Some theories of justice will choose to consider the social and political context, culture and history in order to formulate their principles. For instance, in North America and most Western countries, the eradication of a market economy, if ever necessary in order to advance justice, is not likely to succeed. The principles that will regulate social institutions must take into account that the economies of these societies, to some extent, will be market-based in the foreseeable future. The third level is that of *political feasibility*. A theory that considers feasibility issues at this level is a theory that tackles actual issues about socioeconomic and political policies, such as specific institutional weaknesses and failures, missing or bad laws, and the short-term motivational problems that may prevent the progression of justice.

Political theories can engage with feasibility at all three levels. What is important for non-ideal theories is to be explicit in their inclusion of feasibility considerations. This is less important for ideal theories. Ideal theories need less to engage with the second and third levels, while non-ideal theories would have to engage at least with second level feasibility. This is not to say that the normative value of ideal theory is undermined by the failure to address feasibility issues that are not relevant to its levels of theorising. Ideal theories may be criticized on many grounds, but objections on the grounds of feasibility are normally not the strongest (as seen in Chapter 2). Most feasibility objections should rather be directed to non-ideal theories.

Although they may be explicit about constraints of theoretical feasibility, ideal theories are generally not judged upon whether they are feasible here and now. Rawls thought his theory had to be realisable in a feasible and lasting social world (Rawls

1999b: 153), given what we know about human and social limitations. In the original position, parties have to take into account whether different systems are achievable and sustainable, and whether they will likely draw compliance. These are important feasibility considerations, but not of the kind that presumes a theory to be 'feasible here and now'. What can possibly be said about feasibility here and now needs to proceed from the actual state of institutions, actual legislation and actual impediments on the realisation of justice. The formulation of principles in abstraction, even in a carefully thought-out conception of justice such as Rawls's original position, is a different exercise than one aiming to formulate guidelines in order to influence policy in the world today.

If this reasoning is right, it confirms an idea about why non-ideal theory is more likely to guide action. An ideal theory needs not find ways to design policy here and now, simply because it does not have that function.[74] This implies, as mentioned, that action-guidance is mainly a question for non-ideal theory, understood as a theory aimed at the design of ongoing practices (political organisations, formal and informal regimes) in a reparative, comparative and/or transitional way. I will return to this point in Chapter 4. Feasibility in non-ideal theory is intrinsically concerned with action guidance. This implies that considerations of institutional design are more demanding than what is suggested above by the tripartite distinction put forward by Gilabert. It is not sufficient to take into account the institutional contexts of implementation we live in, but understand the functioning of these institutions to know what is wrong with them and what prevents an effective combat of injustices.

Path-dependence

A final element of non-ideal political theorising, which we should separate from questions of fact-sensitivity and feasibility, relates to claim (iv) about path-dependence. As normally understood in political philosophy, the choice between two institutional settings is made on the basis of its approximation to a social ideal or objective. The likelihood to lead to a given ideal should inform the choice between two institutional

---

[74] But this does not imply the opposite. Non-ideal theory has the capacity to influence the formulation of principles. I develop this elsewhere.

proposals. As I mentioned in section 1.6, the notion of path-dependence is best modelled by its understanding in economics and social sciences than by what is described by the path-dependence argument. Ypi, Goodin and Barry refer to works in economics and social sciences to provide a better definition of the notion.

> Path-dependent theories emphasize, firstly, that founding moments of institutional formation may lead to entirely different developmental paths and, secondly, that institutions further evolve by responding to changing societal conditions in ways that reflect those past trajectories. (Ypi *et al.* 2009: 128).

This will allow me to provide support for claim (iv). In this section, I argue that path-dependence informs what Gilabert calls 'dynamic duties' to support one institutional setting rather than the other, on the grounds that it is the setting that could help bring about the ideal institutional order (Gilabert 2012b: 4.2). We should see that the ideal institutional order is not the guide, or at least not the sole guide, in choosing which institutional settings to promote.

Consider the challenge for governments motivated to raise the tax rates of big corporations operating in their jurisdiction in order to pursue a distributive ideal of greater equality of opportunity, in a context where increases in tax rates may result in a wave of corporate relocations. In today's context of global tax competition, the various forms through which companies can find more attractive tax jurisdictions support the claim that the fear of relocation is at first glance legitimate. The extent to which this fear is legitimate will be determined by the targeted sectors of activity, dependence on qualified labour on site, the social image of the companies within the political community, and the local tax laws. Moreover, as will be shown below, tax competition today allows for different forms of tax abuse, such as trade mispricing, which do not even require corporate relocation. The non-ideal context of global tax competition therefore prevents governments from raising taxes as they see fit. Precisely because of non-ideal circumstances and empirical features of the world, the institutional choice that best approximates the ideal at the level of ideal theorising might not be a choice that promotes social justice. In a non-ideal context, sometimes, path-dependence between a given policy and an ideal objective can only be determined through non-ideal level

reasoning.

This leads to an important point in the discussion of path-dependence. Consider again settings that are not available today but might become available in the future, once the feasibility constraints relax. Some unfeasible but desirable levels of taxation may be attainable in the future once feasibility constraints relax, that is, once tax competition is tackled. Tackling tax competition is thus a necessary step in the realisation of social justice. For potential tax rates and the potential promotion of social justice to become available, tax competition must be tackled. *In order words, we need principles to regulate this practice in order to know what potential revenues and social options will be feasible in the future. This implies that an ideal social order depends on paths paved by non-ideal world regulations. But the converse implication does not hold. The path for non-ideal principles is not given by ideal principles.* In Gilabert's words, we might have dynamic duties to bring about a scenario where a more just world could be implemented, but the exact definition of this ideal is not required to formulate the non-ideal principles to tackle a real world injustice. The content of a non-ideal theory is not path-dependent on the content of an ideal theory. This suggests that claim (iv) is sound. Path-dependence is best determined by non-ideal theorising. Chapter 4 returns to this reasoning.

3.8 Conclusion

Considerations of justice aim among other things at enabling individuals to enjoy their right to freedom. Whilst freedom is mostly realised at the domestic level, within the boundaries of a political community, international interactions and institutions undermine individual's capacities to act upon their plans of life. If such institutions stopped interfering negatively, this would already have immense positive effects. Regulating the problems that are particular to the global level should have an immense impact on individual opportunities in different nations. And the redistributive duties between nations would be considerably lessen in an international context where practices such as tax competition are effectively regulated.

The main task of this chapter was to provide an account of a non-ideal theory of global justice in practice, exploring its methodology and content. It has opened the door to a conception of justice based on the ideas of non-compliance, fact-sensitivity, feasibility and path-dependence. It did so by arguing for four claims (all four claims will receive further support in the empirical analysis of Part II). Sections 3.4 and 3.5 supported the case for (i) and (ii): ideal theory is not analytically prior to non-ideal and a theory of global justice is a pluralist theory pursued by bottom-up contributions. I argued for claim (i) by showing the contributions of a theory of justice that proceeded from the practice. A practice that has a pervasive negative impact on individual opportunities is a sufficient condition for considerations to arise and a theory that proceeds from such practices is likely to give to past and ongoing injustices the normative weight they require to be effectively addressed. The examination of the theoretical value of a theory that proceeds from a specific practice that is the very cause of injustice led to showing (ii). Although giving past injustices enough consideration is a sufficient practical reason to address practices individually, we also saw we have major conceptual reasons to address them in bottom-up contributions, such as the differential demands of agency and the nature of the problem given by the analysis of each practice. The differential demands of agency vary according to the temporal action of agents (past, present, future), the nature of agents (individuals, states, transnational institutions) and the nature of actions (harm reduction, distribution, background regulation, compensation), which suggests that principles should be tailored to individual practices. Furthermore, the analysis of the practice provided key insights about the nature of the problem (such as about the kind of collective action problem faced) which decisively influences the form any response should take. Clearly, theories of global justice should be open to pluralism at the level of principles and to bottom-up approaches.[75] Part II will provide further support to these claims.

Finally, sections 3.6 and 3.7 provided support to claim (iii): determining our duties of justice in a non-ideal world is not a question about non-compliance with ideal

---

[75] Perhaps this is so also at the level of values, but this is beyond the scope of this thesis. Part II will return to this point very briefly.

principles but rather about compliance with non-ideal principles. This shift from the standard interpretation is warranted. Avoiding non-compliance is at the very core of non-ideal theorising. To sidestep the question of compliance by asking whether our duties are affected in cases of partial compliance is to misunderstand the role of non-ideal theories. The non-ideal theorist should aim to bridge the gap between the values underpinning our moral assessment of contemporary problems of justice and the development of regulations that can provide genuine guidance to agents living in the real world as it is today. Regarding claim (iv), we have seen that it is through non-ideal theorising that we can determine how feasibility constraints can relax and therefore what options become available in the future. Claim (iv) will be developed further in Chapter 4, especially with regard to the choice between available options and paths towards more desirable settings.

4. TRANSITIONAL THEORY: CONNECTING IDEALS WITH POLITICAL ACTION

4.1 Introduction

From what we have seen in the three previous chapters, it is hard to argue that normative inquiry is corrupted by practical concerns. We have shown the value of works that are sensitive to practical considerations. And we have tried to model the kind of work that can translate between theoretical considerations and practical problems. This chapter pursues the latter task. It aims at understanding the relation between institutional proposals that are desirable and achievable today and ideal institutional schemes that are not available to us in the near future. It will do so notably by clarifying how non-ideal theory relates to the theory of second-best, by casting light on why non-ideal theory is an action guiding theory of 'transition', and by highlighting the role of the political philosopher at the gates of social science research.[76]

These various contrasts between non-ideal theory and other perspectives will cast light on how non-ideal theory embodies a theory of action guidance. This chapter focuses on what is, positively, an action-guiding theory, i.e. what guides action and not what fails to do so. In chapter 2, after I presented the paradox of ideal theory, I argued that we first need a better definition of ideal and non-ideal theory, which I explored. Then, in the end of chapter 2 and in chapter 3, I attempted to show how non-ideal theory could overcome the paradox. In this chapter, I pursue the exploration of an action-guiding political theory. I will then be in a position to situate the contribution of

---

[76] While the last chapter showed that non-ideal theory is reparative and comparative, this chapter will insist more on how non-ideal theory is transitional. As mentioned, I take these to be three components of non-ideal theory.

my account as a theory occupying the frontier between political philosophy and social sciences.

The chief questions of this chapter are the following. Can aiming at long-term results take us away from short-term gains? Conversely, can short-term results take us away from the path to a more ideal society, or in other words: how can we make the difference between short-term results that are genuine good improvements and short-term results that might undermine the path towards a more ideal society? In short, I argue in this chapter that the way to understand the 'transition' between non-ideal circumstances and ideal institutional settings is not given by ideal theory. Non-ideal theory fills this gap more convincingly. This will provide further insights into a chief notion in this thesis: path-dependence. By understanding the relation between short-term choices and long-term objectives, we will be in a better position to understand to what extent more ideal paths depend on less ideal choices.

The first section of this chapter examines the theory of second best, which illuminates certain aspects of the ideal/non-ideal theory debate (4.2). This section will help frame the debate and rule out certain misconceptions about the interpretation of the question of transition between ideal and non-ideal theory. Sections 4.3 and 4.4 will wrap-up the discussions of path-dependence and transition. Sections 4.5 and 4.6 will put this reasoning into perspective, with regard to history and social sciences. This should cast light on the role of the political philosophy at the gates of social sciences.

4.2 The Theory of Second Best

This section aims at understanding how theories of the second best help inform the question of action guidance. One of the central issues with second best theories is that they have been interpreted in the political philosophy literature as presenting an important challenge to the idea of ideal theory 'as target'. The original formulation of the theorem comes from Lipsey and Lancaster (1956):

The general theorem of the second best states that if one of the Paretian optimum conditions cannot be fulfilled a second best optimum situation is achieved only by departing from all other optimum conditions (Lipsey and Lancaster 1956: 12)

An allocation of economics goods is said to be Pareto efficient if and only if there are no transactions to be made that would increase at least one person's individual utility without also decreasing the utility of another (Wiens 2016: 134).

The three first questions to be addressed here are: how to interpret the theorem; whether the 'optimum conditions' of the theorem are analogous to the kind of state of affairs described by ideal theory; and finally what kind of challenge this poses to ideal level theorising. I leave the second question aside for a moment. I assume here that normative political analysis benefits from the analogy between ideal theorising and the sort of maximization exercise presupposed by the theorem of second-best. I will return to this second question by asking whether the decision procedure set by the second best theorem informs the decisions that are made in non-ideal theory. I will begin by answering the first and third questions.

Regarding the interpretation of the theorem in political philosophy, what is normally understood is that: if one of the background assumptions about social conditions of ideal theory does not obtain in the real world, then the best action under the circumstances might not necessarily be one that would follow the precepts of the ideal principles that would characterise a just society (Goodin 1995b).

For Wiens, this is not entirely right. The theorem applies not when the assumptions fails to obtain but rather when one of the principles that characterises an ideal society fails to be satisfied. That is, if one of the principles that characterise the just society did not obtain, then the best available option under the circumstances would not necessarily satisfy the remaining principles (Wiens 2016: 133). What the theorem is meant to show is that if one optimum condition is not satisfied, then the 'second best optimum' will not necessarily fulfil the ideal setting as much, or as far, as possible. My goal here is less to advocate for one interpretation than to examine what reasonable implications to the ideal and non-ideal theory debate we can draw from them.

The third question I formulated above is whether the theorem challenges ideal theory 'as target', the view according to which ideal theory requires us to satisfy the principles of justice that characterise the ideal society as much as possible. For Robert Goodin, according to the first interpretation of the second best theorem, the theorem applies when we cannot realise some of our basic moral and social values (e.g. liberty, security, equality). When one of these cannot be realised, the theory of second best "warns us against assuming naively that it is better to implement more of our [values] rather than fewer (or indeed to implement each of them to a greater rather than lesser degree)" (Goodin 1995b: 54). If we cannot realise all these values simultaneously, we should not seek to realise as many values as we can.

For Wiens, this should not trouble ideal theory 'as target'. What Wiens in fact argues is that a proponent of ideal theory as target should reply that ideal theory is precisely motivated by the awareness that all values cannot be realised simultaneously (Wiens 2016: 136). And ideal theory normally puts forward ways to determine the relative weight we should attribute to different values. Ideal theories tell us how to balance values (Gilabert 2012a, Gilabert 2012b). Rawls' reasoning for the lexical ordering of his principles is a good example. This way of applying the theorem does not provide a convincing critique of the claim that ideal theory can act as a target.

If we transpose the theorem to the institutional design level, the argument says that if we cannot fully realise an ideal institutional scheme, it is not clear that we should aim to get as close as possible to the ideal. Yet, we can also transpose a similar response than the one given above, to the institutional level. That is, in non-ideal theory we should not try to implement the ideal institutional setting, but the non-ideal institutional setting that best realises the ideal principles. Valentini argues that it is not inconsistent to aim at using ideal principles even if we are not aiming for an ideal setting, "because normative principles do not have any particular institutional implications" (Valentini 2011b) and thus could be satisfied in non-ideal theory by different schemes.

This is very interesting. The implication of these two replies is that there is a

functional connection between ideal theory 'as target' and ideal theory 'as measure'.[77] The measure function is a way to take us closer to the ideal target, even if we do not try to realise the ideal target directly. However, as seen in the last two chapters, I indicated that using ideal theory 'as measure' is not as straightforward as it seems. I return to this here. I wish to show that ideal theory 'as measure' is not what allows us to connect action in non-ideal circumstances with ideal theory 'as target'. We need a better theory of 'transition'.

Recall now the discussion of idealised assumptions in chapters 1 and 2, which I will connect with the interpretations of the theorem by Goodin and Wiens mentioned above. We have seen that making idealized assumptions does not undermine ideal theorising *per se*. There are good idealised assumptions that can be dealt with in non-ideal theories as long as the ideal theory does not mischaracterise the object the principles they are associated with are intended to regulate. However, although good idealisations do not undermine ideal theorising, they do not say much about the action guiding capacity of ideal principles. The second-best theorem warns us against idealisations that do not obtain, in the first interpretation. But more importantly, we know that regardless of which interpretation of the theorem is correct, we should be distrustful of idealisations that cannot be addressed in non-ideal theory.

Moreover, the theorem does not say anything about a duty to create the conditions for some institutional background structures to obtain. Perhaps we have a duty to create the background of interactions that would allow for the implementation of ideal principles (Ronzoni 2009). This is compatible with Wiens being right and the theorem applying only to outputs. It is also compatible with Wiens being wrong and one assessing what assumptions can potentially be addressed in non-ideal theory. That is, the question about assumptions is truly a question about whether idealisations can be addressed in non-ideal theory and whether addressing them will leave us in a better position to implement ideal principles. Whether or not we have a duty to implement

---

[77] Recall these two functions of ideal theory, which are respectively: to provide the target we should aim at when doing non-ideal theory and to provide the measure according to which we can know to what extent a given situation deviates from the ideal.

background conditions for more just institutional settings is not entirely given by the second best theorem. If anything, the theorem really only applies to top-down ideal theorising.

Let me pursue this reasoning by formalising this discussion. The argument from second best theory against ideal theory 'as target', in the first formulation, is as follows (Wiens 2016):

(P1) We should not necessarily aim to satisfy all principles of justice of ideal theory if a background condition assumed in the formulation of the principles does not obtain.

(P2) A given background condition does not obtain in a given case.

(C) We should not necessarily aim to satisfy all principles of justice of ideal theory in this case.

Wiens' interpretation around the idea of outputs gives us the following argument instead:

(P1*) We should not necessarily aim to satisfy all principles of justice of ideal theory if a principle of justice of the ideal set of the principles does not obtain.

(P2*) A given background condition does not obtain in a given case.

(C*) We should not necessarily aim to satisfy all principles of justice of ideal theory in this case.

In fact, Wiens endorses the second interpretation because he claims that, since Lipsey and Lancaster's publication of the first theorem, P1 in the first interpretation has not been proved. However, P1 has recently been discussed. The discussion around the notion of idealisation in chapters 1 and 2 nuances P1 (Valentini 2009 provided the detailed discussion on this point, as mentioned above). Chapters 1 and 2 reached the conclusion that background assumptions cannot be placed in a single category. In some cases, the fact that a background assumption did not obtain, did undermine the value of a theory and the second best theorem would apply. We have seen, such as with the assumption of well-ordered societies in the *Law of Peoples*, why we should not pursue an

ideal theory in the case that a background assumption did not obtain. But in some other cases, theories could more easily deal with assumptions not obtaining. The failure of ideal circumstances is not what prevents us pursuing ideal theory (Lawford-Smith 2009, Simmons 2010, Mason 2012).

Wiens rightly points out nonetheless that in the second interpretation, there is not much to support the claim that even in the case where one principle does not obtain, we will not be able to support normative principles in the real world. The fact that in a given situation we are shown that ideal principles are not satisfied is not sufficient to trigger the second-best theorem (Wiens 2016: 142). In sum, in response to the three questions of this section, we saw that the there are two plausible interpretations (question 1) of the theorem and that none of them rules out that there are ways elements of ideal theorising could be used in non-ideal theory (question 3). Yet, although it provides an interesting conceptual framework, the theorem of second-best does not cast much light into the decision-making process that connects the ideal and non-ideal theory. This relates to the second question that was left aside, regarding how the second-best theorem informs the ideal and non-ideal theory debate. I argue that there are central questions that provide a better framework for a theory of transition than the second best theorem, in particular questions such as: whether (a) we have duties to create the background conditions for more just institutional settings and whether (b) a theory of institutional design can deal with and ideal values and idealised assumptions not obtaining. We should see in the next sections that these questions will help us better to capture what is implied in the transition between available institutional alternatives and more ideal institutional settings.

In order to understand the question of transition, a final element to be addressed in this section relates to the notions of approximation and resemblance, which attempt to elucidate the connection between best and second best. Theories of second best normally do not include transitional elements that connect non-ideal theory to ideal theory. For Simmons,

> A state of affairs can be "second-best" because, say, it is the arrangement that most closely resembles "the best" in the set of those arrangements that are politically feasible (i.e., in cases where the ideal arrangement is simply not feasible), without this second-best

arrangement constituting genuine progress toward actually achieving the ideal (Simmons 2010: 25).

Simmons argues here that although 'second-best' arrangements may resemble the ideal setting, they are not thought of as paving a path that leads to the ideal setting. Simmons is right to criticise decisions based on second-best that 'resembles' the ideal. 'Resemblance' is indeed a vague way to depict the connection between 'best' and 'second best'.

I argue here that 'resemblance' is perhaps not the issue as much as using ideal theory to choose between alternatives. The question of what arrangements must be pursued to connect us to ideal theory has yet not been answered in this thesis. The first element I wish to address is that indeed it is not ideal theory that dictates the choice between two non-ideal level alternatives. This is one notion to be addressed in order to answer the central questions of this chapter, namely about what allows us to choose between short-term actions in order to move towards more ideal settings.

Proponents of ideal theory will say that ideal theory still constitutes a good target for non-ideal theory because we should still aim to satisfy ideal principles to the greatest extent as possible. In other words, we should try to realise the state of affairs that best approximates the ideal setting (Gilabert 2012b: 243, Valentini 2012b: 42). The second best theorem precisely rejects this view. It says we have no reason to expect that a state of affairs that approximates an ideal setting will be one from which we will be able to realise the ideal setting.

In order to reject this conclusion, the ideal theorist has to provide concrete evidence that a set of real-world reforms is likely to bring us the ideal setting. The ideal theorist will have to show which among a given set of alternatives will best allow to realise the ideal social setting. This implies that the ideal theorist will be required to do non-ideal theory. And I have already shown that non-ideal theory does not presuppose ideal theory. That is, whatever reforms are necessary will not be obtained by using ideal theory 'as measure'. It is non-ideal theory that dictates whether to move towards a second-best or a third-best scheme.

Moreover, the ideal theorist will be required engage in non-ideal theorising in

order to "(1) estimate the state of affairs most likely to emerge from efforts to realise the ideal as closely as possible; and (2) compare this state with feasible alternatives from the standpoint of justice" (Wiens 2016: 143). This shows that the comparative analysis of ideal settings will also be given by non-ideal theory. The action guiding elements of a political theory will be informed by non-ideal theorising. The knowledge in this case of the best state of affairs will be determined at the level of non-ideal theory. The kind of work that allows determining what state of affairs is likely to emerge in our effort to move towards more desirable social settings is a work proper to non-ideal theorising.

4.3 Transition and Reflective Integration

The second-best theorem did not provide with much guidance in answering the question of decision-making in transitional theory. We have learned nonetheless that, although it is not excluded that ideals play a role in transitional theory, ideal theory is not what allows choosing between what available options. To know how to realise ideals, we must engage in non-ideal theorising. This does not imply that aiming at long-term results will take us away from short-term gains, but the relation between available alternatives and ideal settings, or long-term and short terms objectives, deserves a closer examination.[78]

Note first, as shown so far, that we need not to be afraid of ambitious goals. There are ways to formulate ideal theory so as to make more explicit elements that need to be addressed in non-ideal theory, and it is paramount that the ideal goals are not the type of objectives that distract us away from understanding real world problems. Taking that into consideration and excluding harmful ideal theories, there should not be that much tension between long-term goals and short-term gains, in the abstract. Whether long-term gains can sometimes take us away from short-term gains is a question that can only

---

[78] We may assume for now that more ideal settings are associated with more long-term goals and non-ideal options are associated with shorter-term goals. The next section will clarify the notion of temporal variation in non-ideal theory and explain how to engage in long-term objectives in non-ideal theory.

be answer by bottom-up contributions of non-ideal theory. The balancing and path-dependence between long-term goals and short-term gains depends on the issue at hand, the nature of the long-term goal and the stage of development of the current institutional context.

This leads us to the two central questions of this chapter, which concern the connection between long-term objectives and short-term gains. Can aiming at long-term results take us away from short-term gains? Conversely, can short-term results take us away from the path to a more ideal society, or in other words: how can we make the difference between short-term results that are genuine good improvements and short-term results that might undermine the path towards a more ideal society?

To answer these question, I wish to draw on the reflective integration thesis developed in Chapter 2. I argued, contra Rawls (Rawls 1999b: 216), that existing institutions are not judged in light of an ideal conception of justice. I argued instead that we should approach the dynamics between ideal and non-ideal theory in terms of reflective integration, meaning we can revise our beliefs about ideal principles once we seek coherence between them in non-ideal circumstances and weigh them against non-ideal factors in a particular field of inquiry. There are two dimensions to this answer. The first is concerned with the mechanics of how institutions track principles. The second is concerned with the normative duties to pave a way towards more ideal settings. The next section develops this second point: the relation between 'ought implies can' and dynamic duties. Why choose an option that is less just but feasible instead of aiming at an option that is more just but not feasible right now? Perhaps we only have a duty to do what we can, but we foresee that we could do more in the near future if we moved to an intermediary institutional setting.

This section develops the first point about the mechanisms of how institutional structures and compliance mechanisms track principles, ideal or non-ideal. Whether the conditions of implementation of a less just scenario influence the conditions of implementation of a more just, more perfect one, have to be determined on a case by case basis. It is likely that in most cases, moving towards a more just scenario would render the ideally just scenario more feasible. But this is not necessarily the case.

154

Firstly, here, it is important to distinguish between objectives of reducing injustices and objectives of comparative judgment (Gilabert 2012b: 115), or as I have formulated, between the reparative and comparative components of non-ideal theorising. [79] As opposed to objectives of reducing injustices, duties related to comparative judgment objectives are more hypothetical in nature.[80] Reducing injustices often makes it the case that some options in the comparative alternatives will become more feasible. The objective of reducing injustices is not necessarily comparative in nature. It can be strictly reparative. One may also argue that in the case of second best options which involve reducing injustices – that is, an institutional scenario that is free from a series of injustices – there is little chance that they will prevent society moving towards more ideal settings.[81] This requires careful examination. As shown in chapter 3, more ideal scenarios are path-dependent on the reduction of injustices (e.g. the eradication of poverty depends on the elimination of a series of systemic injustices such as illicit financial flows). This is a case of a relation between ideal theory and institutional design, when we seek institutional designs that reduce injustices. Reducing these injustices will increase the chances of more ideal principles eventually being realised.[82]

If the comparison takes place between two institutional designs that aim at reducing an injustice, we may ask how these settings contribute to the realisation of principles. As shown, notably because of questions such as agency or collective action, this exercise of reflective integration will take place at the level of non-ideal institutional

---

[79] Sen, for instance, does not make this distinction. For Sen, reducing injustice is always a comparative exercise.

[80] In the Part II of this thesis we will focus on one objective that concerns more reducing injustice, in Chapter 5, and another that is more comparative in nature, in Chapter 6.

[81] This is perhaps less the case in the revolutionary approach of Karl Marx. Marx believe that if we increased wages, it would make the revolution less likely.

[82] In this thesis, most examples presuppose that the goal in justice discussions is one of moving closer to justice, and the question is what is the best way of making progress towards this goal. One may argue that this imposes a fairly positive cast on the discussion. There are cases where we might look at scenarios in which things are getting worse, and what we need to do is decide which of the current goods of justice can be preserved, and to what extent. These highly non-ideal circumstances are not uncommon (think of cases where restrictions on freedom are imposed in the name of national security). Perhaps informative parallels could be drawn between 'second-best' discussions and 'second-worst' discussions, e.g. we could try to assess what path-dependence cases to avoid between a given 'second-worst' solution and the 'worst case' scenario. For the purposes of this thesis, I leave this discussion aside. My examples focus rather on solutions to ongoing problems, which may allow us to avoid some of these harder choices.

design. That is, the comparison between different institutional proposals is given at the level of non-ideal institutional design and the realisation of values is given by their assessment at this level. When it comes to comparative judgments alone, whether a second-best option will pave the way to more ideal settings requires reflective integration and non-ideal theory methodology.

In other words, although ideal settings have an effect on the formatting of second best in reflective integration, the structure of what will be done in non-ideal theory will be determined by an account of the nature of the problem, including its associated question of agency and collective action. Reflective integration allows us to evaluate what is morally desirable and politically feasible because it has a central component that is dedicated to the analysis of practices. And the processes through which reforms are formulated, in reflective integration between values and non-ideal circumstances, are informative on how close to ideals we move. The context specificity of the reforms may well be inspired by values, but the reforms as such will be given non-ideal level analysis which focusses on the nature of the problem and considers its institutional, cultural, and historical challenges.

There is an important caveat here. It concerns the priority between philosophy and democracy. Providing an account of the role of democracy and of civil society in articulating moral ideas is beyond the scope of this thesis. Nonetheless, it should be clear that my model does not require giving priority to philosophy over democracy. That is, at least partly, because sometimes moral ideas and moral concepts are invented and popularized within broader society and only later theorized about by philosophers. Philosophers are not the only ones in charge of the normative weightlifting. Democratic processes, like work from other disciplines, can provide valuable insights for the normative work. I believe a reflective integration model allow us to be sensitive to such inputs.

Moreover, note that the domains of practical feasibility can be shaped historically and moral reasons may spur action to move social history beyond the bounds of current injustices. The intersection between desirability and feasibility is thus best seen as a moving target (Gilabert 2012b: 125). Gilabert and I differ however on the role we assign

to ideal theory 'as target', but we agree that we should see the intersection between the desirable and the feasible as a moving target.

In sum, understanding non-ideal theory as a reparative, comparative and transitional exercise helps us in answering the questions formulated in this chapter: can aiming at long-term results take us away from short-term gains? And how can we make the difference between short-term results that are genuine good improvements and short-term results that might undermine the path towards a more ideal society? Regarding the second question, we have seen that, in comparative cases, non-ideal theory allows identifying genuine good improvements and thus allows moving towards more ideal settings, regardless of a perfectly determined target. We have the knowledge to move towards more just scenarios when reducing injustices in non-ideal theory. In transitional and comparative cases, there is a great deal of knowledge involved in assessing path-dependence between ideal settings and non-ideal reforms. This is especially salient in the case of comparative and transitional assessments; it is less so for injustice reduction objectives. As seen, the knowledge in comparative cases can only be gained through non-ideal theorising. The next section will provide more insights about how this is accomplished in non-ideal theory and will focus more specifically on transitional and comparative cases. Also, the analysis of dynamic duties in the next section will provide further insights to the first question.

I will mention one last point about Rawls' treatment of empirical constraints. For Wiens, directive principles are justified "in light of the extent to which they reflect certain basic evaluative criteria given a set of empirical constraints, which consist of certain assumptions about which states of affairs can be realised" (Wiens 2015: 436). He distinguishes between three components of normative theorising: evaluative criteria (which allows for a comparative ranking of possibilities), empirical constraints (which provide the feasibility curves for jointly realizing the chosen evaluative criteria); and directive principles (which demarcate the lines between which institutional schemes are obligatory, permissible, and prohibited) (Wiens 2015: 433-434). He claims this model clarifies the relationship between moral values and empirical constraints in the process of formulating principles. He notes that there is a distinction between the justification

and the discovery of directive principles. That means that he is not claiming that the evaluative criterion is epistemically prior to directive principles, that it needs to be known for principles to be formulated. It is more that they are logically prior, which means that the realisation of certain evaluative criteria explains why a given principle is justified (*Ibid*. n.8). Pogge and Wiens point out that this can be observed in Rawls's theory of justice. Institutional settings modelled after individual sets of principles are comparatively evaluated by agents in the original position according to the extent to which they realise certain evaluative criteria (Pogge 1989: 36–47). These criteria in Rawls are a fair and stable system of cooperation that respects all individuals' freedom and moral equality.

Rawls believes his principles to be justified because, given the empirical constraints set by the original position, they reflect basic fundamental values. Institutional settings are modelled by principles, which reflect values, under given empirical constraints. Basic moral values are freedom, fairness, equality, peace and cooperation, among others. They are expressed in terms of 'evaluative criteria' according to Wiens. For instance, the value of freedom can be expressed by the idea of non-domination and the value of equality be expressed by the idea of capabilities (Wiens 2015: 437).

In this Rawlsian model, institutions are justified because they realise principles and principles are justified, given empirical constraints, because they realise values. My point is that this model does not say anything about different levels of empirical constraints in less-idealised context and about the normative work required to move from one feasibility curve to the next. The interest for the non-ideal theorist is about the justification of institutions given real world empirical constraints, which differ greatly from the empirical constraints of the original position. The process of bringing down these values and principles to the level of non-ideal theory reveals how dynamic is the exercise of political theory.

4.4 Dynamic Duties and Path-Dependence

This section focuses more precisely on the moral duties involved in cases of second best and non-ideal reasoning. We will examine cases where ideals are partially realised and whether that informs feasibility and path-dependence. More precisely, the last two sections indicated that one central question for non-ideal theory is in virtue of what and by which mechanism we can expand current feasible sets.

The notion of temporal variation will come into play in non-ideal theory work. I mean by this that it is not because we are engaged in non-ideal theorising that we are limited to what can be done here and now. On the contrary, we can aim for a second best solution in hope that it will develop the cultural and institutional changes that will put us in the direction of more desirable social settings. For example, this idea is embedded in the notion of 'progressive realisation' in human rights documents (ICESCR 1976, Nickel 2009, Gilabert 2012b: 132).[83] I have determined in the previous section of this chapter that the right direction is not something determined by ideal theory alone. This claim is not particularly controversial. My goal in this thesis is to show that the non-ideal contribution to determine the right course of action has been underexplored. When we work in non-ideal theory, we work closely with social sciences, economics and natural sciences. The work required to fulfil more ideal settings will be directed by many other elements than ideal theory. Yet, the mechanisms by which this process occurs deserve further examination. This section examines the rise of moral duties in this context.

I begin by reviewing Gilabert's account of dynamic duties. Gilabert asks us to consider a certain institutional scheme S2, which is more desirable than another scheme S1. Yet, S2 is only accessible in circumstances C2 and right now we are in circumstances C1, where it is not accessible. S1 is nonetheless accessible in C1 and S1 is very likely to generate C2. In this scenario, he argues that if the moves to S1 and from S1 to S2 do not

---

[83] The International Covenant on Economic, Social and Cultural Rights and the Universal Declaration of Human rights make several references to the idea of Progressive Realisation.

involve unacceptable moral costs, we have a duty to pursue reforms leading to S1 in C1, and from there we have a duty to pursue S2 in C2 (Gilabert 2012b: 137). For Gilabert,

> "the second segment in this trajectory contributes to the justification of the first, and that the trajectory involves making accessible, at a later time, what may not be accessible now … Let us call duties of this sort, involving the expansion of the feasible sets of political action, *dynamic duties*." (Gilabert 2012b: 137)

I agree with Gilabert that, in the scenario described above, the second segment justifies the first. We do have dynamic duties of justice. Dynamic duties ask us to think a few steps ahead and focus on what background structures can be changed for more desirable settings to become feasible in the future. Gilabert is also right in arguing that when theorising about justice, we must adopt a transitional standpoint. And this is particularly relevant for non-ideal reasoning. Agents must be mindful of the potential process of changing the features of the institutional context in which they evolve. The transitional standpoint differs both from the conservative position according to which the institutional setting is taken as given and from the view according to which we have at our disposal the blueprint for the desirable and feasible settings (Gilabert 2012b: 145).

The central difference between my reading of the scenario above and Gilabert's, is that my reasoning takes place at the level of non-ideal institutional design. By placing this model at the level of ideal theory, we run the risk of assuming path-dependence, in the sense that the ideal provides us with the path. Path-dependence should be determined with great care, as mentioned in the last chapter. The bulk of the challenge is to know what is required to generate the conditions to move to C2 and to examine the likeliness that S1 generates C2. S2 must therefore be a more ideal institutional design that responds to a particular practice and not an ideal society. Otherwise, the scheme would not guide action. S2 must be given through bottom-up reasoning in a non-ideal approach. The conditions to move towards C2 are about non-ideal theory problem solving.

Consider the following example, which would prove claim (iv). Some of the terminology used here and the detail of these institutional proposals will be explored with greater care in the next chapter. Assume we know that an institutional scheme S2 –

think of an International Tax Organization (ITO), a system of global tax governance (which I will present in Chapter 5) – would effectively eliminate harmful tax competition. S2 would provide sufficient enforcement and dispute settlement mechanisms to be effective. S2 would be available from C2, a context of multilateral exchange and collaboration. We are now at C1, a context where we have at best bilateral collaboration and where agents (nations and their respective internal corporate pressure) are only concerned with avoiding double-taxation. S1 is an effective institutional setting that favours multilateral collaboration. We have the option of asking for a demanding Unitary Taxation and Formulate Apportionment (UT+FA) or pursuing the work within the Organisation of Economic Cooperation and Development (OECD). The first option proposes that corporations would have to produce a combined report on their global profits, which would then be distributed to the different tax jurisdictions in which the corporation is active, following a predetermined formula, which is based on indicators of real economic activity. The second option demands more multilateral information exchange for a more rigorous coordination proposal. Now, and I will explain this below, the OECD proposal is a coordination scheme that falls short of global tax governance. UT+FA is a system of governance. The former is available from C1. The latter requires more work in C1 (lobbying, information, cooperation) to become available. However, we know that, given the features of ITO, only UT+FA can lead to it. That is, we have dynamic duties to implement UT+FA. The OECD proposal does not set a path towards the ITO. This is obtained through non-ideal level theorising. In fact, the OECD must be one of the organizations that push for a system the governance that transcends its boundaries. Claim (iv) seems thus correct: path-dependence is not given by ideal theory, but rather by non-ideal theory. The correct target is given by non-ideal theory.

This thus provides further clarification to the two chief questions of this chapter: Can short-term results take us away from the path to a more ideal society? And how to make the difference between short-term results that are genuine good improvements and short-term results that might undermine the path towards a more ideal society? The last section informed on how to answer these questions in reparative justice cases. This

section focussed more on comparative and transitional cases. From the previous example, we can observe that the OECD proposal is one short-term result that does not put us on the path towards a more ideal society (it might not take us away from it, but it might give the false impression we are moving towards it). We can make the choice between the two 'second-best' options on the basis that one is more likely to lead us towards the more ideal institutional setting. This is done through an adequate analysis of the problem, which informs us on the nature of the collective action problem and on type of institutional structure that will be required to overcome it.

The work in non-ideal theory, in a bottom-up fashion, is thus what allows determining what schemes are path-dependent on which proposals. The nature of the empirical constraints will inform the normative work to be undertaken. This suggests that non-ideal theorising allows reducing our epistemic uncertainty as to which reforms are likely to be effective and desirable. We need empirical research and non-ideal institutional proposals, in reflective integration, to determine what paths should be undertaken. Ideal values will also be thought in reflective integration with the non-ideal institutional proposals. Moreover, as argues Gilabert, "full certainty is not something we can secure in political practice. Reasonable, defeasible forms of certainty are all we can, and must, look for." (Gilabert 2012b: 151). Indeed, a non-ideal theorising, reflective integration and the transitional standpoint are part of a cautious consideration, which reduces the need for certitude.

4.5 History and Ideals

One hypothesis needs to be formulated before concluding Part I. Is it possible that today we know enough about political philosophy that we can confidently address injustices without waiting for more refinement of ideal theories? One possible explanation of the recent wave of objections to ideal theorising is that perhaps we know sufficiently about the contours of moral and political ideals that further work in ideal theory is not necessary anymore in order to undertake non-ideal theorising.

This effort to rehabilitate ideal theory 'as target', in a historical perspective is summarized by Swift and Stemplowska.

> Though there are issues about our ability to map possible futures well enough for this kind of consideration to have great weight, though the likelihood of achieving the goal is a very different matter from its possibility, and though it might be argued that we already know enough about the goal to be able to assess whether immediate steps are unambiguously steps in its direction, this defense of ideal theory as setting the target is compelling. (Stemplowska and Swift 2012: 385)

The hypothesis here is that ideal theory provides broader guidelines about where it needs to go, and what are the morally justifiable targets. This hypothesis is weaker than the claim about the analytical primacy of ideal theory. It is weaker in the sense that a precise notion of ideal principles is not required, that ideal theory 'as target' needs only to provide a general direction and that ideal theory is not required 'as measure'. This would mean that non-ideal theorists in fact have a sufficient notion of political ideals to proceed with their work. Whether they might need to endorse a particular conception of ideal theory is a separate question, to which I return below. Put broadly, this first hypothesis says the question about the primacy of ideal theory would thus be one that can only be answered in a historical perspective.

This hypothesis is motivated by the idea that the non-ideal theorist must have in mind some ideals that helps conceptually with the injustice they are addressing. Broad notions of fairness should be summoned when designing international trade rules, such as the TRIPS agreement at the WTO. The duty not to harm is also a moral ideal that has traction in the political sphere and is summoned regularly to address problems of global justice. Notions of domination, oppression and exclusion have also evolved and are constitutive stars of our constellation of moral tools. To go back to a point mentioned in the Introduction, this discussion allows to cast light onto the role of values in non-ideal theory. Moral values have a role to play in solving problems that are particular to the level of non-ideal theory, such as the demands of agency, the nature of collective action problems and the forms of political authority. For instance, in the case of climate change, the idea of responsibility for past emissions is a moral value that will play a role at the level of non-ideal theory to assign duties to agents in the fight against climate

change. The notion of historical responsibility will be assessed in a context where carbon emissions from developing countries are increasing considerably. The role of this principle will be to provide inputs for an institutional design that assign agents responsibilities to reduce emissions, where it will be balanced against other considerations that also aim to explain what ways to reduce emissions are available to us. Although the formulation of a principle of historical responsibility at the level of ideal theory is not what guides action, this notion contributes to the work at non-ideal level. Another example can be borrowed from the context of harmful tax competition. The value of autonomy will help in justifying the notion of fiscal sovereignty, a notion that provides an important basis for the non-ideal institutional design that aims to solve this problem. Values are indispensable for work in non-ideal theory, but their practical role as assigned to non-ideal level problem solving is distinct from the conceptual exploration and reasoning under idealised conditions particular to the level of ideal theory.

Now, can we say that we need an ideal theory about the distribution of rights and liberties in order to prevent violation or rights and liberties? Without our knowing that we are entitled to these rights, any violation could occur. It would be a major regression to disregard all the progress of the last few centuries, in terms of the distribution of rights and liberties. Think for example of due process rights or freedom of speech. Due process rights go back to the Magna Carta. Following what was argued in section 1.4, I do not think the nobles who were challenging King John in 1215 were doing ideal theory. Philosophers that have argued for such rights on a theoretical basis have worked based on rights that were already entrenched in civil society. Nonetheless, although ideals and rights were already in use before being integrated in a philosophical theory, they can very well be constituted in an ideal theory of rights and liberties. Being such, this ideal theory is one that guides action in non-ideal circumstances and must be known prior to non-ideal theory. Once ideals have been around long enough and society is sufficiently evolved, and these ideals become part of an ideal theory, this ideal theory guides action in everyday life. Although philosophers might not be the first to articulate certain ideals and it is not necessarily by the action of philosophers that these ideals are

made actual once the circumstances are suitably favourable, these ideals can become part of an ideal theory, which, like in the case of due process rights, become action-guiding.

This proves the fact that in some respects, society – with its institutions and social ethos – has evolved sufficiently so that ideal theories become action-guiding. This means that ideal theory guides action, although not in all sorts of non-ideal circumstances. It guides action when, while agents may not comply, the institutions have sufficiently evolved through political action and reforms to judge this instance of non-compliance. This process of bridging the society to this level, however, is done by non-ideal theory and political action.

Again, this does not imply that ideal theory has analytical priority over non-ideal theory. Defenders of the analytical priority claim would not be convinced by it. Neither would non-ideal theorists. But it implies that the conceptual space once opened up by ideal theories eventually is filled with legal and political devices that become instruments of the ideal. Society might not be guided by ideal theory in non-ideal circumstances, but in more ideal contexts where the institutional capabilities can respond to the problems, ideal theory has greater action-guiding potential. This historical view of the evolution of political theorising and institutional design is more charitable to the contribution of ideal theory today. I think non-ideal theorists should be satisfied with this conclusion and still maintain that today a considerable portion of the work in political philosophy is situated at the non-ideal level.

A second element of this hypothesis is that ideal theory is not useless, but perhaps we have done enough of it (Stemplowska and Swift 2012: 380 raise this possibility). I suggest one way in which one could agree and one in which one could disagree with this statement. We might have done enough of it to engage in decades of non-ideal theorising in political philosophy without being remotely scared of running out of things to do. In this sense, it is perhaps the case that we have sufficient notion of different ideals towards which we should work, that the centre of the attention in political philosophy should be non-ideal theory. For instance, one may argue that the notions of gender and race equality have recently been complemented by sophisticated debates about affirmative action and measures for us to be know seeking ways to implement a more

just background of interactions. This is maybe intended to show that we know sufficiently about the ideals of race and gender equality but we do have not done enough in non-ideal institutional design to achieve these ideals.

However, we do not have done enough of it to forget the amazing progress that could be achieved by engaging with the always-new challenges that are constitutive of an evolving society in a globalized world, and with the progress that can be achieved by looking at issues of the past. For instance, Ypi's recent work on colonialism – although the fetishism of the unitary value to explain what is wrong with colonialism could be disputed – is a substantial work that better captures some aspects of the problem of colonialism (Ypi 2013). This is a bottom-up contribution to ideal theorising. It is not focussed on institutional design. It is mainly focussed on providing the conceptual tools to better understand and criticize a given situation.

In sum, while we might have done enough ideal theory in the sense of having built enough lighthouses strategically located around the globe, we have not done enough non-ideal theory to make sure people know how to navigate. Yet, this does not guarantee that all lighthouses illuminate far enough, thereby suggesting that there is still room for important ideal theory contributions.

To return to the Everest analogy, ideal theory has been teaching us where to go to find the highest peaks. For Simmons, as mentioned, we need to know where to find the serious candidates for the highest peak in order to endorse any path from here to there (Simmons 2010: 35). We aim for India and once we get there we adjust more finally our route between Mount Everest and K2.[84] This is mostly correct. But I disagree with Simmons about the necessity of knowing which one is the highest peak, for I argued we should drop the notion that ideal theory is perfect justice.

What I have also argued is that we must block access to the routes that lead away from the Himalayas and stop people that prevent other people who want to take the roads that lead to India. What we need to know, when doing non-ideal theory, is where

---

[84] It is worth mentioning that neither Everest nor K2 are in India. Everest is on the border of Nepal and occupied Tibet. K2 is on the border between China and Pakistan. The peaks are situated at opposite ends of the Himalayas. But I follow Simmons here in his analogy for we understand that in India we are closer to the highest peaks than, say, in the UK.

India is. We do not need to know which one is the tallest between the Everest and K2 and we do not need to know the height of the Everest in order to make sure the routes to India are accessible and that people do not run the risk of taking the wrong turns.

To continue the metaphor, in India, the democratic process might allow people to take a few wrong turns. But we need to know that we will remain in India. This leads firstly to the definition of what a minimally just society is, with the various tools that are normally used in non-ideal theory. We will use for instance the principle not to harm, the notion of basic rights, a general ideal of not undermining individual opportunities, and we will do all this in a democratic context. The definition of increasingly more just societies – knowing that we need to move towards northern India – will be given by ideal theory and the evolution of institutions. Ideal theory has taught us that much through the last few decades. By opening a critical space and by shaping the social ethos, ideal theory has contributed to putting non-ideal theory in the right direction. But when engaged in non-ideal theorising, the tools that the philosopher requires are not the ones of ideal theory. The guidance aspect of ideal theory is a very broad theoretical guidance. It is not practical at the level of detail.

4.6 The Political Philosopher at the Gates of Social Sciences

As I made clear in the introduction, I believe the role of the philosopher is still to follow the argument where it leads. This does not change whether the philosopher begins the inquiry in the abstract realms of philosophy of mind or ideal political theory, or in the empirical works of cognitive science or social science. The philosopher will still follow the argument where it leads.

I have argued here for an account of non-ideal theory that aims to guide action. That is the central difference between ideal and non-ideal theory. That is why I insisted that this should be a central element of non-ideal theorising. Of course, to guide action is a complex matter. There are various elements to be taken into account for a theory to be thoroughly action guiding. I included in this concept: path-dependence, fact-sensitivity,

feasibility and compliance. It is possible for a non-ideal theory to focus on only a subset of these elements or on all of them at the same time. For instance, we can theorize about how to ensure compliance with a set of principles for principles that would not be feasible in the near future (Stemplowska and Swift 2012), and leave the feasibility question for another moment. The four notions just mentioned are all elements of an action-guiding theory, but perhaps not of a theory that would guide action here and now, for not all non-ideal theories have to be non-idealised to the extent of guiding action here and now. Like ideal theories, non-ideal theories can be non-idealised to a greater or a lesser extent.

I have argued in the previous section that even if we have done enough of ideal theory, the type of guidance it provides is very different in nature that what can be gained from non-ideal theorising. We get from some of the main arguments in Part I of this thesis that bottom-up non-ideal work on individual practices provide us with the most crucial elements to build an action guiding theory notably with regards to: the different demands of agency (taking into account, for instance, past, present and future agents); the sort of collective action problem we face (e.g. prisoners' dilemma, asymmetric prisoners' dilemma); the kind of authority required (e.g. a coordinating body or a system of governance). All these elements provide valuable insight about how to choose between two non-ideal institutional proposals and how to move towards more ideal settings.

The reason why we do not need ideal theory in order to do non-ideal theory is that improvements to fight injustices in non-ideal theory need to be addressed directly and individually. In cases of reparative justice, ideal theory is not necessary. In cases of transitional and comparative justice, although elements of ideal theorising can play a role, how to move towards more ideal settings can only be known through non-ideal level theorising. Whether addressing these injustices collaborates with a particular ideal theory is an important question but it is wrong to assume that all that is needed is to formulate abstract principles of justice. The structure of the problems with these injustices is to a great extent particular to each one of them.

This conclusion has obvious consequences for the work of the political philosopher interested in non-ideal theory. The non-ideal theorist can still discuss rights and duties. She or he can still talk about responsibility and distribution. But the framing of such debates under less idealised assumptions require philosophers to draw more careful assumptions about the non-ideal world. I argued that the quality of the recommendations will depend on how closer we get to understanding 'how' to do what we 'ought' to do in reflective integration, e.g. knowing what options allow us moving to move towards an institutional scheme that combats tax competition informs us on what we 'ought' to do in non-ideal circumstances.

Now, we may still ask to what extent the work of philosophers should engage with social scientific work and what results are likely to result from that. Stemplowska and Swift agree that perhaps political philosophers should engage with this kind of work. However, for them, "we should be clear to demand that political philosophers deliver concrete action-guidance is to ask them to do something more than philosophy." (Stemplowska and Swift 2012: 386). I rejected this. This is precisely the kind of work I advocate. The non-ideal theorist will obviously not substitute for the social scientist. But she or he will provide a theory that guides action. If the philosopher comes to a conclusion about a real world case by seeking to bring about change, it is hard to argue that this is not part of the philosopher's work (Wilson 2014). On the contrary, the philosopher is perhaps particularly well equipped to be an advocate for real change if ever she or he engages sufficiently with the relevant work of the field.

I think that, in order to determine what could be gained from philosophers engaging in the kind of social science research that could contribute to formulating action-guiding theories, we should look at how philosophers do it. If not enough has been done, this is perhaps a call for philosophers to do it more. I for one believe that there is great potential for philosophical research in engaging in the work of social scientist and policy makers.

Wiens (2015) and Miller (2012) are of this view. They ask for a deeper partnership between the political philosopher and the social scientist in normative political theory (Miller 2012). This collaborative work is also meant to rebut the division according to

which the role of the political philosopher is to formulate principles and the role of the social scientist to implement them. The political philosopher has to better understand the political structures of the realisation of justice, just as the social scientist has a role in the formulation of directive principles.

Moreover, although I have been less concerned with motivation, the political philosopher has to be mindful of the motivational structures that affect social and political life. That is part of the work if one wishes to surmount the status quo. Motivation is an important element, which I have not addressed as such. This thesis is more focussed on the structures that favour compliance and to consider issues such as incentives and sanctions. The work on institutional structures that favour compliance is surely one that can have a positive impact on the motivation of agents. Whether agents are likely to be motivated by the principles is a question that will only be dealt with indirectly in Part II.

4.7 Conclusion

Chapter 4 has provided further exploration on the notion of action guidance and on the ways to conceive the relation between ideal and non-ideal theory. It consolidated the understanding of what makes a non-ideal theory of institutional design action-guiding, notably by exposing ways to understand how to bridge the gap between ideal and non-ideal theory, thereby providing further support to the 'reflective integration thesis' and the 'non-ideal theory thesis'. Once we have ruled out, in the previous chapters, the notions of ideal theory 'as target' and 'as measure' we have opened the door to examining how short-term or second-best gains impact the path-dependence relation between more ideal and less ideal settings. I concluded that the decisions regarding how to assess ways to move towards more just schemes are decisions for the non-ideal theorist.

There is no denial that we need ideals. As Simmons argued

> As *activists* in the cause of justice, ideal theory may come to seem to us simply irrelevant. But it is important to remember that even most non-philosophers who are active in the cause of justice do in fact have in mind, however vaguely, an ideal of justice toward which they take their campaigns to be ultimately directed. (Simmons 2010: 35)

This is true, and I argued that this idea becomes especially compelling if understood in an historical perspective. This nuance casts light on the contribution of ideal theorising, once we see it in an historical perspective: institutions can evolve in ways such as to allow ideal principles and values to have greater action-guiding potential. Ideal theory has more action-guiding potential when circumstances become more ideal (through political action and non-ideal institutional design). The extent to which these ideals need to be coupled with (over-)determined ideal principles is less certain.

I argued that we need non-ideal theory to know when second-best institutional schemes bring us closer to more ideal institutional settings and when they take us away from them. An adequate examination of the practice is what allows to understand the nature of the problem we face (about e.g. agency questions, collective action problems) and the nature of solutions needed (e.g. regulation, coordination, or governance). I argued that the non-ideal institutional design is what renders possible to gain knowledge about the feasibility and desirability of any future ideal setting. This is the case for reparative, comparative and transitional justice; and for works that are essentially reparative, notions about ideal settings seemed to play an even less important role. Our moral duties to implement the non-ideal setting are justified based on knowledge acquired at the level of institutional design. Moreover, we saw that just like ideal theories, non-ideal theories are non-ideal to various degrees. When this degree approaches the level of social sciences it becomes clear whether it yields concrete proposals or not, whether it has a grip on the probabilities of the positive outcomes associated with the various feasible alternatives at hand.

The three case studies in Part II of this thesis should cast light on the extent to which ideal values and principles, as well as non-ideal circumstances and empirical factors, play a role in the formulation of action-guiding principles.

PART II

Will there really be a "Morning"?
Is there such a thing as "Day"?
Could I see it from the mountains
If I were as tall as they?
…
Oh some Scholar! Oh some Sailor!
Oh some Wise Man from the skies!
Please to tell a little Pilgrim
Where the place called "Morning" lies!
- Emily Dickinson

5. TAX COMPETITION

5.1 Introduction

The aim of Part II is firstly to illustrate how non-ideal theory works in the contexts of tax competition and climate justice, which are the focus of Chapters 5 and 6. This will provide us with a solid basis from which to develop a more detailed account of bottom-up theorising in these two cases. Chapter 7 will go beyond the objective of these two chapters and offer an account of a bottom-up theory of climate justice as it relates to carbon pricing. Illustrating how non-ideal theory works and offering new bottom-up theories of climate justice or tax competition are two distinct objectives. They are both important. While the former was the primary goal of this thesis, the demonstration of the meta-theoretical claims via the examination of these three examples will not prevent me from passing judgements about the content of a non-ideal theory targeting these three issues, especially in Chapter 7.

In other words, the central aim of the three chapters of Part II is to provide support to the meta-theoretical arguments of Part I. Tackling in turn tax competition, climate justice and carbon pricing, we are given in Part II a more in-depth elucidation about the central claims of this thesis. These chapters offer further support to the 'reflective integration thesis' and to the central claims of 'the non-ideal theory thesis': (i) against the analytical primacy of ideal theory; (ii) in favour of bottom-up approaches to political theory; (iii) about the question of compliance in non-ideal theory; and (iv) about path-dependence between ideal and non-ideal theory. However, the three chapters of Part II will offer important insights for more thorough bottom-up non-ideal theories targeting tax competition, climate justice and carbon pricing (which will be particularly put into practice in the case of carbon pricing).

Chapter 5 focuses on the issue of tax competition. Today, tax competition is high on the political agenda of international institutions, like the G20 and the OECD, and of national governments (Dietsch and Rixen 2016). In the OECD, the response began to take form in 1998, when the organization commenced its project to target harmful tax competition (OECD 1998). Public awareness of the issue has been raised by recent tax scandals involving multinational companies like Starbucks and Apple. Recently also, especially since the mid-2000's, academic contributions on the topics of tax competition, tax evasion, illicit financial flows and global tax governance have emerged.

This chapter aims to provide an overview of the problem of tax competition and to argue that the way to address this global justice issue is to adopt a non-ideal theory perspective. As seen in Chapter 3, this theory will be non-ideal in its methodology (it will proceed from the real-world practice) and its content (the principles formulated will be sensitive to the non-ideal circumstances and features of the world that are particular to the practice). This chapter argues that the principles that should regulate tax competition can only be given by an adequate understanding of the practice. And, following Chapter 2, this chapter concurs with the view according to which action-guiding principles can only be formulated in reflective integration between ideal and non-ideal theory. I will conclude that bottom-up approaches contribute significantly to understanding and tackling the problems that are particular to global justice today.

This chapter proceeds as follows. After providing a general account of the phenomenon of tax competition (5.2), this chapter offers an assessment of the practice from the standpoint of justice (5.3 and 5.4). It then proceeds to examine the principles of justice and institutional framework that could effectively address this problem (5.5, 5.6 and 5.7). This analysis allows us to conclude that a non-ideal theory of justice targeting tax competition offers specific tools to guide action towards global justice.

5.2 The problem of tax competition

According to the methodology of non-ideal theory, our first task is to provide an account of the targeted practice. From this analysis, the first determination to be made is to assess whether this practice raises problems from the standpoint of justice, i.e. whether it has a pervasive impact on individual opportunities. The content of a non-ideal theory tackling tax competition, including non-ideal principles of justice, should integrate the notions developed in the previous chapters: non-compliance, feasibility, fact-sensitivity and path-dependence.

Providing a full analysis of the problem of tax competition is beyond the scope of this thesis (Avi-Yonah 2000, TJN 2005, Clausing 2016, Dietsch and Rixen 2016 offer thorough and insightful accounts of the phenomenon). I will thus focus only on the questions directly related to the central arguments of this thesis, such as fact-sensitivity, path-dependence and compliance.

Today's globalized economy is characterised by high capital mobility but tax policy is generally treated as a domestic matter. In this context, individuals and corporations can select different tax regimes. Tax competition is broadly understood as the phenomenon of countries lowering tax rates in order to attract capital, corporate investments and labour (Rixen 2011b). Tax competition occurs when tax systems are sensitive to tax differentials. Three forms of tax competition can be distinguished in the areas of: portfolio capital, paper profits, and foreign direct investment (Dietsch and Rixen 2012: 2).

So-called 'tax havens' are an example of tax competition in the area of portfolio capital. Tax havens offer low or zero tax rates to attract capital. They also offer bank secrecy rules in order to hide ownership of bank accounts from external governments. This procedure is illegal when conducted for the purposes of tax evasion in most countries. In this case, individuals hold their capital gains 'offshore'. The curtain of secrecy surrounding this form of tax competition indeed allows many people to evade tax (Dietsch 2016: 233). Notably, 50% the financial wealth of Latin American countries

and up to 70% of the wealth of Middle Eastern countries is held in tax havens (Dietsch 2015: 3).

In the area of paper profits, we observe a number of different practices. One of the most common phenomena is called 'transfer mispricing'. One way through which multinational enterprises (MNEs) manipulate transfer prices is by assigning their profits made in high-tax jurisdictions to their subsidiaries in low-tax jurisdictions. This constitutes a legal form of tax avoidance. Multinational firms manipulate prices on intra-firm transactions. Consider the case of a company that sells office supplies in the US and declares very low profits, because it manipulates prices so that is appears to buy its pens, printer paper and laptop bags from its subsidiary in the Cayman Islands, a tax haven, at an extremely high price.[85] No money actually circulates between these affiliate companies. The Cayman Islands company in fact registers this transaction as if it has bought these from another subsidiary company, a manufacturer in Honduras. The Honduras subsidiary appears to sell the goods to the Cayman Islands subsidiary at a very low price, again declaring very little profit. This way, the US company and the Honduras subsidiary do not declare significant profits. However, the Cayman Islands subsidiary makes a gigantic profit, for it apparently bought pens, printer paper and laptop bags at a low price from Honduras but sold them at an exorbitant price to the multinational office furniture company in the US. MNEs do not need to actually buy and sell goods from its subsidiaries. Most of this is only written and calculated for tax purposes.[86] The overpriced items will produce revenue in subsidiaries located in low tax countries and reduce them for subsidiaries in high tax countries. This is one example of how transactions are manipulated so that the bulk of profits of multinationals is declared in low-tax jurisdictions. Another example in the area of paper profits is what is today known as the 'Double Irish' tax structure, used by Google and many other multinationals. Also known as 'earnings-stripping', it involves taking out loans from foreign affiliates and paying tax-deductible interests (Dietsch 2016: 233-234). As

---

[85] Whether this is or can be regulated by the 'arms-length principle' is a question to which I return below in section 5.6.
[86] I thank an anonymous reviewer at MOPP for pointing this out.

178

opposed to tax evasion, most of these strategies in the area of paper profits are legal. To get an idea of the magnitude of the problem, without claiming that all of it is transfer mispricing, the OECD in 2002 has shown that 60% of all international trade is done intra-firm (Dietsch 2016: 234). In the first two forms of tax competition, governments 'poach' capital from the potential tax revenue of the governments where individual and corporations are actually located.

The third form of tax competition occurs in the area of foreign direct investment (FDI). This is the competition to attract companies. MNEs choose the location of their company according to various factors such as labour qualification and quality of local infrastructure. Yet, one important factor is the tax burden in the chosen country (Clausing 2016). Governments wishing to attract FDI can for example lower their general tax rates or design preferential tax regimes for particular sectors of activity. One example that comes to mind in this case is Ireland. With its low corporate tax rates, Ireland attracted 25% of US companies' FDI in Europe between 1990 and 2004 (Peet 2004). The direct investment stock in Ireland is greater than in Germany and France combined (Taylor 2015). This form of tax competition involves a real relocation of economic activity. A government that engages in this third form of tax competition 'lures' the tax base of foreign governments to make it a legitimate part of their tax base. Whether the luring government will actually tax these revenues is an open question. The available data (see below) indicates that these taxes will be considerably lower at the 'luring' destination.

Thinking about the relation between real tax competition and virtual tax competition, that is the relocation of real economic activity versus the relocation of profits, firms will be more inclined to relocate their real economic activity as the possibility to shift profits diminishes. In other words, the response of multinationals depends on the possibility of tax avoidance (Clausing 2016). There is a substitution effect between portfolio capital and foreign direct investment in response to tax incentives (Clausing 2016: 38). We observe it between poaching and luring. If MNEs know that they can reduce their tax rates by shifting their profit to low tax jurisdictions, they will have low incentives to look for FDI opportunities in low-tax countries (Dietsch

2016: 235). For instance, if virtual tax competition were more heavily regulated, corporations would have more incentive to seek to relocate their economic activities to low-tax jurisdictions. It will be important to bear in mind that banning virtual tax competition would increase the incentive to relocate real economic activities.

There is substantial evidence that multinational firms are indeed tax-sensitive in their economic decisions. In fact, evidence shows that multinationals are even more tax-sensitive in financial decisions than in economic decisions. This comparison is striking. Take the example of subsidiaries of multinational corporations based in the US in figures 5.1 and 5.2.

Figure 5.1. Top income countries for affiliates of US Headquartered multinational firms in 2011 (Share of total of foreign profits of US affiliates abroad).
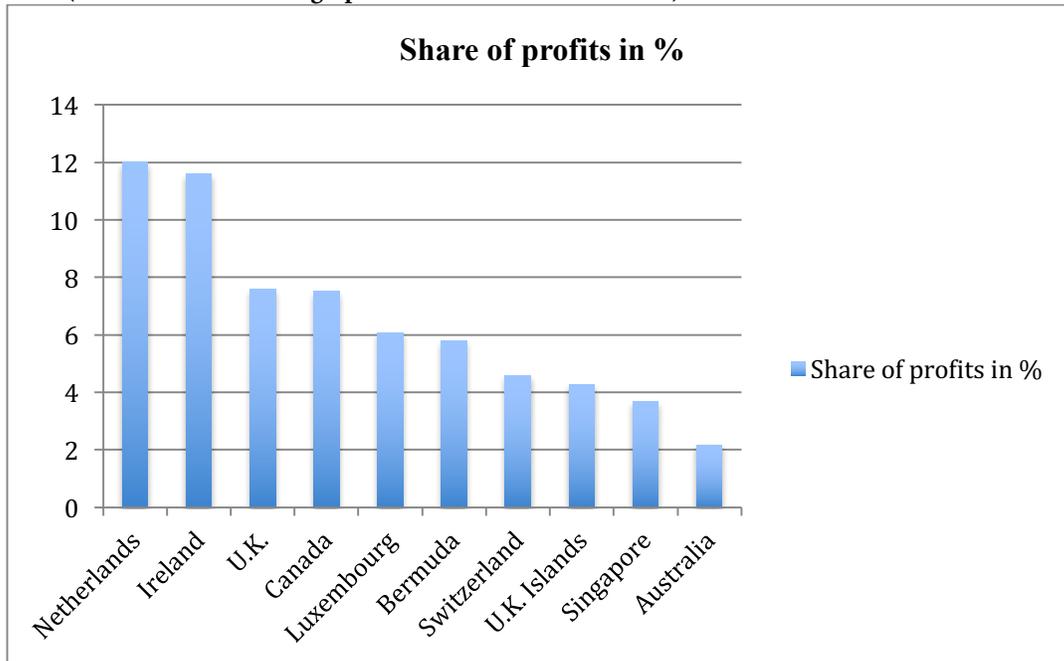
Source: Data from the US Bureau of Economic Analysis (Clausing 2016)

Figure 5.2. Top employment countries for affiliates of US Headquartered multinational firms in 2011 (Share of total of foreign employment of US affiliates abroad).
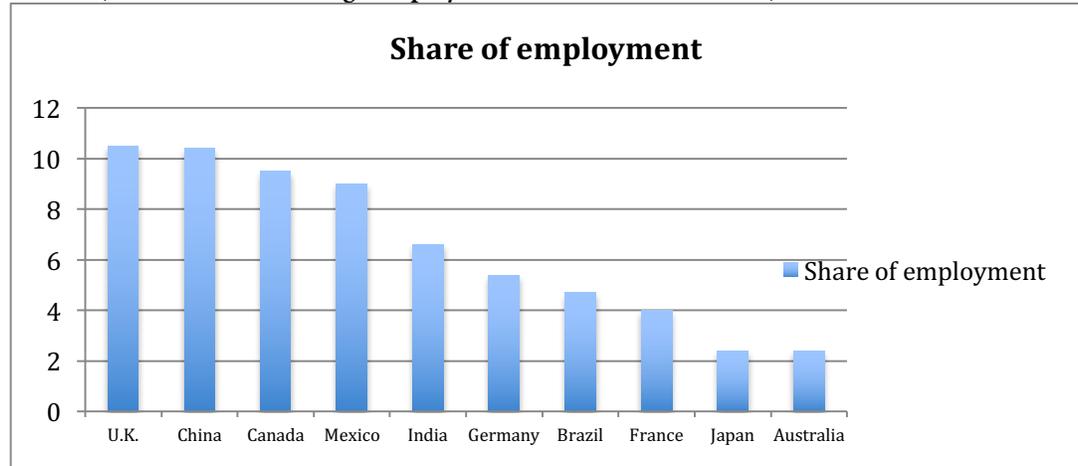
Source: Data from the US Bureau of Economic Analysis (Clausing 2016)

As shown in Figure 5.1, seven out of ten destinations with the highest level of foreign profit of all US headquartered multinational corporations have an effective tax rate of less than 6.5%. The percentage of all foreign profits of these 7 destinations is 46.5%. However, the latter only account for 5% of all foreign employment. In comparison, if we look at the largest employment countries for affiliates of US-based multinational companies in Figure 5.2, no company in the top 10 has a corporate tax rate of less than 6.5%. These are the big markets where US multinational firms have operations abroad for real economic purposes (the UK, Canada, and so on). However, it is not in these countries that most profits are actually made.

This data shows that profits are in fact very sensitive to tax rate differences. An immense portion of profits of multinationals are made in countries where there is little economic activity. Examining the consequences of tax competition for different governments should allow for a more precise grasp of the problem.

5.3 Impacts of tax competition

An in depth analysis of the consequences of tax competition is of course beyond the scope of this chapter. It would certainly be interesting to map its consequences in terms of employment, investments and revenues. For the purposes of discussion, I will focus on the question of revenues. I will distinguish between rich and poor countries; and within more affluent countries I will distinguish between the impact of tax competition on capital and on labour.

Indeed, it is helpful in this context to distinguish between the impact of the phenomenon on richer and poorer countries. Empirical studies have shown that nominal corporate tax rates in OECD countries (essentially the world's 35 richer countries) decreased from an average of 50% in 1975 to an average of 25.7% in 2010 and the nominal top personal income tax rates decreased from 70% to 41.4% over the same period, according to the OECD tax data base. Since the tax base in these countries expanded over this period, overall revenues as percentage of GDP remained stable (Dietsch and Rixen 2012: 5). What studies show is that, internally, the burden shifted among tax payers. Bigger corporations benefitted more from the decrease of tax rates than did small and medium companies. Also, the tax burden shifted from capital to labour. It has been observed that, if richer countries wished to prevent significant revenue losses, they could adopt regressive fiscal policies, notably by shifting the tax burden from capital to labour, and from taxation on revenue to indirect taxation of consumption (Dietsch 2015: 47-8).

In contrast, in developing countries, we do observe losses in revenues. The losses of revenue have not been compensated by a broadening of tax bases (Clausing 2016). Corporations and corrupt officials take advantage of weak or non-existing rules governing financial transactions. They not only become rich, they also weaken the institutions that are meant to sustain the jobs, living conditions and overall rights of the world's poor (IBA 2013). Tax abuses deprive agents, many of them poor countries, of the capabilities to create or strengthen the institutions that uphold political, cultural and socio-economic rights. And they augment countries' dependence on foreign assistance

and thereby diminish their financial autonomy. In sum, tax competition creates important distributive problems in developed countries and considerable revenue losses in developing countries.

In absolute numbers, in 2008, income shifting by multinationals deprived the US government of an amount somewhere between $60-$90 billion. Although data in a context of bank secrecy is hard to obtain, globally, estimates are that about $8 trillion (a lower-bound estimate) are located in tax havens, which represents about 8% of total wealth (Zucman 2013, Zucman 2014).

5.4 Assessment from the standpoint of justice

The empirical analysis of the previous sections will inform the normative analysis to follow. In lowering their tax rates to attract investments and capital, countries exert pressure on other countries to lower their tax rates in order not to lose capital and investments. Two points are worth mentioning. Internally, agents benefit unequally from it: corporations more than individuals and big corporations more than small ones. Globally, developing and poor countries are deprived of significant revenues. Although countries maintain their capacity to set the tax rates as they wish, they lose their '*de facto* fiscal self-determination' (Dietsch and Rixen 2012). Without regulation, there are risks of this phenomenon being exacerbated and countries being dramatically deprived of their capacities to act upon justice internally. A state's capacity to implement justice within its borders can thus be undermined. Whether this practice leads to a race to the bottom – all states entering competition and lowering their tax rates to a minimum – is debatable (Ronzoni 2009, Ronzoni 2014). What the data suggests is that the race to the bottom has been prevented in richer countries precisely because the tax burden shifted, exerting more pressure on immobile factors like labour and smaller companies. Even if the race did not end completely at the bottom, injustices have nonetheless been created (with regressive fiscal policy in rich countries and shrinking tax revenues in poorer countries as shown in the previous section). This has heavy consequences. Illicit

financial flows and tax competition infringe state sovereignty in forcing states to forego a welfare system, or at least in reducing significantly their chances to implement a welfare system. Undermining their internal capacity to act might lead states to lose the power to control their economies or to monitor external interference. Empirical evidence seems to point in this direction (TJN 2005, Avi-Yonah 2007, Dietsch and Rixen 2012, OECD 2014). If countries wish to prevent revenue losses, they have to adopt regressive fiscal measures.

In sum, if countries respond to tax competition by diminishing the burden on capital income relative to labour income, they would be undermining attempts to address income inequality through the tax system. In more affluent countries, this change in the ownership patterns of taxation is not always reflected in the overall level of tax revenues. Smaller companies tend to bear an always-heavier tax burden than bigger corporations and similarly labour tends to bear an always-heavier burden than capital. The internal injustice of these shifts in ownership patterns created by tax competition reflects the loss of governments' capacity to act upon justice.

To pursue the ethical analysis, we may categorize the first two forms of tax competition as virtual tax competition, as opposed to the third form (FDI), which we may call real tax competition. Virtual tax competition is composed of the two first forms of tax competition mentioned above: for portfolio capital and for paper profits. "From a legal perspective, there is a difference between these two cases – the first is considered evasion and illegal whilst the second constitutes legal tax avoidance – but from an ethical perspective, there is no difference between the two." (Dietsch and Rixen 2016: 14). I would rather say that there is no difference except the fact that FDIs stimulate other aspects of the economy (which in turn may raise other tax revenues).

This is only a brief summary of the phenomenon of tax competition, but it indicates we have reasons to believe tax competition raises a problem from the standpoint of justice. As mentioned, the task of this chapter is less to formulate original principles of justice to tackle this problem, than to show the normative elements that must be considered by any theory of global justice wishing to address it. The primary objective of this chapter is more to show how bottom-up non-ideal theory works than to

present a full-blown non-ideal theory to tackle tax competition. I want show how to do non-ideal theory, notably by examining the role of actual non-compliance, feasibility and path-dependence for this level of theorising. The presentation of the problem in this section should help us in observing in the following sections that the complexity of the problem of tax competition reveals normative elements that theories of justice must consider.

Now, one relevant question is why are governments entitled to tax the revenues produced within their national boundaries. Dietsch invokes in this context one crucial idea that is common in the international tax law literature: one has a duty to contribute to the public goods where one conducts one's economic activities, because individuals and firms are part of a system of cooperation (Dietsch 2016: 235). This system is only possible because public goods and infrastructure are provided by governments. Therefore, natural and legal persons should pay taxes in the countries where they benefit from infrastructures. This underlines that what is under scrutiny here is not how agents spend their money, but how they receive it (Dietsch 2015: 90). Profits are not created by the single hand of capital owners. I take the idea that governments are entitled to tax the revenues produced within their national boundaries to be considerably weighty from a normative standpoint.

5.5 Non-ideal principles for tax competition

This section presents insights about the ways non-ideal theorising contributes to solving the problem of tax competition. I will begin by quickly reviewing a few issues with ideal theorising that are avoided by non-ideal theory. One idealisation avoided by non-ideal reasoning is conceptualising the reflection about global taxation by abstracting from the world as divided into states, as the division of states would presumably consist in a morally arbitrary division which would itself be the cause of much injustice. With all that has been said so far about action-guidance, I do not need to dwell much on this point. But it is important to remind ourselves here that making such an idealisation

would make us mischaracterise the very subject of our inquiry. Whether or not the world would be more just without it being divided by national borders is speculative and irrelevant. What is pertinent to the discussion is to know that the problem of tax competition arises because capital mobility is global and tax policies are determined domestically. Not taking seriously the existence of states, with their interests and prerogatives, capabilities and submission to private interests, is to misunderstand the problem. Whatever regulation, especially of international coordination, has to be sensitive to the powers and limits of nation-states today. A second potentially harmful ideal principle of institutional design is the harmonization of global tax rates as a simple solution to tax competition. Preventing countries from adjusting their tax rates in moments of economic growth and recession is probably dangerous economic policy (Ronzoni 2014, Dietsch 2015: 118). Governments require flexibility to raise taxes in moments of great expenditure and to diminish taxes in times of unemployment. Arguing for the harmonization of global tax rates as a simple solution is potentially damaging. More sophisticated principles such as unitary taxation and formula apportionment (UT+FA), which I mentioned above and I will return to below, are more tailored to guide action in non-ideal circumstances.

Moving to the level of non-ideal theory, recall we distinguished the three forms of tax competition – in relation to portfolio capital, paper profits and foreign direct investments (FDI) – and two ways via which the autonomy of states is undermined: the 'poaching' (when the capital owner does not follow the investment) and the 'luring' (when countries attract FDI by making investments part of their tax base) of taxes. For the large affluent countries, if governments wish to prevent important revenue losses, they have to adopt regressive fiscal policies, notably by shifting the tax burden from capital to labour, and from taxation on revenue to indirect taxation of consumption (Dietsch 2015: 47-8). For less affluent countries, the result was more directly a net and massive loss of revenues.

It follows from the assessment of the practice from a normative standpoint that the involuntary deprivation of resources needed to act upon justice internally, and its consequences on aggregate welfare, is a sufficient reason to regulate this practice. States

should not suffer from undue external interference, and thereby lose their capacity to express a collective self-determination that should have a say both on the size of their public budget and on their redistributive capacities.

Dietsch and Rixen formulate two principles to regulate the three different kinds of tax competition: the membership principle and the fiscal policy constraint (Dietsch and Rixen 2012, Dietsch 2015: 80). The membership principle targets competition for portfolio capital and paper profits. It requires that natural and legal persons pay taxes in the countries in which they benefit from the existing infrastructure. The membership principle is associated with a demand for transparency. Individuals and corporations must make their income data available to tax authorities. The consequence of eliminating the competition for capital – recall the substitution effect mentioned in 5.2 – is that it would increase the competition for real investments. This is when the second principle comes into play. The fiscal policy constraint forbids tax policies that are simultaneously unjust and strategically motivated. It states that:

> A tax policy is legitimate if it does not produce a collectively suboptimal outcome. A collectively suboptimal outcome is here defined as one where the aggregate extent of fiscal self-determination of states is reduced (Dietsch and Rixen 2012: 13).

Respecting states' fiscal self-determination is morally required for states to act upon justice domestically and for states to be able interact freely at the global level. Following what has just been said, a practice that infringes on the fiscal determination of states is a practice that constrains states' capabilities to implement programmes and provide public goods its society desires. The intuition behind regulating tax competition is to promote justice by preventing injustice in the first place. Its objective is to allow for better national distributive capabilities, including the capability to maintain the overall size of a national budget and the capability to determine the internal distribution within this budget.

We could pursue this reasoning and promote the regulation of tax competition based on other notions than the reduction of states' resources, and its associated pervasive impact on individual opportunities and the autonomy of states. For instance, an argument could be developed centred on the notion of fairness of interaction. We

may assess the practice of tax competition and regulate it on the basis of fairness of interaction by arguing that states should remain capable of interacting fairly in the international arena (Ronzoni 2009). States as moral agents should be able to act as free and effective agents. Moreover, it is the case that there are various responses to these situations depending on one's conception of fiscal autonomy. Whether we agree on an ideal theory of autonomy is not a precondition as to whether we will find an effective mechanism to address the problem (Dietsch 2016). In other words, there are different conceptions of autonomy and different normative bases (e.g. impact on individual opportunities, or fairness of interaction) that allow us to move forward to a non-ideal level analysis. Again, ideal theorising does not seem to provide the way to set the target for non-ideal theorising. The action-guiding normative frameworks will have elements of ideal theorising and a solid non-ideal analysis based on the nature and on an account of the practice in real-world circumstances. We can thus move beyond the justification of why we should regulate tax competition and point out rather how non-ideal theorising contributes to its regulation.

5.6 Interpreting principles in non-ideal circumstances

It is clear now that the methodology of non-ideal theory is effective in pressing us to address the problem of tax competition directly and individually. We have sufficient reasons to believe tax competition is a cause of injustice in the world today. No ideal theory is required for us to address this problem – I take that the intuitions of injustice noted above are grounded following 5.3, considering the regressive policies and loss of revenues they cause. And we know enough about the specificities of the practice to guide us in the formulation of principles (much more than the thin lines exposed in this chapter). Now, I wish to point out how the content of a non-ideal theory, including the principles formulated above, needs to be fact-sensitive and target compliance.

First and foremost, the principle of states' self-determination needs to be balanced against the principles of transparency and information exchange. Values of transparency

and self-determination cannot be balanced against each other abstractly, in ideal theory. For instance, one idea put forward by the OECD consisted in providing tax 'information exchange on request' (OECD 1998). Yet, for a country to gain access to other countries' information on individual cases it needs to provide initial proof of tax evasion. Until recently, such proof was very hard to build. More robust measures based on routine verification and multilateral exchange of information are necessary (Dietsch and Rixen 2012: 21). This shows why addressing actual non-compliance is a constitutive part of a non-ideal theory of tax competition. Asking agents to report on audited financial statements and to foster banking transparency is central to the formulation of this conception of justice. That is, the precise design of principles necessary to enhance states' distributive capabilities requires the balancing of their self-determination and their engagement for financial transparency. Today, many countries have agreed that automatic exchange of information (AEOI) should be the new standard in global tax matters (OECD 2016).[87] Mechanisms favouring compliance with non-ideal level regulations are constitutive of a theory wishing to guide action towards more desirable social settings.

An example hinging on feasibility, fact-sensitivity and compliance, to which I promised to return, is the 'arm's length principle'. In the example about transfer mispricing in section 5.2, the office supply company based in the US declared low profits in the country. It also manipulated transactions so as to appear to have sold its goods and made a huge profit – tax-free – by 'paying' its subsidiary in the Cayman Islands an exorbitant price for pens, paper and laptop bags. In principle, such transactions should not have been permitted because of the arm's length principle, according to which transactions have to be conducted as they would have been with an arbitrary third party. This has as a consequence that companies can only buy and sell at the fair market price. Not only the 'arm's length principle' is a bottom-up principle about fairness of interaction, but we need a non-ideal theory level of analysis to assess its adequacy. The OECD transfer pricing guidelines are based on this principle (OECD 2011). This is a

---

[87] I thank an anonymous reviewer at MOPP for pointing this out.

very specific principle, based on impartial fairness in transaction that needs to be enforced as part of a general proposal for curbing tax competition. We will need to be sensitive to facts and non-compliance, not only to make sure the OECD guidelines provide effective means for countries and courts to assess transfer pricing practices of multinationals, but to make sure more complex cases can be addressed. This principle is indeed particularly complex to monitor and to enforce (Avi-Yonah 2016). It does not effectively prevent companies from engaging in profit shifting. Also, there are complex cases that pose a challenge to it even if it was appropriately enforced. Consider the case of intellectual property rights, a feasibility challenge for the successful implementation of this principle. A company could sell coffees in the US and write in its books that it paid its subsidiary in Ireland a high price for the intellectual property rights on its famous 'Makkachino'. Intellectual property rights are more difficultly assessed under the arm's length principle. These are questions that a non-ideal theory of tax competition needs to discuss in order to guide action in the non-ideal world.

This allows me to return to claim (iii) and the other elements that define the content of non-ideal theory. Clearly, in order to prevent injustices caused by tax competition, the role of non-ideal theory is not to determine what to do when agents do not comply with ideal principles, but how to make agents comply with non-ideal principles. The question of compliance in non-ideal theory is motivated by actual non-compliance as the core of the problem. It is not a question that can be avoided by non-ideal theory, but one that should structure it for the identification of more desirable settings to be possible. Moreover, for regulations to effectively target compliance, they have to be fact-sensitive. It is in relation to real world facts that we assess the desirability of proposals such as the harmonization of tax rates. Also, for principles of justice that target tax competition (such as Dietsch and Rixen's two principles) to be feasible, mechanisms that make the implementation of principles feasible have to be enforced. The agreements on AEOI of the OECD have to be in place for membership taxation to be monitored. This is the kind of non-ideal theory feasibility issues that we must keep in mind. In the same vein, it is crucial to understand why a principle such as the 'arm's length principle' is ineffective in the present context. In today's context, too much is

based on a principle that has very little chance of succeeding. A question for a non-ideal theorist is whether to build mechanisms around this principle to make it effective or find an alternative principle. I turn to this analysis now.

5.7 Institutional design and path dependence

For Avi-Yonah, we should indeed favour a system of unitary taxation and formula apportionment (UT+FA) instead of the arm's-length standard. UT+FA is the proposal according to which corporations would have to produce a combined report on their global profits, which would be distributed to the different tax jurisdictions in which the corporation is active, following a predetermined formula based on indicators of real economic activity (Avi-Yonah 2016: 289). Avi-Yonah notes that this system is not feasible today given the resistance from the OECD and many countries. He argues for an intermediary step that use FA only in the cases where profit shifting could be foreseen. These cases are those where the 'arm's length principle' fails. They are, therefore, the cases that can be resolved by looking at the profits realised in different countries. The more ideal UT+FA setting is path-dependent on this intermediary system. Contrastingly, relying on the arm's-length principle does not allow for the required path to be paved. It has been observed that the OECD's (2013) efforts to repair the arm's length standard have in fact contributed to a cat-and-mouse game, where the regulators are always one step behind (Avi-Yonah 2016, Dietsch and Rixen 2016). Following the arguments about dynamic duties and path dependence in Chapter 4, and the normative assessment of tax competition as a problem of justice that must be addressed, if Avi-Yonah is indeed right about the effectiveness of UT+FA in tackling tax competition, we may formulate the following argument.

(P1) We have moral reasons to seek to implement UT+FA, for it allows tackling tax competition effectively, but UT+FA is not feasible today;

(P2) UT+FA is path-dependent on the implementation of FA;

(P3) FA is feasible today;

(C) We have a moral duty to implement FA.

This allows me to return to claim (iv). We observe that assessing path-dependence is crucial for the formulation of an action-guiding theory and that path-dependence is best determined by non-ideal level theorising. (C) is justified on the basis of (P1), just like the second section of the argument S2 in section 4.5 was justified on the basis of S1. (P2) and (P3) concern respectively path-dependence and feasibility. This is an example of a non-ideal theory moral argument in favour of the regulation of tax competition.

I will now pursue the reasoning on path-dependence and institutional choices in order to show that establishing premises like (P2) above play a crucial role in non-ideal theorising. Note that path-dependence is observed not only in the choice of a system of regulation, but also on the very way to approach the governance of international tax interactions. "Global tax governance thus consists of the set of institutions governing issues of taxation that involve cross-border transactions or have other international implications" (Dietsch and Rixen 2016: 3). Governance involves hard choices such as the choice between limiting the mobility of capital by shifting power to an international body and finding mechanisms of global governance that enhance the capacity of states to tax mobile capital. Exactly what level or what dimensions of a nation's power to tax need to be shifted to the international level is a question beyond the scope of this chapter. What is central for the purposes of this thesis is to observe that the way to organise collective action in the case of tax competition can only be given in non-ideal theory.

It is clear from what has been discussed so far that given the externalities and interdependence created by global capital mobility, a number of issues about taxation cannot be adequately addressed within the nation-state. Central to the institutional design of the response today are the OECD non-binding recommendations around the provision of technical expertise and the diffusion of good practices. However, we must observe that this is a response to a coordination game with a distributive problem, to wit, a bilateral approach to double tax avoidance (Rixen 2011a), codified in the international double tax avoidance (DTA) regime (aimed at avoiding double taxation). Contrastingly, the problem of tax competition today should rather be seen as an asymmetric prisoner's

dilemma and not simply a coordination problem.[88] The response to this problem should focus on multilateral governance with enforcement capabilities instead of coordination (Rixen 2011a, Dietsch and Rixen 2016: 6-7). The analysis of the practice in a non-ideal world is crucial to understand the type of response needed.

The fact that global tax competition exhibits path-dependence is particularly important in the context of this thesis. In other words, following Rixen, tax competition should not only be seen as a natural consequence of globalization, for careful attention should be paid to prior endogenous institutional choices (Rixen 2011a). The central institutional response is path-dependent on the measures implemented because the very problem of tax competition exhibits path-dependence. In double-tax avoidance (the international DTA regime), governments did not want to endanger the solution they had institutionalized before tax competition became a real problem. A coordinating function was established and countries did not want to jeopardize it. But an unintended consequence of the institutional design of double-tax avoidance is under-taxation, which comes with the negative consequences already depicted above. This contributed to the creation of an asymmetric prisoner's dilemma scenario given the conflict of interests among big and small countries (and the powerful business lobbies of these countries). That is, the roots of the problem of tax competition can be traced back to a response to a previous problem, from which agents did not want to resile (Rixen 2011a: 198). The response to a coordination game with a distributive effect (avoiding double taxation) is not the response required to an asymmetric prisoner's dilemma.

Path-dependence is observed between the individualistic and selfish actions of individual agents and the lack of adequate institutional responses. Global tax governance

---

[88] Following Rixen (2011a) and Dietsch and Rixen (2016), an asymmetric prisoner's dilemma is characterised by situations with strong conflict of interests between the parties. The following strategic structure can be observed: "one party (tax haven) has deadlock preferences, i.e. it not only prefers defection over co-operation in individual strategies but also prefers the outcome of collective defection over the outcome of collective co-operation. The other party (big, developed country) has prisoner's dilemma preferences, i.e. while it prefers defection over co-operation in individual strategies it prefers the outcome of collective co-operation over that of collective defection" (Dietsch and Rixen 2016: 7 note 6). This structure is distinct from the strategic structure of the regular (symmetric) prisoner's dilemma, in which both parties prefer the cooperative outcome over the uncooperative outcome.

has played an important role in creating the problem of tax competition. This shows that assessing path-dependence is indeed the province of non-ideal theory. The distinction between the justified and the not justified targets is given through non-ideal theorising. Claim (iv) thus indeed provides a better understanding of the notion of path-dependence than the path-dependence argument formulated in 1.6, according to which ideal theory identifies the target that we should pursue. How to understand a problem and how to respond to a problem in non-ideal theory is what dictates what we will be in a position to accomplish next. The more desirable institutional setting certainly depends on what has been done previously. And awareness about how to pave a way is given by a non-ideal institutional response. That is, if anything, non-ideal theory can create a path to ideal theory, but not the other way around, thereby confirming (iv). Moreover, the questions related to path-dependence in non-ideal theory are sufficiently interesting and complex to be dealt with independently of ideal theory.

To conclude, in terms of global tax governance, a more ideal institutional proposal could take the form of an International Tax Organization, the ITO, as modelled by Rixen and Dietsch (Dietsch and Rixen 2012, Dietsch 2015, Rixen 2016). The ITO would have the objectives of overseeing and enforcing the regulations targeting the different forms of tax competition. Given the complexity of the problem, its globalized scope and its asymmetric prisoner's dilemma frame, only an institution with global reach, strong monitoring, enforcement mechanism and comprising a dispute-settlement procedure, could effectively tackle tax competition. The model must replace the bilateral mode of interaction prevalent in international taxation and must offer a legalized dispute-settlement mechanism similar to that of the WTO.

5.8 Conclusion

This chapter provided a general account of the phenomenon of tax competition (5.2) and offered an assessment of the practice from a justice standpoint (5.3 and 5.4). We have seen that tax competition has a profound effect on states' fiscal autonomy,

which allowed us following the normative framework of chapter 3 to address this practice individually. Sections 5.3 to 5.7 supported the case for (ii): the complexity of tax competition supports the case for a bottom-up approach to global justice, i.e. principles should target practices individually. I have offered in this chapter an example of bottom-up non-ideal theorising, while specifying elements that should constitute the content of a non-ideal theory of justice that targets tax competition.

In section 5.5, 5.6 and 5.7 we saw how an adequate account of the practice revealed the normatively relevant elements that a theory of institutional design should consider, and what principles of justice and institutional framework could effectively address this problem. This analysis allowed me to conclude that a non-ideal theory of justice targeting tax competition offers specific tools to guide action towards global justice. We saw in these sections that an adequate analysis of the practices provided insights for the formulation of principles (on the grounds of fiscal self-determination, transparency and information exchange). Not only the complex nature of individual practices require principles tailored to address these complexities (think about the overcoming the deficiencies of the arm's length principle to curb transfer mispricing), but the formulation of these principles can be motivated by the values of fiscal sovereignty, of aggregate welfare and of fairness in transaction (we could argue that the formulation of non-ideal principles can be justified and motivated by different ideal values). The three values of fiscal sovereignty, aggregate welfare and fairness in interaction can play complementary roles in curbing tax competition.

Sections 5.5, 5.6 and 5.7 provided support to claims (iii) and (iv). Determining our duties of justice in a non-ideal world is not a question about non-compliance with ideal principles but rather about compliance with non-ideal principles. This shift from the standard interpretation of the question was justified. Non-compliance is at the very core of non-ideal theorising. To interpret the question of compliance by asking whether our duties to comply with ideal principles are affected in cases of partial compliance is to misunderstand the role of non-ideal theories. Also, we saw that non-only tax competition exhibits path-dependence – in the sense that prior endogenous institutional choices condition the actual state of tax competition – but the possible paths towards

more ideal settings have to be determined by non-ideal level analysis, such as in the choice between an UT+FA formula and OECD coordination mechanisms. The non-ideal theorist should bridge the gap between the values underpinning our moral assessment of contemporary problems of justice and the regulations that can provide guidelines to agents today.

6. CLIMATE JUSTICE

6.1 Introduction

This section presents a case study in support of the 'reflective integration thesis' and the 'non-ideal theory thesis'. It aims to offer a more detailed account of the internal mechanisms of the reflective integration thesis and claims (i), (ii), (iii), and (iv) of the non-ideal theory thesis. Sections of this chapter have improved on material that has been presented elsewhere (Gajevic Sayegh 2016).

Central to this chapter is the interpretation of principles of climate justice in non-ideal circumstances. As in the previous chapter, the claims composing the non-ideal theory thesis will be supported by an examination of the components of an action-guiding theory of institutional design: compliance, feasibility, fact-sensitivity and path-dependence.

Many central features of the climate justice debate are beyond the scope of this chapter. For instance, I leave the discussion between integrationist and isolationist approaches and the distinction between harm-avoidance justice and burden sharing justice for the next chapter. My contribution operates at a different level from many of the main contributions to the climate justice debate. In the spirit of this thesis, the objective of this analysis is to determine whether inputs from facts and other disciplines (climate science, economics, social sciences) provide normative supplementations to the philosophical inquiry.[89] I will argue that they do. This chapter focusses less on whether principles of climate justice – such as the 'equal emissions per capita principle', the 'equal sacrifice principle', the 'historic responsibility principle' and 'the capacity

---

[89] Again, this claim is not controversial, but not sufficient attention has been paid to how this happens. I provide in this thesis examples that show how inputs from other disciplines constrain the normative inquiry of political philosophers.

principle' – are justified morally, than on what kind of real world challenges we face when making sense of them, should we choose to implement these principles. Again, the primary objective of Part II of this thesis is to show how non-ideal theory works. Yet, just like in the previous chapter, elements that should constitute the content of a non-ideal theory of climate justice will become apparent. Illustrating how non-ideal theory works and offering a new account of bottom-up climate justice are two distinct objectives and both are important. The former is the focus of this chapter and the previous one, while the latter objective is pursued in Chapter 7.

6.2 Climate Change

Climate change consists in 'any systematic change in the long-term statistics of climate elements (such as temperature, pressure or winds) sustained over several decades or longer' (American Meteorological Society cited in Dessler 2012: 4). It involves alterations in the distribution of weather patterns or, in other words, changes in the patterns of statistical description of weather over a period of time.

One aspect of the climate challenge is that it poses problems that have no national boundaries. Damage caused by GHG emissions is largely independent of the location of emissions (Aldy and Stavins 2007). The emissions of polluting gases by one country end up in a globally shared atmosphere.

Moreover, we know today that climate change is not only a fact, but the cause of the phenomenon is proven to be anthropogenic (IPCC 2013b). For the first time in history, the human activity on earth has a significant and dramatic impact on the planet's climate. An increase of 3 or 4 °C in the atmosphere will change the way we live on this planet.

This chapter does not offer a solution. It aims rather at better understanding one aspect of the problem: bridging the gap between ideal principles of climate justice and real-world considerations that seem to heavily influence the institutional design of any solution.

6.3 Scientific Constraints

An initial element to consider in this case is a general understanding of the scientific constraints on climate objectives. The way in which scientists frame the problem defines a feasibility constraint. Recent studies have attempted and succeeded in defining the planet boundaries of the earth, that is: 'the safe operating space for humanity based on the intrinsic biophysical processes that regulate the stability of the earth's system' (Steffen *et al.* 2015: 1-2).

The planet has natural boundaries, and life on earth depends on our capacity to live within them. With increasing certitude, it is known that the safe operating space of carbon concentration in the atmosphere is situated below the threshold of 450 parts per million (ppm) (*Ibid.*). The concentration stood at 391 ppm of carbon dioxide in the atmosphere in 2011, increasing at a rate of 1.9 ppm annually (the 1995–2005 average) (IPCC 2007). We were at 397 ppm in the beginning of 2015 (Steffen *et al.* 2015: 3) and the 400 ppm mark was crossed in the course of the same year.

Establishing the correlation between carbon concentration and temperature increase is challenging. Yet reliable models allow us to say that there is a 99% probability that a 550 ppm concentration of carbon dioxide in the atmosphere will result in a 2 °C increase in average global temperature (Posner and Weishbach 2010). The chances of a 450 ppm carbon concentration inducing a 2 °C atmospheric temperature average increase are also considerable (78%). The 2 °C limit reflects the political compromise translation of the scientific constraint, which is the current one endorsed by the UNFCCC: to limit global warming to 2 °C above pre-industrial levels (UNFCCC 2014) with an effort to trying to keep it under 1.5˚C (UNFCCC 2015). Moreover, it has been pointed out that current economic models tend to underestimate the impact of dangerous climate change (as well as the benefits of a net-zero carbon economy) (Stern 2016).

In sum, no robust moral theory for climate justice can be formulated without considering the scientific imperatives that should constrain human action. I take it that,

given that a carbon concentration of 450ppm has a 78% chance of inducing a 2 °C atmospheric temperature average increase, the climate justice debate should use this a hard feasibility constraint. In other words, proposals of climate justice must make sure they respect the constraint from climate science.[90]

6.4 Climate Justice and Ideal Theory

Following this (overly brief) scientific summary, we may now focus on how the climate challenge becomes a problem of justice. In *A Perfect Moral Storm*, Stephen Gardiner offers a description of the field of climate justice. For Gardiner, climate change is a perfect moral storm, for it combines three major challenges that make it very difficult for agents to do the right thing (Gardiner 2011). We can postpone the challenge to future generations (the intergenerational storm). The fragmentation of agency means that we can only address this challenge in a collective effort (the global storm). And we do not possess a moral theory capable of guiding our political institutions (the theoretical storm). I will focus first on the theoretical challenge: no moral theory at the moment is sufficiently strong to guide collective action.

In order to meet this challenge, it is useful to distinguish between two aspects of the response to climate change (Armstrong 2012, Dessler 2012): the mitigation of climate change (to reduce its pace and magnitude), and adaptation to climate change (to reduce its impact). Mitigation seeks to prevent the negative effects of climate change. Adaptation refers to the ways we need to find in order to live with the consequences of climate change. The main objective of mitigation policy is the reduction of GHG emissions, notably changing the way we use energy, changing our ways to practice agriculture and limit deforestation (Risse 2012: 125). The central role of reducing emissions explains why, in the philosophical debate, principles about the just distribution of emission rights have taken such a central place (Neumayer 2000,

---

[90] This brief account of the feasibility constraint from climate science is sufficient for the purposes of this chapter. The next chapter provides further insights about the interpretation of the climate constraint.

Athanasiou and Baer 2002, Singer 2004, Gosseries 2005). Adaptation measures consist in controlling flood and droughts, building barriers against sea level rising and developing crops that resist droughts.[91]

Climate change becomes a concern of justice because mitigation and adaption impose burdens on agents around the globe to act. Theories of climate justice propose principles to effect the distribution of these burdens. They aim at formulating the principles and values that should guide us in society. These principles and values could be associated with particular institutional designs in non-ideal theory or be formulated as general principles. Two prominent positions in the climate justice debate have articulated two different principles for the distribution of our burdens to mitigate climate change: the 'principle of equal sacrifices' and the 'principle of equal emissions per capita'. Note, the first principle focuses on the distribution of the effort to mitigate (and potentially adapt to) climate change, whereas the second centres on the distribution of emissions rights (Caney 2011: 86-97 reviews different approaches for the distribution of GHG emission rights the climate justice debate). [92]

David Miller advocates the principle of equal sacrifices: countries should make an equal sacrifice to their standards of living in order to fight climate change. Miller argues that by emitting polluting gases, societies impose serious harm on other humans, especially on those in societies who are already poor. The first step of this argument consists in waiving poor countries (those that cannot lift their citizens above the poverty line threshold) from the responsibilities to decrease their emissions (Miller 2009: 125). They should be able to increase emissions in order to develop economically (Miller 2009: 146). A second step is to distribute emissions among the remaining countries. Miller

---

[91] Dessler places geo-engineering alongside mitigation and adaptation as a third potential (and risky) response to climate change (see Dessler 2012: 181). Less controversially and less risky, we should also mention the responsibility to fund the development and transfer of clean energy. Caney (2012: 259 n.6) also mentions it. I will return to responsibilities of this kind in the next chapter. The other strategy I will also address in the next chapter is compensation. I leave geo-engineering aside. I do not believe is a route we should envisage taking.

[92] As mentioned, in this chapter, I leave aside the important distinction between isolationist and integrationist approaches put forward by Simon Caney (Caney 2012). Caney provides an insightful critique to the 'equal emissions per capita' view. I take that some of the empirical analysis of the next sections could contribute to an isolationist and an integrationist perspective. I will return to this distinction in the next chapter.

rejects an equal per capita right to emit, since he observes that some countries have a greater capacity to fight climate change than others. If we stipulated that some countries should have the right to emit, say, one metric ton of carbon dioxide per person per year, some countries could do that fairly easily. The remaining countries would alone bear the cost of fighting climate change. Hence, the principle of equal sacrifices:

> the action that is required involves some sacrifice, and all those who can contribute without harm to their vital interests should do so on an equal basis. Some will need to do more than others physically, but the sacrifice, in terms of income or consumption forgone, should be the same for all (Miller 2009: 150).

All countries that can fight climate change without harming their vital interests should do so on an 'equal sacrifice' basis.

More conventionally, the standard egalitarian position for just emissions is the principle of equal emissions per capita: everyone should be able to emit the same amount of carbon dioxide per year (Neumayer 2000, Athanasiou and Baer 2002, Singer 2004, Gosseries 2005 defend different versions of this principle). The general reasoning behind this principle goes as follows: assuming that it could be determined what degree of climate change is acceptable and what scientific measure of emissions corresponds to this level, we could pinpoint the amount of carbon dioxide emissions to which everyone is entitled. In other words, we would determine what is the global 'GHG budget' and we would split this budget equally. This egalitarian principle states, drawing on the idea that the atmosphere is owned by all and no one should deprive others of the ability to use it, that there are no reasons why anyone should have a greater claim to pollute than anyone else (Singer 2004: 35). As opposed to Miller's account, this principle does not argue that we should distribute equally the cost of mitigating climate change, but rather that we should distribute equally the right to pollute (Armstrong 2012: 202). According to Peter Singer, developed nations should not have the right to pollute more simply because it would be costlier for them to reduce their emissions. Singer accepts nonetheless that the principle of equal emissions per capita can be paired with some form of emissions trading. Richer countries could still buy the parts of poor countries' emissions quotas. This would presumably still allow us to reach our pollution reduction objectives.

The next section provides qualifications on these two principles. My goal is not to support one of these principles, but rather to present the non-ideal circumstances that our theories of climate justice have to take into account and show how these non-ideal circumstances should shape the way we conceive our principles.

6.5 Historical Responsibility and Agents' Capacity to Act

There are two important factors that could be summoned to complement one or the other moral principles of responsibility to limit our emissions of GHGs. The 'equal per capita' principle or the principle for 'equal sacrifice' can be modulated by these two considerations. The second of the two is already explicit in the principle of equal sacrifice. And is a consideration that could be in tension with the 'equal emissions per capita' principle.

These two normative considerations are: agents' historical responsibilities for past emissions and agents' capacity to pay to mitigate climate change.[93] Countries have indeed different historical responsibilities with regard to the GHG already emitted (Gardiner 2011, Armstrong 2012, Risse 2012). The 'historical responsibility' line of argument maintains that the historical weight of past actions conditions agents' fair share of the burden. Secondly, not all countries have the same capacity to respond to the problem of climate change, understood as changing their ways of producing and consuming so as to reduce their emissions. The 'capacity to pay' line of argument holds that the more capable countries have a duty to help those less capable of meeting the climate challenge. I will discuss these two principles in turn.

Some countries are far more responsible for the current levels of carbon dioxide in the atmosphere than others. It is also true that some nations' wealth and levels of industrialization can be correlated with past emissions. The argument for the

---

[93] To these two normative basis for principles to distribute the burdens to mitigate climate change, we may add the idea of a principles based on the agents' benefit from the activities that have caused climate change. I return to this idea below.

responsibility for past emissions consists in claiming that some countries have used more than their fair share of total carbon emissions. Therefore, current emission targets must take past emissions into account (Shue 2009, Shue 2014). Not everyone agrees with this principle. Some argue that the distribution of emissions right should be 'history-insensitive' (Vanderheiden 2008: 229-230). The argument for including past emissions is meant to challenge, or nuance, the principle of equal emissions per capita, which unduly favours developed countries by forgiving them for past emissions.

Regarding past emissions, from a moral point of view, one must justify from which moment we should consider that countries became responsible for their emissions. For instance, Miller and Risse prefer not to hold countries accountable for emissions in periods they could not possibly have been aware of climate change. Risse (2012: 137) claims that, at the moment of the IPCC report of 1990, all countries were already aware of the climate challenge and of their duties to act upon it. I will return to this in the assessment of this principle in the next section.

The second principle I mentioned we may wish to take into account is a principle targeting agents' different capacity to act. This is based on what Miller calls the principle of capacity: "that remedial responsibilities ought to be assigned according to the capacity of each agent to discharge them" (Miller 2001: 460). Some countries have a greater capacity to tackle the problem. This ongoing inequality should make us more sensitive to the demand of differentiated responsibilities: countries that now have greater capabilities to address the issue should be held more responsible to do so and those that have citizens with greater basic needs unfulfilled should have less responsibility.[94] The principles of capacity is thus clearly a component of the 'equal sacrifices principle'. The more capable countries will contribute more to our fight against climate change than less capable countries, but the sacrifice in terms of consumption or income forgone will be the same for all.

However, the principle of capacity seems to be in tension with the 'equal emissions per capita' principle. Yet, as Singer suggests, the equal emissions per capita principle

---

[94] Caney distinguishes between different ways to interpret the notion of capacity. I return to this principle in the next chapter.

could be coupled with a mechanism for emissions trading. This would allow money transfers between richer and poorer countries. Moreover, since developed countries are normally high emitters per capita, an 'equal emissions per capita' principle should already be demanding on richer countries.

My point is not to argue for any of these principles, but to understand whether they are affected by empirical considerations and whether their real-world interpretation informs our normative reasoning about climate justice.

6.6 Interpreting Historical Responsibility and Capacity

Despite its strengths, some questions were left unanswered in the Kyoto Protocol. The Protocol only placed reduction objectives on developed countries. It was the feeling in 1998 that developing countries would not commit to a treaty that could undermine their more competitive position in world markets within a globalised economy. Today, although this feeling has changed, the distribution of responsibilities to mitigate climate change between nations change still raises numerous questions.

One of the challenges for an action-guiding theory, as mentioned in the previous chapters, and a characteristic of a feasible theory, is to be able to assess conflicts between principles and values. A theory of climate justice might need to strike a balance between a principle of equality (in one of the two forms seen above, for instance), a principle of historical responsibility, and a principle of capacity to act. Before balancing principles, a real-world interpretation of them is warranted. This is an attempt to make sense of these principles in the world today, not an in depth moral discussion.[95] My contribution is thus distinct from some of the interesting discussions on the topic on the climate justice debate. It is not to show whether these principles are justified morally but rather to show, should we choose to include them, what kind of real world challenges we face when making sense of them. However, in the spirit of this thesis, the objective of this

---

[95] Gardiner (2011: 414-420), Caney (2011) and Miller (2009) provide an insightful analysis on this.

analysis is to know whether inputs from facts and other disciplines (climate science, economics, social sciences) provide normative supplementations to the philosophical inquiry. In order words, I maintain that the empirical examination will provide useful inputs for the normative work. I will explore this in the next two sections.

In order to understand the data that is relevant to the analysis of historical responsibility, it is helpful to set three parameters. We must choose the period of emissions, beginning in say 1990, 1950 or 1850. We must then determine if we will opt for a per capita analysis, or take countries' total emissions. And we must choose to include or exclude land use and deforestation from our calculation, or to limit ourselves only to pollution from fossil fuel combustion.

Let us choose for instance 1990 as our benchmark, the year the first IPCC assessment report was published, which was used as the basis of the UNFCCC.[96] One may note that this date has relevance from a moral standpoint, as mentioned above, for around 1985 or 1990 the risk that GHG emissions would cause harmful climate change has already become common knowledge (Miller 2009). Say we include all carbon emissions and not only those originating from fossil fuels. And let us look only at total emissions per country. China would be our top emitter, with the United States, Russia, Brazil, Indonesia and India in order closing the top six. Japan, Germany, Canada and the UK would close the top ten (World Resources Institute 2015). These data are relevant. Five countries in the top ten polluters are developing or newly industrialized countries. If we excluded land use and deforestation, the same developing countries would be in the top ten, but more evenly distributed across it (with China second, Russia third, India fourth, Brazil seventh and Indonesia tenth). Note that including land use is relevant in many respects, such as in the interest of preventing deforestation to produce biofuels, which overall will not have a particularly beneficial effect on emissions reduction.

If we would choose instead the per capita emissions since 1990, Kuwait, Brunei, Belize, Qatar and Equatorial Guinea are the top five emitters including land use.

---

[96] As mentioned above, one may note that this date has relevance from a moral standpoint, for around 1985 or 1990 the risk that GHG emissions would cause harmful climate change has already become common knowledge (Miller 2009: 130).

206

Excluding land use, Kuwait, Brunei, Qatar, Oman and Trinidad and Tobago would occupy the first five positions. It seems the per capita ranking would not be so different if we took as our benchmark the year 1850.[97] We will see below what is the per capita emissions of the world's top polluters.

I only mentioned top emitters in all the examples. Many countries would have a valid claim that they did not emit so much in the past. Burundi, Chad, Mali and Ethiopia have emitted very little per capita since 1850. For another example, in total emissions, excluding land use, Kiribati, the Cook Islands, Nauru and Comoros are among the lowest emitters.

The point here is that when looking at the real-world data, it is not that straightforward to separate between developed and developing countries based on historical emissions. This analysis has to be more fine-grained than the two-block style analysis of the Kyoto Protocol, to which I will return in a moment.

Let us for a moment assume that 'distributing' emission rights on a per capita basis is morally justified (Caney 2012 challenges this) and that we should be operating within a carbon budget. The equal emissions per capita view should thus consider the emissions available per capita based on what nations have emitted since 1990 and, looking at real world data, it has to determine what counts as emissions. Now, whether it chooses or not to be sensitive to past emissions, it cannot differentiate between richer and poorer nations. The real-world analysis in this sense aligns with the moral analysis that the equal per capita view, even in a historically sensitive approach, is insensitive to the capacity of different agents to address the problem. In other words, by overly focusing on the distribution of emissions rights, an 'equal emissions per capita' view might neglect the legitimate right of poorer countries to develop. If this is indeed the case (Caney 2012, Caney 2014, Shue 2014 provide a compelling case for it from a normative

[97] One complication however is that the developing countries among the world's top emitters since 1990 in absolute numbers (with China, Russia, India, Brazil, Indonesia in the top ten), which will also be among the top emitters in the foreseeable future, are not important emitters since the pre-industrialized era. For us to choose 1990 as a starting date for considering people morally responsible for their emissions, we must thus insist on the idea that before that time, people were not aware of the risks posed by GHG emissions. Also, we have to rule out the idea that these people have benefited from the wealth created by these emissions. Their current level of industrialization should become morally relevant elsewhere.

207

standpoint), this view would face a problem that is familiar in climate negotiations roundtables and commonly known as the development challenge (Schelling 1992, Caney 2011: 83). The development challenge means that we could not simultaneously realise the goal of climate change mitigation and the goal of allowing poorer people to develop. There are ways to overcome the trade-off of the development challenge and we should note, as Caney rightly points out, that the protection of the atmosphere is a precondition of development (Caney 2011: 84). It seems nonetheless that the empirical and the normative analysis are aligned: the 'equal emissions per capita' view cannot overcome this tension. The tension between development and climate change mitigation is indeed at the core of climate justice discussions (and it will be examined further in this chapter and the next). This brings us to the real-world interpretation of the principle of capacity.

Should the countries that contribute more to mitigation be those that can afford to do more? The view that rich countries have more demanding obligations to mitigate climate change is one shared by many countries, inclusively at the table of negotiations for climate treaties (Posner and Weishbach 2010: 73).[98] A common view is that even if they should not be assigned disproportionate responsibilities in terms of reductions, they should at least lead the way in climate negotiations (for example, China stated in the context of the Kyoto protocol that its participation would be contingent on developed countries taking the lead in reducing their own emissions).

United States, China, India, Brazil, Indonesia, Mexico, South Korea, South Africa and the European countries are responsible for almost 90% of annual GHG emissions. In 2007, GHG emissions from the developing world exceeded the emissions from developed countries (World Resources Institute 2015). Today developing countries, home to many of the world's poorest, are responsible for much of the global emissions. Together, developing countries which account for 80% of the world's population are responsible for 63% of total emissions (2015 data), as opposed to 41% in 2004. China, today, is alone responsible for 23% of global GHG emissions. India is responsible for 5,8%. Per capita, China emissions (7.91 tCO_2e) are comparable to the UK's (8.64 tCO_2e),

---

[98] Posner and Weisbach are not in favour of this view.

which are both less than half of the US's per capita emissions (18.55 tCO₂e) and also less than Russia's (15.74 tCO₂e). India, in comparison, has less than a third of UK's emissions per capita (2.33 tCO₂e). Brazil's emissions per capita are slightly higher than China's (9.17 tCO₂e) and Indonesia's are somewhat equivalent (8.02 tCO₂e) [99]. The average for Europe's 28 richer countries is 8.22 tCO₂e. South Korea (13.22 tCO₂e), Mexico (6.19) and South Africa (8.86) also have considerable emissions per capita.

The data shows that the countries that are responsible for 90% of the world's total emissions have also somewhat considerable emissions per capita, with the exception of India and Mexico, which are the only countries below the European average. The development challenge is thus clearly observable, because many of the world's poorest people live in India, China, Indonesia and Brazil (over 700 million people in these four countries alone live under the US$1.25 per day poverty line). Given that emissions targets are set by countries, the question of how to differentiate between countries that should have emissions reduction objectives and those that should not cannot be based on poverty indicators alone.

The principle of capacity therefore also faces an important challenge. In theory, the principle of capacity, according to which countries that have citizens with greater basic needs unfulfilled should have less responsibilities, does not need to take into account the real state of global emissions. We know now that some of the world's top emitters today (which to a large extent are also the top emitters from 1990 in absolute numbers and per capita) have a substantial portion of their population living under the poverty line. Yet, it is clear that an exemption can be made for the individual countries which both are among the world's lowest emitters and suffer from endemic poverty.

That is, given that the global carbon budget is fixed, if the principle of capacity applied directly to differentiate between individual countries emissions reduction targets, a nearly impossible task (considering even the most optimistic political feasibility and the fastest speed for the socio-economic transition out of the fossil fuel

---

[99] All data in total GHG emissions including land use per capita available from the CAIT Climate Data Explorer. 2015. Washington, DC: World Resources Institute. FAO 2014, FAOSTAT Emissions Database. Available online at: http://cait.wri.org. http://www.wri.org/blog/2014/05/history-carbon-dioxide-emissions

era) would be given to the world's richest countries. One way out of this conundrum is not to apply the principle of capacity to the determination of emissions reduction targets, but to have countries with greater capacity to assist developing and poorer countries in their climate change mitigation efforts in some other way. It would appear that, as opposed to the 'equal emissions per capita' view, the challenge for a principle of capacity (or an equal sacrifice principle) is less that it would have unjust consequences than how to appropriately operationalise it. I will return to this point and to the development challenge below. Before, I wish to address two other real-world complications.

In sum, this section has shown that the principle for historical responsibility and the principle of capacity have to be coupled with real world data in order for us to know how the operationalization of the differentiation between richer and poorer countries could be morally justified.

6.7 Non-Ideal World Challenges

As mentioned, the focus of this chapter is not to know whether our obligations under ideal principles change in situations of partial compliance, a question with which discussions of non-ideal theory are overly concerned. Framing the question this way implies a presupposition of the primacy of ideal theory. As I argued in claim (iii), non-ideal theory should focus on finding ways for agents to comply with regulations in order for justice to progress. That is, the question is not whether our obligations change in situations of partial compliance but rather what explains partial compliance and how it can be resolved. In the same vein, theoretical challenges such as whether to factor in historical responsibility and agents' capacity in our principles of climate justice are informed by the way it can possibly be done in order to favour compliance with climate objectives in a way that is fair. The real-world data examined in the previous sections already provided insights about parameters to be considered for an institutional design of climate treaties. I wish now to point out two features of the world that should also

constrain our thinking in our design of climate treaties.

Problem 1. One first non-ideal world challenge concerns what I will call in this chapter and the next 'mitigation efficiency', in the following sense: the strategies that will allow GHG emissions reduction to be maximized given the feasibility curves here and now in a way that feasibility curves are expanded in the future to allow meeting global emissions reduction targets. The feasibility curves should be given by the thorough consideration of the emissions reduction strategies that are cost-effective today and these curves expansion in the future will be given by how the strategies today affect the cost of initiatives in the future. The next section provides a graphic support to discuss this idea.

Given the increase in emissions of past decades, the industrialisation of developing countries and the growth of the global economy that could exacerbate increase in emissions, achieving climate objectives must be done in the most efficient way possible. Time is of the essence. And today, it is in developing countries that we find the best opportunities, the most cost-effective ways, of reducing emissions. The overall cost of climate change will be drastically reduced by the inclusion of developing countries (Olmstead and Stavins 2010: 7). For instance, in order to achieve the necessary reduction of GHG emissions, it is much more efficient to develop an economy based on sustainable energy sources[100] than to go through the profound fossil fuel-based industrialisation rich countries experienced and then change all polluting facilities, power plants and goods into less polluting ones (Posner and Weishbach 2010: 88). The energy, agriculture and transport sectors all have good cost-effective mitigation options in developing countries. Favouring options which are climate efficient is a way to avoid the problem of 'wrongheaded industrialisation'.

---

[100] I use the notion of sustainable energy, or green energy, to refer to energy sources like solar power, wind power, geothermal power, hydropower and tidal power, which are renewable and have little impact on the environment. Hydropower is considered to be green energy source despite its resulting submersion of extensive areas upstream of dams. Nuclear power is normally not considered to be a sustainable energy source. I do not consider it to be.

Mitigation efficiency also motivates market-based approaches to climate change mitigation. The IPCC has also pointed out that the overall cost of mitigation will be reduced by the inclusion of non-OECD countries in the effort of mitigation. Indeed, emissions trading within and/or with non-OECD countries could lead to lowering the cost of mitigation and lead to more mitigation in these countries (IPCC 2001). Although it is unlikely that a global emissions trading scheme would be manageable, regional carbon pricing initiatives in places that have no such mechanisms could diminish the cost of the emissions reduction effort. The next chapter is entirely devoted to the question of carbon pricing.

The normative value of mitigation efficiency needs to be established in greater detail. Efficiency in this case has normative value because avoiding 'wrongheaded industrialisation' is a way to diminish the overall sacrifice to mitigate climate change. We have seen that developing countries such as China, India, Brazil and Indonesia had no emissions reduction objectives under the Kyoto Protocol. It seems to be the consensus today that these countries need to commit themselves to emissions reductions if we wish ultimately to reach our global objectives (IPCC 2001). In order for a principle of climate justice to provide guidelines that will respect our planetary boundaries and lead us forward in the architecture for agreement of climate targets, industrialised and developing nations need to be involved in a meaningful way (Olmstead and Stavins 2010). This is mainly due to a concern for mitigation efficiency: diminishing the overall sacrifice required to mitigate climate change is one of the best ways to ensure that global emission reduction targets can be met. The challenge, as we have seen, is to implement mitigation efficiency and avoid wrongheaded industrialisation in a way that is fair. I will argue that thinking mitigation efficiency and fairness together allows for an optimal way to pursue these two objectives. In other words, there are ways to thinking about investment in climate related projects that allow to relax the trade-off between efficiency and fairness.

Problem 2. A second non-ideal circumstance relates to 'carbon leakage'. If only some countries reduced carbon emissions in their energy sector, carbon-intensive

energy might become cheaper in other countries and encourage them to consume more. If emissions restrictions become too differentiated, countries that agree to take action will see an increase in the cost of carbon-intensive goods, while countries that have no emissions restriction will have a comparative advantage in producing them. One consequence might be a shift in the production of carbon-intensive goods to non-participating countries, a problem known as 'emissions leakage' (Olmstead and Stavins 2010). For instance, we could observe a capital flight to higher emitting jurisdictions. This could undermine the efforts to decrease the production of carbon-intensive goods, and it might not help lowering the cost of greener products, which could have detrimental economic effects. Yet, carbon leakage is a potential problem and should not be thought as a problem that should inhibit sub-global efforts of climate change mitigation (Farber 2013). However, the design of climate policy must be sensitive to the problem so that national and regional initiatives will not be undermined by carbon leakage, and the global emissions reduction target remains achievable.

One potential consequence related to carbon leakage is that too much differentiation puts developing countries on the path of producing carbon-intensive goods thereby diminishing their incentive to join a treaty, which would also be inefficient. The more countries increase carbon-intensive production, the more their economy relies on this production and the less they have incentives to join a treaty (an additional problem to the problem of wrongheaded industrialisation, for the countries in this case might not necessarily be developing countries). Principles of climate justice need to be sensitive to the question of emissions leakage in order to guide action in a meaningful way. We must seek reflective integration between our idealized principles and problems such as carbon leakage to identify feasible paths to reduce emissions and distribute responsibility to all relevant agents.

6.8 Insights for Reflective Integration

This chapter focusses on reflective integration and on some of the desiderata

associated with path-dependence, compliance and conflicts between values, which express the action-guiding capacity of a theory of climate justice. Tackling climate justice by adopting a method of reflective integration between ideal and non-ideal theory implies asking how non-ideal circumstances and empirical features of the world affects the formulation of our principles. At the level of ideal theory, we identified two versions of a principle for equality of emissions, a principle for historical responsibility and a principle based on agents' capacity to pay.

The non-ideal theory analysis was given in two steps. In the first step, we saw that the interpretation of the principles of historical responsibility and agents' capacity is not straightforward. Regarding historical responsibility, determining the starting historical point for consideration of agents' responsibility is a moral discussion, involving the consciousness of wrongdoing. We can make a strong case about the wrongdoing in the activity of polluting from 1990, although polluting gases have been emitted by human activity in significant quantities since 1850 and even more since the 1970s. Also, we have seen that by setting different pollution parameters and examining the relevant data on the matter a historical responsibility analysis does not allow labelling developed countries as 'historically responsible' and developing countries as 'not historically responsible'. If we take 1990 as our starting date, the date of the publication of the first IPCC report when climate change was presented to the world as an imminent challenge, both developing and newly industrialised countries are responsible for a considerable portion of GHG emissions.

Regarding the principle of capacity and the distribution of burdens to mitigate climate change, we have seen that the question of how to differentiate between countries that should have emissions reduction objectives and those that should not cannot be based on poverty indicators alone, given that some of the greater emitters today are also the home countries of many of the world's poorest people. We saw the tension, identified as the development challenge, where countries that are responsible for a large share of global emissions past (since 1990) and present, per capita and in absolute numbers, are also countries where a large percentage of the world's poorest live today.

Regarding the 'equal emissions per capita principle', we have seen that regardless

of it being sensitive for past emissions or not, this principle is insensitive to agents' capacity to respond to the problem, which would create considerable problems of justice. Contrastingly, the challenge for a principle of capacity (or an equal sacrifice principle) was not that it would have unjust consequences, but rather to find a way to operationalise in a way that would respect constraints for climate justice and climate science. The empirical analysis concluded that the principle of capacity should not apply directly to the determination of emissions reduction targets. I suggested that countries with greater capacity should assist developing countries in their climate change mitigation efforts in some other way.

In a second step of the non-ideal theory analysis, we saw two empirical facts of the world that should also constrain the way we design our institutional response to climate change. These problems were labelled 'wrongheaded industrialisation' and 'carbon leakage'. These problems warned us against too much differentiation in emissions reduction targets between countries. We would run the risk of observing carbon-intensive industries migrate to countries with no GHG emissions reduction legislation and provide incentives to carbon-intensive production within these legislations. On the contrary, we saw that the concern for mitigation efficiency requires taking advantage of low cost mitigation efforts in developing countries and the inclusion of developing countries in international mitigation policies, such as carbon pricing-related initiatives. The non-ideal theoretical analysis showed that the method of reflective integration contributed normatively to the philosophical work.

Following the reflective integration methodology, one of our tasks, a first desideratum of an action guiding theory of climate justice, is to assess conflicts between values and courses of action. The development challenge warned us that distributive justice considerations could conflict with our climate imperatives. Also, we have seen that if we applied the principle of capacity to differentiate between emissions reduction targets, and place the emission reduction responsibilities on affluent countries, we would be favouring an institutional design that is much more expensive and less politically realistic. Another action guidance desideratum requires that the non-ideal circumstances that influence the compliance of agents inform the formulation of

regulations. Carbon leakage could influence compliance in the effort of curbing global emissions.[101] But so will the abandoning of less affluent nations in their effort of climate change mitigation. A third desideratum demands that we clarify what feasibility constrains might relax in the future in order to know how it affects the appropriate timeframes for regulation and the paths of feasibility for future action.

The UNFCCC demand for 'common but differentiated responsibilities' could be interpreted as an attempt to fulfil these desiderata. This is an authoritative principle in climate negotiations not only in Lima (2014) and in Paris (2015), but since the 1992 Rio de Janeiro Earth Summit. This principle, in a way, is obtained by seeking reflective integration between ideal principles and non-ideal circumstances, for it tries to balance concerns of justice with the feasible paths for attaining climate objectives. The vagueness of this principle can be overcome by an interpretation of its 'common' and 'differentiated' aspects. Reflective integration shows us that an action-guiding interpretation of the principles of historical responsibility and agents' capacity highlights the importance for almost all countries to have emissions reduction targets. Only a few countries can be reasonably excluded from the global mitigation effort (such as Kiribati which is on the verge of disappearing, or Burundi which has virtually not contributed to climate change), provided their exclusion does not create or exacerbate the problem of carbon leakage. The 'differentiated' aspect and the concern for fairness will be given by a sensible timeframe of inclusion of all nations, with poorer nations given more time to comply, and by assistance mechanisms, such as technology transfers and funding for clean energy (the Green Climate Fund, about which I will say more below, is one major initiative of the sort, aiming to assist developing countries with climate change mitigation and adaptation). As I suggested, mitigation efficiency and fairness can be

---

[101] Caney (2014: 135) argues about the importance of designing the social, economic and political contexts in ways to induce agents to comply with ideal principles of justice or, more precisely, what he calls 'first-order responsibilities'. I argue that we also have to design these contexts in order for agents to comply with non-ideal principles of justice. Caney does not explicitly equate first-order responsibilities with ideal principles. These are responsibilities for agents to perform certain actions. Second-order responsibilities are those that some agents have in order to ensure that agents comply with their first-order responsibilities. I will argue below that Caney is right about the importance of designing these contexts, even in cases when we might not have a clearly determined view about what first-order responsibilities entail.

pursued jointly. The perspectives for mitigation efficiency in funding clean energy supply in developing countries are considerable. The moral duties of developed countries to mitigate climate change and to fight poverty can be fulfilled by assisting developing countries in their climate effort. Consider now figure 6.1 about the development challenge and mitigation efficiency. I consider for the purposes of discussion that we can apply the question of fairness to the climate change effort, which should cast light on the question of development more generally.

Figure 6.1. The Development Challenge and Mitigation Efficiency
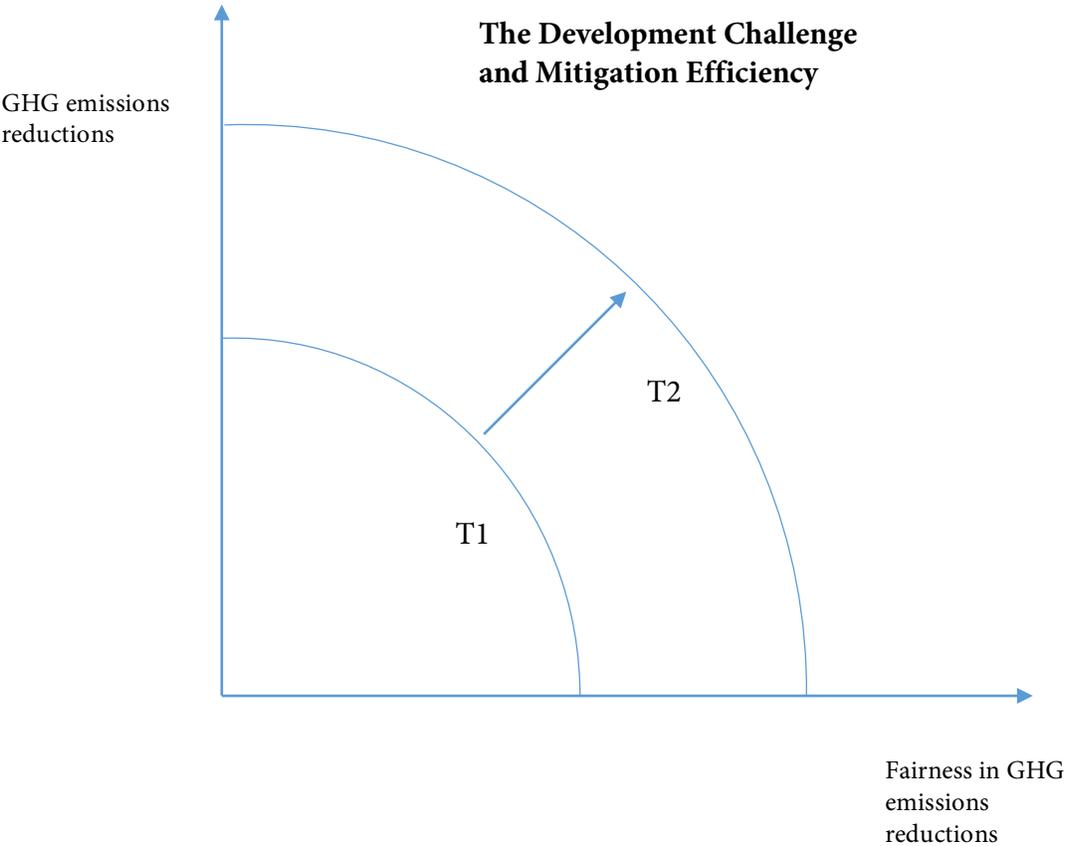
I defined 'mitigation efficiency' as the strategies that will allow GHG emissions reduction to be maximized given the feasibility curves here and now in a way that feasibility curves are expanded in the future to allow meeting global emissions reduction targets. The development challenge is modelled by the general existing trade-off between the two axes. There are various responses to the development challenge. The way

suggested in this chapter is that, given the low cost opportunities for climate change mitigation in developing countries and given the moral duty of more affluent nations to have more responsibilities in the climate challenge, providing developing countries with the means to develop, for instance, sustainable energy sources or sustainable agricultural practices, is a way to overcome the development challenge. That is, this is a way to jointly realise the goal of climate change mitigation and the goal of allowing poorer people to develop. GHG emissions reduction and fairness can be pursued jointly.

For the purposes of the argument, let us consider only the case of energy. Firstly, we must take into account that investments in sustainable energy do not achieve as much emissions reduction when they are an installation to augment the energy consumption in a given region as when they replace already existing carbon-intensive energy sources. However, there are also many sustainable alternatives today that already have important economic returns or are cost neutral (Stern 2007). Also, investments in green technologies drive down their cost, making the cost associated with choosing green alternatives considerably lower in the long run. There might always be alternatives that favour more one of the two objectives in figure 6.1. My argument is that the joint considerations of these different options for investing in sustainable energy that allow pursuing emissions reduction and fairness together – with alternatives that are cost neutral, profitable, less expensive, more expensive, including those that drive the cost of sustainable energy – should allow us moving from feasibility constraints of curve T1 in the figure to feasibility constraints of curve T2, rather than only moving on or under the curve T1. Integrating the emissions reduction strategies that are cost-effective today allows to expand the feasibility constraints for emissions reduction and poverty relief in the future, allowing to do more of the two. The curves expansion of the future will be given by how the strategies we choose today affect the cost of initiatives in the future.

The interpretation of the common and the differentiated aspects of the principle is therefore not given by a diminishing of emissions reduction responsibility, but by the assistance mechanisms that developed countries will provide to developing countries to fight climate change, among other options that also allow pursuing the objectives of GHG emissions reduction and poverty relief jointly. For instance, another way to

differentiate between responsibilities of richer and poorer countries is by an adequate timeframe of response.[102] The question of differentiated responsibilities can be addressed, at least partially, by robust assistance mechanisms, such as the Green Climate Fund, and by a transitional timeframe where developing countries are given more time to begin reducing their emissions, whenever both asked for and necessary.[103] It is unclear whether the Fund will succeed in its mission of raising US$100-billion annually by 2020, but this is clearly an avenue which allows for a better integration of scientific climate imperatives and distributive justice. In reflective integration between ideal and non-ideal theory, an appropriate interpretation of the principle of 'common but differentiated responsibilities' is a central piece of an action-guiding institutional design that addresses the relevant non-ideal circumstances in order to promote compliance, efficiency and fairness. This could result in the action-guiding principle of climate justice of institutional design we are looking for. In sum, our principle of climate justice should not seek for ways to waive developing countries of reducing emissions but finding ways for richer countries to help them meeting their emission reduction targets.

This chapter wanted to show what kind of conclusion could be reached by adopting the method of reflective integration. It seems that this method allows for integrating a principle of capacity and a concern for past emissions, while taking seriously climate imperatives and globe distributive justice, in a way that considers the expansion of feasibility constraints for future action.

---

[102] I thank Daniel Esty and Jeffrey Sachs for pointing out to me that allowing poorer countries more time to begin complying with emissions reduction targets, and doing so in a progressive way, is also an effective way to differentiate between responsibilities of more affluent and less affluent countries.
[103] Carbon pricing (as a way to promote cheaper and available alternatives to polluting energy sources) and divestment are two avenues to complement our principle of common but differentiated responsibilities, in order to make sustainable energies (like solar, wind, geothermal and tidal) some of the most accessible and affordable energy sources for all countries.

6.9 Conclusion

Today, the UNFCCC acknowledges that treaties should consider 'common but differentiated responsibilities and respective capabilities', and prominent positions push for climate treaties that seek differentiated but meaningful ways of including developing countries, such as through global assistance mechanism (e.g. the GCF), extended time path of targets and flexible market-based policy instruments (Olmstead and Stavins 2010). The objectives of fairness, cost-effectiveness and GHG emissions reduction can, and to a large extent must, be pursued jointly.

Discussions of climate justice cannot take place only in the abstract realm of ideal theory. The real world interpretation of ideal principles in reflective integration provides us with a real sense of their strengths and weaknesses. Reflective integration between ideal and non-ideal theory allows us to include an adequate temporal framework to address the problem. Climate negotiations require long-term objectives (the Kyoto protocol has been regarded as trying to accomplish 'too little too fast'). Emissions reduction could be distributed in a sensible timeframe, so as to leave more time for the emergence of greater incentives to meet these objectives (with the development of new technologies, carbon pricing and the shifting of consumption away from carbon-intensive goods), once emissions reduction targets are agreed upon.

Work in ideal theory remains relevant. It provides us with strong reasons for action, and presses us to consider difficult moral questions such as our obligations to future generations. But the formulation of action-guiding principles requires both ideal and non-ideal theory considerations. I have not argued against ideal theory. Rather, I have given an account of the use of ideal principles in non-ideal circumstances. Seeking reflective integration between ideal and non-ideal theory casts light on how a conception of justice may guide action in the real world. Although a full-blown theory of climate justice could not be developed here, we were provided with further insights about how the relationship between ideal and non-ideal theory works. This chapter focused on the role of reflective integration in the formulation of principles and on the normative contribution of non-ideal theory to political theorising.

Whether the premise that only a universally global effort is necessary for national initiatives to be meaningful is valid is a question I leave for the next chapter. Yet, although the lack of a concerted effort should perhaps not inhibit sub-global initiatives, a concerted global effort can clearly contribute to a cost-effective and fair way to address climate change mitigation.

7. CARBON PRICING

7.1 Introduction

Nations have come together to debate what should be done about the risks posed by climate change to our environment, economies and societies.[104] Market-based approaches feature prominently around climate roundtables and progressively become an important aspect of mitigation policies. The two principal market-based policy instruments are the carbon tax and the cap-and-trade system (Stiglitz 2006, Weitzman 2013). They have two features in common: efficiency (achieving emissions reduction at a lower cost) and distributive implications (they generate revenues) (Olmstead and Stavins 2010, Bowen 2011). Both approaches involve putting a price on carbon. The core justification for pricing carbon rests on the economic analysis of 'negative externalities': situations where the effects of the production and consumption of goods and services impose costs on others, which are not reflected in the price charged for these goods and services (Bowen 2011).

The literature on economic instruments to address greenhouse gas (GHG) emissions dates back to the 1970's. Yet, the existing literature lacks materials that integrate the ethical and the distributive components of market-based instruments. This absence is salient, considering that the distribution of burdens in the emissions reduction effort between countries is among the key obstacles to implementing an international agreement.

In order to bridge this gap, this chapter aims to model a theory of climate justice in practice. As opposed to the two previous chapters, this chapter develops in more detail

---

[104] This chapter benefited from invaluable comments from participants at the second ESRC Seminar Series on Climate Ethics and Climate Economics, held at the University of Nottingham, 13-14 April 2016.

the content of a bottom-up theory of climate justice as it relates to carbon pricing. It will track the 'reflective integration thesis' and the 'non-ideal theory thesis' in order to observe how theoretical considerations of climate justice applied to climate policy could result in an action-guiding theory of climate justice. This chapter focuses on one particular case that connects climate justice and climate economics. It addresses the following question: to what extent can market-based instruments for climate change mitigation (MBIs) respond to requirements of justice? Or, similarly, do regional carbon-pricing policies need to be aligned with principles of global climate justice? Distributing the agents' share of the burden in emissions reduction is an important step in the architecture of climate agreements (Bell 2008, Posner and Sunstein 2008, Miller 2009, Posner and Weishbach 2010, Shue 2014).

The contribution of this chapter is twofold. Firstly, it is to provide a sound normative foundation for a bottom-up approach to climate justice, and for carbon pricing mechanisms, around the notions of a 'right to energy', the 'duty not-to-harm' and a 'capacity principle'. I argue that we can structure these three notions in a cohesive proposal. Secondly, it is to identify the normative elements from theories of climate justice that should constrain the design of MBIs so that these become instruments of justice. This chapter aims to pave the way for a design of MBIs that balances requirements of climate ethics with the emissions reduction potential and the social co-benefits of different distributive alternatives. In line with the conclusions of Chapter 6, I will argue that, once we consider jointly emissions reduction targets, efficiency and fairness, the best course of action is to design MBIs so as to invest and provide incentives to lower the price of green alternatives, in order to assist developing populations in their climate change mitigation effort, and fund the transition to a low-carbon economy. MBIs help to expose and relax the trade-off between efficiency and fairness in climate policy debates.

This chapter proceeds as follows. Firstly, in sections 7.2-7.4, I will outline a broad-brush portrait of theoretical considerations about climate justice that have not been exposed in Chapter 6. I will situate climate change mitigation in this context and contrast the view championed in this chapter – a bottom-up approach – with other

approaches to climate justice (holistic, atomist, integrationist and isolationist). Secondly, in sections 7.5-7.6, I will explain the basics of carbon pricing mechanisms. In sections 7.7-7.9, I will argue that carbon pricing is a strategy that allows balancing the imperatives from climate justice (for the effort to mitigate climate change to be fair), climate economics (for the effort to be cost effective) and climate science (for it to reach the emissions reduction objectives and respect our planetary boundaries).

Basically there are three steps to my argument where I will raise ethical considerations: at the level of the normative foundations of MBIs, at the level of solving the internal problems of justice it raises, and at the level of solving external problems of justice. Understanding how to balance ethical questions in practice should inform us how to devise a better theory.

A few preliminary remarks are warranted. It is important to note that MBIs are normally only one initiative within a larger emissions reduction strategy. MBIs only contribute to the reduction of part of a nation's GHG emissions. Secondly, my goal is to explore how principles of climate justice are realised in practice. It is not to develop solutions, but to see whether solutions already proposed realise requirements of justice. I wish to point out ethical questions that could guide the design of MBIs. Thirdly, following the demonstration of the claims (i), (ii), (iii) and (iv) of the non-ideal theory thesis in previous chapters, this chapter should allow us to better capture how to make progress in the climate justice debate by adopting a bottom-up approach. This chapter is a bottom-up contribution to the climate justice debate. It is bottom-up in the sense of claim (ii), as it looks at one specific practice where considerations of climate justice arise. It aims at balancing questions of climate justice at the level of climate policy. I want to see what form these questions take in the context of climate policy, focussing on one policy in particular. In this sense, this chapter is different from the previous in what the interpretation of the principles in non-ideal circumstances is less what is at stake than the modelling of policy based on considerations of justice.

This particular policy was not chosen randomly. Many prominent scholars working on climate economics and climate ethics issues agree that we should put a price on carbon (Stern, Nordhaus, Stiglitz, Weitzman, Caney, Hepburn, Dessler). And many

market-based schemes are currently being implemented or are already in advanced stages of implementation e.g. the Western Climate Initiative (WCI, between California and the province of Quebec), the Regional Greenhouse Gas Initiative (RGGI in the Northeastern United States) and the European Union Emissions Trading Scheme (EU ETS). Finally, given its many distributive aspects, it seemed like a fruitful ground upon which to make progress in the climate justice debate.

I leave aside here discussions about the social cost of carbon. Economic theories would bring into play here notions such as the social discount rate and the social cost of carbon in order to make the exercise of determining how much we make our emissions reduction effort more quantifiable (Broome 2012, Rezai 2016). I consider that it is sufficiently rigorous for the purposes of this chapter to use the inputs from stabilisation curves and the timetable of emissions reduction to determine the price of carbon.

7.2. Climate change

I pursue here the brief summary offered in the previous chapter. Anthropogenic climate change is the result of the accumulation of greenhouse gases in the atmosphere. This is explained by several factors. The amount of fossil fuels used in industrial production, electricity generation and transportation has increased dramatically since the industrial revolution. The absorptive capacity of our planet has decreased substantially, notably because of deforestation.

This has global consequences, because the pollution released in any part of the planet will have an effect on the atmosphere that is globally shared. It is a problem of global justice since actions in one place have repercussions for others anywhere in the planet. And this has consequences over time. GHG's such as CO_2 stay in the atmosphere for decades. This thus raises questions such as the extent to which future generations have rights against current people.

One key prerequisite to frame the ethics and economics debate is to be clear on what the constraint from climate science is. The evaluation of the severity of the threat

posed by climate change has been associated in the scientific literature with discussions about temperature increase, facilitating understanding of the problem. Nations have agreed in Paris in 2015 to keep climate change under a 2˚C average increase in comparison with preindustrial levels, with an effort to trying to keep it under 1.5˚C (UNFCCC 2015). This reflects the aim of avoiding catastrophic climate change – i.e. droughts, floods, biodiversity reduction, ocean acidification, among other grave dangers – associated with a more than 2˚C increase in temperature. The likelihood of exceeding temperature levels is given in terms of concentration of GHG in the atmosphere, expressed in the form of CO_2e. As mentioned, consensus scientific estimates suggest that at concentrations of 450ppm of CO_2e in the atmosphere, there is a 78% change of exceeding a 2˚C rise. It seems that at the very least we should aim to keep carbon concentrations under 450ppm.[105]

According to the IPCC, the concentration was 391ppm in 2011 (IPCC 2013a). The best estimate of carbon concentration today is that we reached the 400ppm mark in 2015, with an annual increase of more than 2ppm. That means that carbon emissions have to be curbed very soon for the world to be carbon neutral before it reaches the 450ppm mark. This is best expressed in terms of stabilisation trajectories (Stern 2007, see graph figure 7.1 below).

---

[105] Following what we have seen in the last chapter and as I will suggest, I do not use this data to determine the 'carbon budget' that should be distributed between countries.

Figure 7.1. BAU emissions and stabilisation trajectories

The figure below shows illustrative pathways to stabilise greenhouse gas levels between 450 ppm and 550 ppm CO_2e. The blue line shows a business as usual (BAU) trajectory. The size of the mitigation gap is demonstrated for 2050. To stabilise at 450 ppm CO_2e (without overshooting) emissions must be more than 85% below BAU by 2050. Stabilisation at 550 ppm CO_2e would require emissions to be reduced by 60 – 65% below BAU. Table 8.2 gives the reductions relative to 2005 levels.
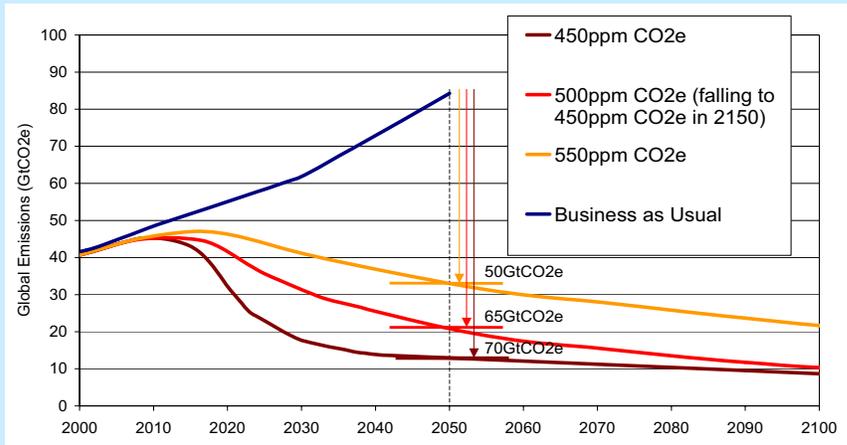
Source: Stern (2007: 205)

The implication of this is that if the stabilisation trajectories of some climate justice proposals do not allow reaching the emissions reduction objective imposed by the scientific constraints, they should be discarded on that basis. However, following Chapter 6, it is possible that proposals today cannot as such reduce emissions to the adequate level (say they would put us on the path not to exceed a 500ppm CO_2e concentration). Yet, given that this proposal is clearly in the right direction, it is possible that it could be revised in, say, 10 years so that a more ambitious target could be envisaged.

7.3 Climate justice: distributing rights and duties

In order to introduce what I believe to be the central constraint from climate justice, it is warranted to give a little context to the methodology of the position

defended in this chapter. To do so, I will begin by reviewing three distinctions put forward by Simon Caney (2012, 2014).

Simon Caney argues that issues of climate justice are too closely connected with wider concerns of global justice to be dealt with in isolation (Caney 2012). To consider these intersections seriously would be to adopt what he calls a method of integration. In Caney's words, to know what a distribution of GHGs entails, one must start with an account of distributive justice that includes a principle of what is owed to all persons, a principle of intergenerational justice and a principle of responsibility for historic injustices (Caney 2012: 291).

For Caney, the method of integration can be understood in a maximalist or a minimalist way (following the distinction I put forward in Chapter 2 and Caney's own account). That is, one might wish to determine what global justice, intergenerational justice and historic responsibility demand in a perfectly egalitarian way in order to determine how emissions should be distributed. Alternatively, one may endorse a minimalist position and hold that global and intergenerational justice require that the basic needs of all current and future generations are met. Caney's argument for integration is compatible with the two views (Caney 2012: 292).[106]

The contrasting view is what he calls the isolationist approach. According to the latter, given the complex web of intersections just mentioned (global justice, intergenerational justice, action under uncertainty), it would be practically impossible to make progress in the climate justice debate if each contribution had to consider every single intersection. According to isolationists, an integrationist approach would place unrealistic demands on a theory of climate justice.

The second distinction concerns the relation between three strategies of climate policy or components of climate action, that is: mitigation, adaptation and compensation. The two first have been briefly mentioned in Chapter 6. To mitigate

---

[106] Note in passing that issues of climate justice intersect with global justice and intergenerational justice, but also with institutional design and action under uncertainty Zellentin, A., 2015. How to do climate justice. *In* Brooks, T. ed. *Current controversies in political philosophy* London: Routledge : 1). These last two issues are not as such distributive questions. As with other questions of political philosophy, it involves discussions such as about individual, corporate and national responsibilities, or the contribution of non-ideal theorising.

climate change means to reduce its scope and magnitude, such as by reducing polluting activities. To adapt to climate change means to reduce its impact on human populations. This can imply for instance providing assistance to climate refugees. And to compensate means to provide assistance for past and present harms, but also to provide assistance to adapt and to mitigate (Caney 2012, Dessler 2012).

A holistic approach is one that considers these different strategies jointly. The atomistic approach maintains that we should pursue these strategies individually. A holistic approach has the intuitive appeal of stressing the importance of the different aspects of our response to climate change, and not only mitigation. For instance, it allows for questions such as, should the duties of one country with regards to mitigation be alleviated if this country did more than its fair share with regards to adaptation? In other words, if a country takes very seriously its duty to host climate refugees, does its duty to reduce its emissions lessen?

I introduce now a third distinction, also taken from the work of Simon Caney. There are approaches to climate justice concerned about what Caney calls *Burden-sharing justice*. Some of the distributive principles advanced in the literature are: that the burdens should be distributed according to the agents' capacity to pay, or that they should be distributed according to their historical responsibility, or still following those who have benefited from the activities that have caused climate change (Caney 2014: 125-126).[107] Indeed, not all countries have emitted equally in the past. Some nations have emitted much more than others. Not all countries have the same capacity to pay for mitigating climate change. And not all countries have benefited equally from these past emissions. *Burden-sharing justice* approaches are contrasted with *Harm Avoidance Justice* approaches. The latter view "takes as its starting point the imperative to prevent climate change, and it works back from this to deduce who should do what" (Caney 2014: 126). This view is compatible with the position adopted in this chapter. Before turning to this position in the next section, I will say a brief word about burden-sharing

---

[107] These two first principles have been explained and interpreted in the last chapter. In this chapter, I will try to model a more concrete application based on these principles. Also, as mentioned, some of these principles can potentially be couple with principles seen in the last chapter e.g. the 'equal per capita' and the 'equal sacrifice' principle.

justice, which I have not mentioned in the last chapter.

We have seen that when looking at climate justice through the lens of distributive justice, one way of formulating what is the goal of distributive climate justice is to ensure the patterns of distribution of a scarce good, including the benefits delivered by the good and the burdens in maintaining the pattern of distribution, are fairly shared by all parties (McKinnon 2015: 377). As seen in the last chapter, one prominent good on which the debate has focussed so far was GHG emissions (Singer 2004, Jamieson 2005, Vanderheiden 2008). This implied a 'right to emit'. This view has considerable problems. I concluded in the last chapter by arguing that the differentiation in the distribution of efforts to mitigate climate change should not apply to 'emissions rights'. The way to be sensitive to historical responsibility for past emissions and to agents' different capacity to pay was not by applying these principles to emissions rights.

Now, with regards to intergenerational justice, it has been pointed out that 'equal per capita' principles of emissions rights are insensitive to future generations. In the case of climate justice, as in a few other cases, the distribution of burdens involves aspects of intergenerational as well as intra-generational justice. Because a portion of CO_2 emitted stays for decades in the atmosphere, and people not yet born can suffer the consequences of CO_2, it is plausible that we do indeed have duties of justice to future generations. Yet, an argument based on 'equal per capita' emission rights could imply that people alive today could use the entirety of the carbon available. It remains silent about what future generations should do. To make this line of reasoning more plausible, we might feel compelled to include future generations. Perhaps we would have to divide the allocation of emissions between a very large number of people in future generations, because these generations matter from a moral standpoint. This would definitively allocate negative emissions to the present generation and urge us out of a carbon-based economy. Any plausible argument based on the right to emit must say that present generations have no right to use all the carbon budget. I will argue given that this notion has little action guidance potential (giving negative emissions to present generation tells us very little about what to do) that the concern for future generations is better taken into account by arguing a 'right to energy', which I develop below.

Also, I wish to clarify that the 'ability to pay' principle is one particular form of the more general principle (mentioned above) coined by Miller as 'the principle of *capacity*': "remedial responsibilities ought to be assigned according to the capacity of each agent to discharge them" (Miller 2001: 460). There could be other ways to interpret the 'principle of capacity' (Caney 2014 puts forward what he calls the Power/responsibility principle). For the purpose of this chapter, I will mostly use the principle of capacity in its general form.

7.4 The case for a bottom-up approach

Before exploring what I call a 'bottom-up approach', I will begin by mentioning one worry with a position that primarily focusses on distributing emissions, i.e. one that adopts a strictly 'burden-sharing justice position' based on the distribution of emissions, and does not pay sufficient attention to a 'harm-avoidance position'. Caney rightly points out the importance of focussing on what actions would most effectively prevent dangerous climate change and what responsibilities follow from that (Caney 2014). Caney argues that we should also consider harm avoidance-based perspectives because they are effective and because burden-sharing perspectives are incomplete.

I should distinguish here between a weaker version and a stronger version of the critique against a burden-sharing position. On the weak reading, as Caney suggests burden-sharing approaches that focus on the right to emit are incomplete. The strategies – e.g. the identification of actions and agents – to avoid catastrophic climate change are beyond the scope of these approaches. On the stronger reading, the principles given by a burden-sharing positions might be incompatible with the distribution effected within a harm-avoidance approach. In other words, an equitable sharing of the burden to mitigate climate change requires a certain distribution of burdens, would not coincide with the distribution of burdens required to achieve the necessary emissions reductions.

In the two cases, the worry is that by focussing on a burden-sharing positions, we run the risk of losing sight of what is really urgent: avoiding catastrophic harm posed by

unchecked climate change. Or, more technically, these positions might not take into consideration stabilisation trajectories that avoid catastrophic climate change, as shown in figure 7.1. What is at stake with these positions is whether, when moving to the non-ideal world, there would still be a chance of succeeding in keeping the CO_2 concentration in the atmosphere below the required threshold to keep the temperature increase from the preindustrial below 2˚C. The planetary boundaries defining a safe operating space are very clear about it and political philosophy like any other discipline has to take this as a hard constraint.

Whether or not the stronger version of the critique is sound, the incompleteness of burden-sharing positions should be sufficient for us to develop another approach which focusses more specifically and directly on avoiding catastrophic climate change. Caney argues, after showing that sacrifices have to be made to mitigate and to adapt to climate change, that we cannot simply assume that agents will spontaneously comply with their responsibilities (Caney 2014: 134). This supports the idea that burden-sharing positions are incomplete (I will return to this point at the end of the next section). Given this incompleteness, we have good reasons to develop an account based on a commitment to avoid catastrophic climate change.

I return now to the principle of capacity. Caney offers an account of second-order responsibilities, which in his definition, are formulated to ensure that agents comply with first-order responsibilities. Caney's account is based on what he labels the 'power/responsibility principle'. It attributes responsibilities to those who can make a significant difference. The important point for Caney in this respect is that we should distinguish between different ways to understand the principle of capacity stated above "remedial responsibilities ought to be assigned according to the capacity of each agent to discharge them" (Miller 2001: 460). The 'power/responsibility principle' and the 'ability to pay principle' are two versions of the capacity principle. For Caney, the important difference between the principles is that the former applies to second-order responsibilities whereas the second applies to first-order responsibilities. Also, he argues that not all the capacity to make a difference is about financial resources. Finally, it is not

grounded on an appeal for equitable burden-sharing, but rather on the commitment to avoid catastrophe. Caney concludes that:

> Given this then: *since* there is a prospect of disastrous effects on people's lives and *since* some agents not only can play an effective role, but their action is critical to avoiding these disastrous impacts; and, finally, *since* these agents lack compelling countervailing reasons for action we are, I think, driven to the conclusion that those agents with the power to discharge *second-order responsibilities* have a duty to do so. (Caney 2014: 146)

I agree with the reasoning that supports his conclusion and with the conclusion. I return to the distinction between first-order and second-order responsibilities in a moment. I will argue that we have strong second-order responsibilities even if we do not have a perfectly defined notion of what first order responsibilities entail.

I clarify now that I adopt what I called in previous chapters a 'bottom-up approach' as it relates to climate justice. The methodology of the bottom-up approach is characterised by the focus on the normative aspects of specific strategy, issue or policy, which in this case is a strategy to address climate change (e.g. one single policy, or one strategy such as adaptation or mitigation). A bottom-up contribution is not hermetic. For instance, it does not exclude that global justice considerations can gain normative relevance in a climate justice debate. It is important to note that the what a bottom-up approach does not map into what has been labelled above as isolationist positions. One reason is that the method of isolation overly focusses on the distribution of 'emissions rights', whereas the bottom-up approach does not argue that distribution of 'emissions rights' is what is at stake. Another reason is that a 'bottom-up approach' admits that there are morally relevant connections between global justice and climate justice, whereas the isolationist approach does not. The bottom-up contribution I put forward has the characteristics of (a) not relying on the distribution of GHG emission rights as a 'focal variable' and (b) structuring the duties 'not to harm' and 'a capacity principle for responsibility' in a cohesive proposal.

Given this is a bottom-up contribution, I will have to bite the bullet that I would not know beforehand what would be required by other principles of global justice. I do not know how this proposal would fare in relation to other rights, but I expect it will have a positive effect on fulfilling rights to subsistence, health and a clean environment.

Moreover, I will clarify how my position aims to link the duty not to harm with compensation to poorer nations. I think this gives a strong link between one specific harm and its associated compensation. Relatedly, although I will not determine the exact portion of historical responsibility for past emissions, MBIs implemented in rich countries and industrialised countries would compensate for part of that.

Besides avoiding the conceptual and practical problems with Burden Sharing approaches based on a 'right to emit', an approach that proceeds from what is required to address the problem has other appeals. Firstly, we get to work out ethical aspects of actual climate policy. Secondly, it allows us to build a theory based on imperatives and recommendations from other disciplines that should constrain our own philosophical proposals. These two advantages suggest that there are chances of making our theory relevant for political decision-making, without losing the conceptual rigour of philosophical work. This is what I wish to develop in the next sections.

It remains very probable that the more fair the distribution of the burdens is, the more it is likely to induce compliance: the more states see others doing their part the more they are likely to do theirs. Although, we might not have a perfect principle of justice to distribute the responsibilities to mitigate climate change, there are ways to make the effort more just. As I will clarify below, sub-global initiatives can be undertaken without a perfectly coordinated global movement. I will look at only one tool of climate policy that takes seriously the need to mitigate climate change and that has the power to contribute to climate justice.

My concern here is less to reject integrationist, isolationist, holistic or atomistic positions than to provide a strong case for bottom-up approaches. I reject however the *primacy* of 'maximalist integrationism', just as I rejected the *primacy* of ideal theory. We do not need to know in advance the conclusions of a maximalist theory that integrates global justice and climate justice. My view only takes the minimalist stance that the basic needs of all persons current and future must be met. I take it that not requiring agreement on a robust position about global justice to be another advantage of bottom-up approaches. The extent to which this minimalist global justice position affects the normative work in a bottom-up approach will be clarified below. Moreover, a bottom-

up contribution allows to show that neither atomist and holistic approaches are right. The atomist approach is wrong in claiming that contributions should focus on mitigation, adaptation and compensation individually. Yet, again, a bottom-up contribution needs not to accept the primacy of a holistic position. Morally relevant problems can be addressed when considering these strategies jointly.

In sum, bottom-up approaches: focus heavily on the actions that have to be undertaken to avoid catastrophic climate change; they integrate robust considerations of climate justice such as the 'duty not to harm', a 'right to energy' and a 'principle of capacity'; they do not require settling on a maximalist integrationist and holistic view to take considerations of justice seriously.

7.5 Normative grounds for pricing carbon

A sensible place to start our practical enquiry of climate justice is to think about how rights and duties relate to carbon pricing. Our first task is to understand what imperatives can be deduced from a moral standpoint. I only mentioned so far the climate science constraint. This section will focus on the climate ethics constraint. There are two key aspects to be explored in this section. Firstly, I will clarify how my position is grounded on the 'right to energy' as opposed to contributions that were based on a 'right to emit'. Secondly, I will show that a bottom-up contribution targeting MBIs allows to structure rights and duties in a morally relevant way: we can justify MBIs based on the 'duty not-to-harm' and use the revenues generated to compensate for the harm done. These compensatory duties can be based on the 'principle of capacity'.

As mentioned, a number of contributions to the climate justice debate are premised on the idea that we should distribute the right to emit GHGs fairly. For Henry Shue, the question of climate justice is how we can achieve the greatest possible emissions reduction without plunging more people into poverty (Shue 2014). I take that the imperative from political philosophy is that this should not plunge more people into poverty and should provide the means to take people out from poverty (whether there is

an economic imperative of Pareto optimality is something I set aside here). In my view, the principle of climate justice is that we should achieve the greatest possible emissions reduction while not preventing people from lifting themselves out of poverty. My view is less minimalistic than Shue's in this respect. People have basic rights, among which we find a right to subsist, which cannot be meet without development.

Yet, and this is central, the right to subsist can be met by way of low-carbon development (Shue 1994, Shue 1995). Regarding rights and the benefits of the distribution, Shue rightly pointed out that what matters is not really that people have a right to emit, but rather that people have their energy needs met. He argues that instead of distributing emissions, we should make sure developing populations have access to clean energy (Shue 2014). That has an implication of tremendous consequences and allows to refine the conclusion of Chapter 6. The fact that we need to pollute to meet these needs is contingent. People's energy needs could be met by non-carbon based energy production.

This conclusion must be taken seriously: in order to move forward in the climate justice debate, it is better to argue for a right to energy, not emissions. Emissions are polluting. Energy needs not to be. Instead of giving people the right to pollute as a matter of fairness, we will assist them to develop in a less polluting way, as a matter of ecological consciousness and distributive fairness. The solution is thus morally justified and has practical appeal: to reduce poverty by allowing for access to clean energy.[108]

This movement is crucial. If we combine the imperatives from climate justice and climate science, it is urgent that (a) we secure individual rights to subsistence with particular attention to the poor (b) by achieving a low or zero carbon development society. The 'distribuendum' or the 'currency' of climate justice should thus not be emission rights but rather 'energy rights' in the sense of ensuring that individual rights to energy are met through low carbon development (Shue 2014).

Therefore, I will ground my account of MBIs on a right to energy, and not on a right to pollute. Today, clean energy must be subsidised for it to become an even more

---

[108] This corroborates with the conclusion of the previous chapter.

viable alternative. As we will see, carbon pricing is a tool to do so. The design of MBIs must be so as to respect individual rights to energy, to not prevent people from having these rights fulfilled and to help fulfilling these rights.

I will now discuss the duties aspect of the question. We have distinguished so far between harm avoidance duties and duties based on the capacity principle (Caney 2014). Henri Shue defends a harm avoidance perspective (Shue 2014: 156, 265). In one sense, this means avoiding actions that result in suffering. In another sense, it means compensating for harms previously caused, which could be interpreted as endorsing a 'polluters-pay principle'. I wish to link this reasoning to my account of MBIs, which have two main features. We should see that MBIs are based on the duty not to harm in the first place, for they are sensitive to the idea of paying for harms caused. Yet, they open the door to the duty based in the ability to pay (one version of the principle of capacity) in a second step. In fact, in the case of MBIs, the polluters pay principle on which these instruments are based generate revenues that could be used according to the 'capacity principle', which we could interpret for the sake of simplicity as the 'ability to pay principle'. In sum, MBIs structure the two duties – the duty not to harm and the capacity principle – in a cohesive proposal. In other words, besides being grounded on the duty not to harm, there is a second step in the design of the policy that concerns distributing the revenues generated, such as to make it sensitive to considerations of justice. This could be done based on the principle of capacity. At this stage, developed countries should help poorer populations having their rights to energy met without compromising the global emissions reductions effort.

My account thus proceeds from the idea that MBIs are a way to put into effect the duty not to harm, in the sense defined by Henri Shue, which I mentioned above. Market-based instruments make agents pay for the harm done – and do it by including *all agents* involved in polluting activities. That means that emissions reductions will be achieved by minimizing the number of polluters which do not contribute to the emissions reduction effort (whether this will be translated strictly in terms of carbon

prices or of more general reduction strategies has to be shown). The next section will focus more specifically on the functioning of MBIs as a negative incentive to pollute.

I return now to the distinction between first-order and second-order responsibilities. As mentioned, first-order duties are responsibilities for agents to perform certain actions. Second-order responsibilities are those that some agents have in order to ensure that agents comply with their first-order responsibilities. Regarding the 'ability to pay principle', Caney associates it with first-order duties and this is how it is understood in the climate justice literature.

In one sense, one may argue that in the case of carbon pricing mechanisms, we are not at the level of first-order or second-order duties, in Caney's terminology. That is because we are applying our moral reasoning directly to one specific policy. However, it seems that in the case of MBIs, the harm-avoidance basis of the 'polluters-pay principle' is clearly in line with second-order duties. They are ways to help discharge first-order moral duties about climate change mitigation. Moreover, this is a second-order responsibility that provides means to fulfil a first-order responsibility more directly. The justification of the use of the funds generated by MBIs as a way to compensate for harms done could be coupled with the 'ability to pay principle'. This is a way to discharge a portion of a first-order duty. That is, even if in my bottom-up approach we do not need to determine exactly what the first duties of each agent are, we generate ways to fulfil these duties. MBIs are not only a way to reduce the harm done, but also a way to fulfil individual rights to energy.

Also, I argue that MBIs should be modelled such as to respect the climate justice constraint, which is to achieve the greatest possible emissions reduction while not preventing people from lifting themselves out of poverty and, more positively, by helping them fulfilling their right to energy. What is central in the context of MBIs is that the mechanism through which we provide incentives for agents to avoid causing harm is the same mechanism that generates the means to make the climate mitigation effort more fair (and the last sections of this chapter envisage some ways through which this can be accomplished).

In this sense, when thinking about justice in practice in this context, market-based instruments for climate change mitigation allow us to link a mitigation strategy with another strategy of climate justice: compensation. Compensation normally connects to the question about how we can link duties with the outcomes of specific actions (Zellentin 2015: 8). This implies showing that there is wrongful loss and establishing who is responsible for counteracting it.

In other words, all moral duties that can be connected to the design of MBIs are not only 'allocative' in nature. There is another level of debate beyond allocation. There are also corrective duties (I assume allocation and correction fit under the label of distributive duties). Corrective duties ensure that wrongs are repaired (McKinnon 2011, McKinnon 2015: 377). They connect those causing wrongful harm with the people they harm such as to generate compensation claims by the latter (Adam 2011; whether we need to factor uncertainty in here is a question I leave aside).[109]

Corrective duties aim at repairing the wrongs done. Normally, any party suffering wrongful harm has a rectification claim that has to be met by the one harming. Usually, this is done through compensation, which regularly takes place in the context of climate change (Hunter 2007). And corrective compensatory justice can also have an intergenerational scope. Normally, there is no liability before the causation of harm. But in the case of climate change, the liability under corrective justice should also respond to the imposition of impermissible risk (McKinnon 2011, McKinnon 2015). In this case, an agent imposing risk should be ipso facto liable for providing the means for compensation.

There is widespread scientific agreement that CO_2 emissions create impermissible risks for future people and some present people in at risk areas. These risks will likely mature into harms (McKinnon 2011, McKinnon 2015). Therefore, present generations have a compensatory duty to future people. This is the intergenerational element. The intra-generational compensation comes from the fact that a number of developed and newly industrialized countries are polluting far more than developing countries. Yet,

---

[109] This connects with the notion that non-ideal theory involves reparative, comparative and transitional considerations.

future people in these developing countries will also suffer (in many cases even more) from the effects of climate change. This is not only about the harm that we will cause but also that we might even prevent future people from acting ethically for they will have no means to think about their future generations.

The kind of compensatory duties that can be associated with MBIs, that aim at helping people to act ethically, can be linked to what Caney calls a second-order responsibility of 'enablement' (Caney 2014: 137). Caney rightly points out the importance of focusing on the responsibilities that target specifically how to solve the problem. If MBIs contribute to helping people discharging duties of justice and acting morally, they would be fulfilling this second-order duty. The different ways that compensatory duties can be interpreted as responsibilities of enablement will be seen in the last sections of this chapter. For another example of a second-order responsibility of enablement, Maltais argues that states have the responsibility to create effective international cooperative frameworks, and even that it would be justified that some countries went ahead of others in doing so even if they would except weak initial reciprocity (Maltais 2014).

In order words, this normative assessment of MBIs assigns a duty not to harm *ex ante* – i.e. before actions (and targets on emissions reductions) are agreed upon – and allocative and corrective duties *ex post* – once what has to be done to achieve emissions reduction has been determined. The corrective duty is in principle based on the notion of compensation for the harm done. The allocative duty is the one associated with fulfilling the right to energy based on a principle of capacity. The duty not to harm *ex ante* provides strong reasons that justify the implementation of carbon pricing mechanisms and these two duties *ex post* provide strong moral grounds for a design of MBIs.

This allows to support the argument of the previous section. One feature of this strategy is to avoid the potential problems of incompleteness or of incompatibility of burden-sharing approaches. By arguing for compensatory distribution *ex post* we contribute to a design of distributive duties to mitigate climate change that does not prevent the emissions reduction effort from reaching its target. For this climate policy in

particular, there is a way to balance between moral principles that relaxes the tension between achieving emissions reduction objectives and distributing the effort to do so fairly.

Let me conclude this reasoning by pointing to one caveat about the weighting of principles of climate justice, which support adopting a bottom-up approach. Consider, in the context of the allocation of duties, seeking balance between the 'polluters pay principle' and the 'ability to pay principle'. Firstly, we know that in the polluters pay principle, those who currently contribute the most to the problem bear the greatest burdens in addressing it (they will pay more). Moreover, I justified that the second step of the design of MBIs concerned with distributing revenues could be based on the allocative principle of capacity or on the notion of reparative duties, or on a combination of the two. In the context of the institutional design of MBIs, these principles were compatible.

We might wish to include in this discussion other principles such as the 'historical responsibility principle' or the 'beneficiary pays principle', mentioned in section 7.3. We might wish to consider whether the current cohort in the anthropocene is responsible for the harm done by past generations (Rosen *et al.* 1999). I mentioned that a way to solve this problem is to consider emissions from 1990, date of the publication of the first IPCC assessment report. This problem could potentially also be solved by adopting a 'beneficiary pays' perspective, where although we do not need to claim that present people in affluent countries are responsible for past harms in the climate context, we can argue that they are benefiting from it. This could be a satisfactory answer to factor in historical emissions. However, if we interpret the ability to pay principle so as to claim that, regardless of past responsibility, people who have benefited from past pollution should be required to contribute more to addressing the problem, we have to determine the causal relation between pollution and actual wellbeing by distinguishing it from other causes (such as labour). This is perhaps a problem about over-determination in backwards looking perspectives. Yet, just because finding the ultimate principle that balances between these considerations in abstraction is not easy does not mean it should not be tried.

I believe that by using the principle of capacity to regulate the compensatory component of MBIs we avoid the problem of over-determination. This distributive component allows to be sensitive to considerations of justice even without determining the full extent of first-duties of justice. If this practical solution allows us to distribute rights and duties in a fair but not perfectly just way, and at the same time is more likely to put us on a path to achieve meaningful emissions reductions, it provides sufficient grounds for this to be pursued in a bottom-up way. The exact measure of historical responsibility and the exact measure of the principle of capacity will not be determined in abstraction, but will nonetheless be relevant for the practical reasoning concerned with the distribution of responsibilities around international climate roundtables.

7.6 The economics of pricing carbon

With some of the normative underpinnings for MBIs in place, I now present an overview of carbon pricing instruments from an environmental economics standpoint. The core justification for pricing carbon rests on the economic analysis of 'negative externalities': the situations where the effects of the production and consumption of goods and services impose costs on others, which are not reflected on the price charged for these goods and services (Bowen 2011). Because the costs of emission are not imposed on the emitter, the polluter has no incentive to make any effort to reduce it. When a price reflects the cost of emitting pollution, i.e. when it internalizes the negative externality, it directs investment and consumption away from polluting activities. Moreover, carbon-pricing mechanisms allow us to cut emissions efficiently, precisely because they allow agents to cut the emissions where it is less expensive for them to do so (Stern 2007, Dessler 2012, Aldy and Pizer 2015).

As mentioned, market-based approaches feature importantly around climate roundtables and are increasingly an important aspect of mitigation policies. These approaches are in place under initiatives such as the WCI, between California and the province of Quebec, the RGGI in the North-eastern United States and the EU ETS. The

two principal market-based policy instruments – the carbon tax and the cap-and-trade system (Stiglitz 2006, Weitzman 2013) – have two features in common: efficiency (economic, achieving emissions reduction at the lesser cost, and administrative companies know what emissions to cut) and distributive implications (they generate revenues) (Olmstead and Stavins 2010, Bowen 2011). Both approaches involve putting a price on carbon. For the purposes of discussion, I will focus more on cap-and-trade mechanisms than on carbon taxes. There have been many cap-and-trade programmes being implemented or recently implemented, and they are large in scope. It is important to begin by reviewing the factors explaining why MBIs are more efficient than regular control and command approaches (CCA).

MBIs are advertised as more administratively and economically efficient than CCA. They are more administratively efficient because governments do not have to develop control standards for each facility: "The facilities themselves, with their internal knowledge of industry operations, make the critical decisions about whether and how to reduce emissions." (Kaswan 2014: 237). I use the term administratively efficiency here to refer to efficiency in the decision-making procedures to determine the way to achieve the emissions reductions.

Secondly, they are also touted as more economically efficient, in the sense of cost effectiveness, than CCA. That is because they reduce aggregate emissions at the lowest industry cost taken globally (Dessler 2012). Indeed, CCA require all facilities to do the necessary adjustments to reduce emissions to the same amount, even if some plants could reduce at a lower cost than others. Cap-and-trade programs allow plants that could reduce emissions at a lower cost to sell permits to those with more expensive marginal reductions costs (see table 7.1 below). For the firms, MBIs reduce the costs to comply with emissions reduction targets. For society, that means that fewer resources are devoted to achieving climate objectives.[110]

---

[110] I present here MBIs as they would ideally function. I assume for the purposes of this paper that these instruments work from an administrative and economic standpoint, as mentioned above. I will only present potential problems with these instruments as they relate to justice considerations.

Contrasting with the question of climate justice, the question of environmental economics is how can we achieve the greatest possible emissions reduction at the least possible cost. MBIs are an instrument to achieve this objective, precisely because they cut the cheapest emissions. Normally, MBIs will charge a price for a ton of carbon, which normally increases every year or every so often. Fewer permits are emitted every year, so that the reduction in GHG emissions is progressively increased.

An initial cap on emissions is established and the implementing agency will distribute permits summing to the cap to companies. Each company can choose "to reduce emissions to meet its allowance allocation, or to reduce emissions by more than the allocation and sell the remainder, or maintain existing emissions and buy allowances to make up the difference between the number distributed and actual emissions" (Kaswan 2014: 237). All companies must demonstrate annually that they had enough permits to cover emissions.

The logic behind these instruments is that companies will cut emissions until the marginal cost of cutting another ton of carbon equals the price of the permit. In principle, MBIs respect the environmental economics constraint. Consider figure 7.2. Table 7.2 models a scenario with two power plants. Imagine the price for carbon permits is £4 for a one-ton permit. Plant A emits 10 tons of CO_2 into the atmosphere. Column 3 indicates the marginal cost to cut one ton of carbon. The first ton of carbon will cost Plant A £1. The second ton will cost £2, the third £3, the fourth £4. That is, for the first three tons, the price for not emitting them is less than the price of the permit. Plant A will thus cut these three tons, and probably the fourth one, it they can cut it for the same price as buying the permit. Consider now Plant B and its associated marginal cost for emission reductions (say it is an older plant). The marginal cost to cutting emissions is given by column 5. Plant B also emits 10 tons of CO_2 each year. Together, the plants emit 20 tons of CO_2. Now, say that the government wants to reduce emissions by 6 tons of carbon.

Figure 7.2. Cost of reducing emissions for Plants A and B

Emissions reduced	Tons of carbon emitted	Plant A Cost in £		Plant B Cost in £	
		Marginal	Total	Marginal	Total
0	10	-	-	-	-
1	9	1	1	2	2
2	8	2	3	4	6
3	7	3	6	6	12
4	6	4	10	8	20
5	5	5	15	10	30
6	4	6	21	12	42
7	3	7	28	14	56
8	2	8	36	16	72
9	1	9	45	18	90

Source, Dessler (2012).

In a CCA approach, if we have two plants, it means each plant would have to cut emissions by 3 tons. That means: Plant A will cut 3, which will cost £6 in total. Plant B will also cut 3 tons, which will cost £12 in total. The total cost of cutting 6 tons of carbon in a CCA approach in this case is £18.

In a carbon market, the government will only issue permits for 14 tons of CO_2 that year. One permit costs £4. Plants A and B will buy (or will be given) 7 permits each (I leave aside here the original auction of permits which could have the same result). The price of the permit is £4 and the Plants have to cut emissions by 6 tons together. They each have to buy permits for their emissions. Company A can cut four tons for less than the price of the permits. It only needs 6 permits of one ton each. It will cost Plant A £10 to do so and it will have one permit to sell. Company B can only cut 2 tons for less than £4. It will cost Plant B £6 to cut its two tons. But it will need 8 permits to do that. Company A will have the incentive to sell one of its 7 permits to company B. That is, under this model, the same 6 tons of CO_2 will not be emitted. But the total cost for achieving this emissions reduction is £16. Under this model, achieving the same emissions reduction will cost £2 less than in the CCA model.

It is in this sense that MBIs are economically efficient. This is a way to cut the cheapest emissions for a lesser aggregated cost. It does not exclude the possibility that more expensive emissions reduction can be achieved by another policy. But for this

portion of the emissions reductions effort, we should expect to observe an economically efficient way of cutting emissions.

This allows to put aside one concern that has been raised from a philosophical perspective in the literature. It has often been argued that carbon permits are just another way to tell the rich that, as long as they pay, they can keep polluting. This argument is flawed for various reasons (Caney and Hepburn 2011 provide a thorough discussion on this point). Briefly, firstly, putting a price on carbon directs production and consumption away from carbon-intensive goods. Secondly, in a cap and trade system, there is a cap on emissions. Fewer permits are emitted each year. GHG emissions are set to diminish every year.

7.7 Internal Problems of Justice with Market-Based Instruments

The next sections of this chapter address some important trade-offs between fairness and emissions reduction at the global level. I will begin by examining problems of justice which might be created by the very implementation of MBIs. These must be addressed for MBIs not to become a cause of injustice. Carbon-pricing mechanisms raise 'internal' justice-based issues and, in this context, trade-offs between efficiency and fairness arise. This section addresses two trade-offs that are particular to the implementation of the policy. I will raise a third problem at this end of this section.

Firstly, from a domestic standpoint, there is one important tension between pursuing mitigation efficiency and fairly distributing the burdens to mitigate climate change. Polluting facilities are often located closer to poorer populations. If emissions are traded freely, it is possible that the plants located closer to poorer communities will reduce their emissions less than other plants. In other words, there are no guarantees that poor communities will not continue to bear a high cost for living in polluted areas.

A similar problem arises if emissions offsetting is allowed in the implementation of a given market-based mechanism. If affluent countries purchase international offsets permits like offsets from planting trees in other countries (and if we associate emissions

offsetting with market-based mechanisms), there are no guarantees that polluted areas will experience less pollution in the short to mid-term. The poorer communities in these countries could continue to suffer from pollution-associated problems (Kaswan 2014: 244). The design of MBIs must be sensitive to this trade-off.

A second important tension between MBIs and distributive goals concerns administrative efficiency. There is a trade-off between administrative efficiency of MBIs and participatory democracy. That is because MBIs are designed to maximize private autonomy and administrative efficiency. Public involvement in auctions and trading is minimized.

> In a cap-and-trade program, government entities would set the emissions cap, but they would not design a system of industry-specific requirements through a public rule-making process. At the individual facility level, the public would continue to have a role in initial siting decisions. But since there is no opportunity for public participation in private allowance trading decisions, the public would not have any input into subsequent changes in GHG emissions unless those changes were substantial enough to trigger co-pollutant regulatory proceedings (Kaswan 2014: 244)

Cap-and-trade programmes are administratively efficient, but come at the cost of governmental and public involvement.

In response to this second point, I should emphasise that MBIs should be seen as a complement, as a constituent part, of a more overarching climate initiative. They normally complement a regulatory system instead of replacing it (Kaswan 2014: 246). This opens the door for more political participation, notably by setting related or non-related regulatory standards. One example of a related regulatory standard that could be required on the grounds of political participation is the demand that facilities themselves install continuous emissions monitoring systems that facilitate government overview and enforcement of the MBI programmes. This was observed in the Acid Rain Program, a trading programme implemented in the United States from 1995 to reduce atmospheric levels of sulphur dioxide and nitrogen oxides, which has proven to be administratively efficient (*Ibid.*). This kind of regulatory standard could avoid a time-consuming administrative process of monitoring. Another regulatory standard that

could be demanded by public participation is an information campaign to smaller companies that are unaware of technological alternatives available to them.

I do not want to suggest that the public should have a say in auctions and trading, except perhaps in cases where emissions reductions were not observed in very polluted and poor areas. I want to raise the idea that controlling bodies have to press companies and industries to find mechanisms (about the number of permits, quantity of emissions, publishing of results) to facilitate monitoring. This is meant to allow easy access to the public about the progress of emissions reduction and the circumstances of the trade. This one potential way to argue for more public intervention. Also, I think that appropriate monitoring could facilitate combining emissions trading with other strategies.

In response to the first point, I argue that there are ways to design MBIs such as to have an appropriate balance between efficiency and fairness. In order to reduce the fairness problem of having more polluting facilities in poorer areas, within and across nations, we could envisage distributing fewer allowances for facilities in already polluted areas than in less polluted areas. "If allowances were auctioned, facilities in polluted areas could be allowed to purchase only a certain percentage of their baseline emissions." (Kaswan 2014: 249). Alternatively, instead of making fewer permits available, a higher price could be asked for allowances in these areas. Although there will be no guarantees of reaching the objective, this could create a negative incentive to reduce polluting emissions in already polluted areas. These are questions of design, that might affect efficiency negatively, but this negative effect could be outweighed by the benefits of making very polluted areas less polluted for the residents.

One may note that another positive effect of reducing the costs of mitigation is that there will be more resources to help poorer populations cope with the costs of climate objectives and that poorer populations will benefit the most from rapid action.

> "As the Intergovernmental Panel on Climate Change (IPCC) has noted, residents of developing countries are likely to be more adversely impacted by climate change's consequences than those in the industrialized world. […] if economically efficient policies lead to higher reduction goals, they could mitigate the climate change impacts on the globe's most vulnerable regions and communities." (Kaswan 2014: 239-40).

Yet, as the discussion about the development challenge and mitigation efficiency in Chapter 6 has showed, the different trade-offs involved in this kind of reasoning about justice in practice are complex. The rest of this chapter is also dedicated to addressing it.

Now, there is a third problem, which I also consider to be 'internal' to the implementation of MBIs. It is worth noting that the increase in the costs of carbon emissions is likely to be reflected all the way on the consumer chain. It is possible thus that it has a regressive effect, having a disproportionately greater impact on poor populations (Caney and Hepburn 2011). One may thus argue that the revenues generated by MBIs should be used to compensate poorer populations for the price increase of carbon-intensive goods. This trade-off is similar to the trade-offs that will be addressed in the next section.

7.8 Designing MBIs to become instruments of justice

I argue in this section that efficiency in emissions reduction and fairness can be pursued jointly. I thus follow on the conclusion of Chapter 6, as it applies to the design of one specific climate policy. The following sections aim to go beyond what was accomplished in the previous chapter, however, in order not only to show how non-ideal theory works, but also to offer an account of a bottom-up theory of climate justice as it relates to carbon pricing.

There are at least two trade-offs between fairness and efficiency in this debate; one is particular to the domestic context and the other arises at the global level. Domestically, this trade-off takes the following form: we have a choice between allocating the revenues generated by market-based mechanisms to incentivise the development of existing green technologies in order to make them more competitive and increase the positive impact of market-based instruments, or allocating them as a compensatory measure to lower income families who will suffer most the increase in price of carbon-intensive goods (Nordhaus 2009, Aldy and Pizer 2014, Aldy and Pizer 2015). Globally, the trade-off has the following form: we have a choice between using the

revenues generated by MBIs to compensate poor populations (say, that are less historically responsible for polluting emissions or less capable to reduce their emissions) for past harms in non-climate related projects, or investing in research and dissemination of sustainable alternatives (e.g. energy, agriculture, transport) that maximise emissions reduction. This last trade-off is analogous to the 'development challenge' addressed in the last chapter.

This section will first review different alternatives that can be pursued with the revenues generated by MBIs. It will propose a framework to assess these options based on its emissions reduction potential, its political traction, its economic return and a moral assessment. Every single judgment in this analytical framework requires empirical support beyond the scope of this chapter. This is only an overview of different options based on the argument developed in the last two chapters.

This section is premised on the idea that the funds generated by MBIs are of moral significance for various reasons, following the reasoning of sections 7.4 and 7.5. These funds are generated by a price put on polluting activities. As argued above, the justification for putting a price on carbon is based on the duty not to harm. The revenues generated could be used to compensate for the harm done. This implies that it would be justified to invest these funds in the alternatives that are part of our response to climate change, which would help people fulfil their right to energy. Also, the principle regulating the distribution of burdens to contribute to this effort could be based on the principle of capacity. Moreover, as pointed out by economists, MBIs affect powerful businesses, which have considerable political influence, and which will feel the consequences of putting a price on carbon. The following options I review are not investments which are exclusively funded by MBI revenues. Nonetheless, given the particular moral significance of MBIs and the relation they maintain with these alternatives, the revenues generated should be attached to one of them or a combination of them.

One first proposal mentioned in the conclusion of Chapter 6 is that part of the revenues of MBIs implemented in developed countries should be directed to the Green

Climate Fund (GCF). This should, say, be automatically written into the design of MBIs so as to respect global climate justice constraints. The GCF aims at making funds available to less developed countries for them to pursue their own emissions reduction initiatives. The COP21 has settled that developed countries should contribute at US$100 billion annually to the GCF. These funds will be used notably to subsidise sustainable energy sources, agricultural practices and transport to poorer populations.

This option might have less of an economic impact on poorer population in the short run than a direct compensation for price increases. However, the effects in emissions reduction would be greatly augmented, for the pricing of carbon would at the same time favour the development of clean energy and other sustainable alternatives. Moreover, given that it is in developing countries that we find the most cost-effective ways of reducing emissions (Olmstead and Stavins 2010: 7), these initiatives will be better funded. These initiatives aim at providing green jobs training, sustainable development practices and financing energy efficiency improvements in less advantaged communities. Linking the revenues of MBIs with the GCF seems to be one option which respects the constraints from climate ethics, climate economics and climate science. It would be helpful now to distinguish more clearly between an assessment of this proposal from the climate perspective (in terms of emissions reduction), its political traction, its economic return and its moral assessment.

Climate assessment: investing in the GCF should help developing countries commit to climate objectives and get on track with emissions reduction. It is plausible that the results in terms of emissions reduction with this proposal will be positive. This option is 'mitigation efficient' in the sense it diminishes the overall sacrifice to mitigate climate change.

Political traction: investing in the GCF should have political traction given that developing countries would be likely to increase their climate efforts. This is a powerful argument in the context of climate negotiations. One could argue that, historically, providing assistance to developing countries is not a particularly popular idea in developed countries. This is partly because it is thought that the money is used ineffectively. An effective use of revenues by GCF would certainly contribute to a

positive public reception of this kind of international transfer.

Economic return: more studies are required to know exactly what kind of economic return can be excepted from investing in the GCF. One possible positive element in this respect is that it is normally much less costly for countries to develop green economies from the start than to develop a polluting economy and then making green investments to change the form of development, as pointed out in Chapter 6. Consider the case of mobile phone technology in Africa. Africa has developed mass telecommunications without the expense of introducing outdated land-line technology.

Moral assessment: this is perhaps the most straightforward of judgments for this option. Contributing to the GCF allows us to link the polluters pay principle with compensation for the harm done. We would link the polluters to people suffering from the effects of this pollution. Considering that poorer nations will be among those who will suffer the most from the effects of climate change, providing them with the means to mitigate and adapt to climate change is morally justified. The compensation for the harm done will be effected directly by helping people fulfil their right to energy. This example illustrates one way to think about climate justice in practice.

A second policy option is to invest in developing and diffusing sustainable alternatives (e.g. energy, agriculture, transport) and in local climate initiatives, which are not done through the GCF and do not necessarily target developing countries. The motivation to invest in technological development and diffusion of sustainable alternatives is to make them cheaper in the long run and to make sure more resources are put at the disposal of fighting climate change. What is paramount in this case is to disentangle the different costs and benefits associated with this option. Recent studies indicate that the benefits of investing and diffusing green technologies have been underestimated (Green and Stern 2016). Insights from these studies would allow us to determine to what extent there is a trade-off between economic gains and emissions reduction.

Climate assessment: of the four options reviewed here, this is one which, jointly with the previous alternative, has some of the greatest benefits from the point of view of

emissions reduction. The funds will directly serve the case investing in options that aim at reducing emissions, but not only in developing countries. This option is also 'mitigation efficient' in the sense it will make sustainable energy and practices more affordable in the long run.

Political traction: investing in green technologies like wind and solar has been depicted for a long time by politicians as something that is contrary to economic growth.[111] The political traction of green investments is gaining more and more sympathy from public opinion in a large number of countries and communities, although perhaps not a majority in places that matter, and in political spheres. This could shift from a solid account of the economic return and especially a political understanding and dissemination of the results.

Economic return: I will return to this below, but it is paramount to take into consideration the fact that investments in green technologies drive down their cost, making the cost associated with choosing green alternatives considerably lower in the long run. There are many green alternatives today that already have important economic returns or are cost neutral (Stern 2007). There are also great savings associated with spending less on fossil fuels.

Moral assessment: besides the positive moral impact of reducing the impact of climate change, recent studies show that the co-benefits of this kind of investment are significant, in terms of health and employment. Especially, avoiding the health related costs of pollution should be something to bear in mind (Green 2015).

A third option is to compensate poorer populations for the price increase of carbon-intensive goods. This third avenue raises the important trade-off mentioned above. Poorer populations will suffer from the price increase of carbon-intensive goods all over the planet where MBIs are implemented (taking this factor separately from other factors that could also raise the price of these goods). They could be compensated by the revenues generated through MBI programmes. The central problem with compensating

[111] For instance, Stephen Harper, former prime minister of Canada (2006-2015), has in a number of occasions and debates declared that fighting climate change is detrimental to economic growth.

lower income families is the opportunity costs incurred: investing in developing and diffusing green technologies have an important impact in diminishing even further polluting emissions and have profound health and economic co-benefits (Kaswan 2014, Green 2015). Moreover, this would not necessarily create the desired disincentive to purchase carbon-intensive goods for a portion of the population. By compensating people for the price increase, it is unclear whether people will turn away from these goods. The mitigation efficiency of the measure would be diminished at the benefit of more fairness (Aldy and Stavins 2011).

Climate assessment: given the opportunity cost and the potential mitigating effect on the desired disincentive, this is perhaps not the option that will maximize reductions in emissions.

Political traction: returning the money to part of the population is likely to be a popular measure. In British Columbia, Canada, the implementation of a carbon tax was done at neutral cost. All the money generated returned to the population. This is perhaps a way to draw political support to the measure.

Economic return: there is no particular economic benefit (job creation, investments) associated with this option.

Moral assessment: this targets directly the poorer populations that will be the most affected by the implementation of these measures. This option could help to make sure that the 'right to energy' of poorer populations is not compromised.

A fourth option is one put forward notably by John Broome. Broome argues that one avenue is to compensate emitters in order to ensure they will not become the big losers of carbon pricing schemes. Compensating emitters would in this sense lead to a Pareto improvement scenario in the sense that no one is made worse off by introducing this new policy (Broome 2012). From my knowledge, it is not clear that the empirical evidence demonstrates that there are more chances of achieving the desired emissions reduction by compensating the emitters instead of using the money to make developing countries more likely to reduce their emissions or to make alternative energy cheaper.

Climate assessment: there are no particular climate benefits in choosing this

option. In fact, it is not impossible that there will be negative climate effects in comparison to any other option, for this will not undermine the power of companies in polluting sectors of activity.

Political traction: presumably, this would be the biggest advantage of this option. The polluters, that have so much influence in political circles, would not be against this measure, for their losses would be compensated. However, in Australia, the idea that big polluters were being compensated was not well received (Cullen 2011).

Economic return: all things being equal, this option will not undermine dramatically the economic capacity of polluting companies and therefore employment in these sectors will not decrease very much. In the energy sector, given that many polluting companies are not limited to fossil fuel production and operate in various fields in the energy sector, this option will potentially create the necessary incentives for these companies to shift part of their production and workforce to less polluting fields.

Moral assessment: this option is not particularly positive from a political morality point of view. Although it would be a Pareto improvement to make polluters better off and not making anyone worse off, this measure would compensate those the most responsible for harming others in the first place.

7.9 Reflective Integration and Action Guidance

I will review in this last section a few practical elements from climate economic studies which should inform how we conceive justice in practice. The last section showed how much empirical evidence is required to make all things considered judgments. I am suggesting that the philosopher can contribute considerably to political debates by balancing between policy avenues and helping determine the best course of action to be undertaken.

Firstly, one may rightly ask if ideal theory was of any help in distinguishing between the viable and desirable options in non-ideal theory. As mentioned in the sections on reflection integration in Chapter 2 of this thesis, political theories will use, at

least minimally, theoretical elements proper to ideal level theorising. In this bottom-up contribution, elements such as 'the duty not-to-harm', the 'capacity principle' and the 'right to energy' contributed centrally to formulation of guidelines for the design of market-based instruments. Reflective integration allowed us to select these elements over ideas such as 'the right to emit', thereby offering a moral justification to our institutional design with greater action-guiding potential. This chapter focused more on an application of principles, and less on an interpretation of principles, as it was the case with the previous chapter.

Secondly, the just course of action to be undertaken in the climate justice context has to be informed by how much more emissions reduction can be achieved by investing in sustainable alternatives (for instance energy, agriculture and transport) in comparison to other options, by how much more needs to be invested in the GCF for developing countries to be capable of fully cooperating with the emissions reduction effort and by what options really help the plight of the most vulnerable populations of the planet.

I will now follow on the analysis of Chapter 6 in order to assess whether investment in development and dissemination of green alternatives is perhaps what allows relaxing the trade-off between maximising emissions reduction and fairness. The idea is that the investing in options that I called 'mitigation efficient' – that is, options that aim at maximizing reduction in GHG emissions and expand the feasibility curves for more mitigation in the future – will relax the tension between mitigation efficiency and fairness. These options will (a) reduce the risk of catastrophic climate change for everyone and especially vulnerable populations, (b) have important co-benefits, especially in terms of health and economic opportunities, and (c) they are sufficiently sensitive to fulfilling the right to energy of poorer populations.

If well designed, MBIs could be efficient (administratively and economically), they could be fair (by being more demanding with facilities in already polluted areas and subsidizing alternative energies for poorer communities) and they could achieve important emissions reduction (by the selling of permits that decrease in number every year and by making more resources available to invest in green energy and other technologies). MBIs generate revenues that could be used to curb some climate

injustices. And although countries would have self-motivated interests to invest in emissions reductions, some countries should benefit from international distributive justice measures, based on a 'principle of capacity'.

Given that the objective of the GCF is to target developing populations specifically to fulfil their right to energy, there is an overlap between options 1 and 2 in the previous section in the joint pursuit of fairness and emissions reduction. Note that this kind of overlap is welcome given that contributing to the GCF is not sufficient to reduce the risk of catastrophic climate change, not only because it does not pretend to do this but also because it is unlikely that any single policy could do it.

Now, there are other key findings in empirical research that inform the climate justice debate. For instance, we have indications that, generally, the national benefits of decarbonizing the economy outweigh the costs. Nations have incentives to reduce GHG emissions: the assumption that nations have incentives to free-ride is for the most part mistaken (Green 2015). There is no global prisoners' dilemma in this case as opposed to the case of tax competition, in the sense that it is not in nations' advantage in the long run to ignore emissions reduction targets. There is a collective action problem, for no agent alone could achieve the common objective. But there is no prisoners' dilemma, in the sense that agents' incentives in cooperating are dependent on the action of other agents. In the context of advocating economic investments in low carbon initiatives, there should be more effort made by governments to identify what economic sectors present net-beneficial gains and cost-free gains, what sectors may be net-costly, and what should be done in terms of international cooperation to tackle the latter (Green and Stern 2016).

We must keep in mind that MBIs are normally implemented nationally or regionally (it is unlikely that a global emissions trading scheme will be implemented in the near future). Domestically, poorer populations will be affected by the price increase of carbon-intensive goods. Their right to energy will be compromised if the price increase of carbon-intensive goods goes uncompensated. They deserve compensation. One solution that will diminish the opportunity cost incurred in not reducing emissions and not enjoying the co-benefits in reducing GHG emissions associated with this option

is to argue for 'indirect' compensation, through targeted subsidies for alternative energy sources and products. The design of MBIs has to be sensitive to the poorer populations of richer nations in order to respect their right to energy. Again, in order to overcome the trade-off between fairness and emissions reduction is to ensure that poorer populations have access to affordable sustainable alternatives.

Once we consider efficiency and fairness jointly, the best course of action is to invest in developing and diffusing sustainable alternatives, provided the design of MBIs is sensitive to communities living in already polluted areas. It is important to point out in this context that the current economic models underestimate the benefits of reducing fossil-fuel pollution, as they also underestimate the impacts of dangerous climate change (Stern 2016).

7.10 Conclusion

In another attempt to engage in reflective integration this chapter delivered on the two contributions it set out to do, that is: (a) to provide a sound normative foundation for a bottom-up approach to climate justice, and for MBIs, around the notions of a 'right to energy', the 'duty not-to-harm' and a 'capacity principle'; and (b) to identify the normative elements from theories of climate justice that should constrain the design of MBIs so that these become instruments of justice. We have seen that 'the duty not to harm', the 'capacity principle' and the 'right to energy' can be integrated in a cohesive way both in the justification for implementing MBIs and in the determination of how the revenues it generates should be distributed.

The conclusion of this chapter was aligned with the conclusion of Chapter 6: the integration of inputs from other disciplines (in this case climate science and climate economics) contributed to the normative work of the philosopher. Yet, going beyond the goal of Chapter 6, we have seen in this chapter that thinking about climate justice as it applied to one policy in particular allowed balancing objectives in order to overcome important trade-offs, notably between emissions reductions and fairness.

The general argument of last chapter retained its form in the conclusion of this chapter. This allowed for an institutional design of MBIs that could potentially be efficient (administratively and economically), fair (such as by subsidising alternative energies, especially those targeting poorer populations) and could achieve important emissions reduction (by the selling of permits that decrease in number every year and by making more resources available to invest in green energy and other technologies). MBIs generate revenues that could be used to tackle injustices related to climate change mitigation and adaptation. MBIs can be designed based on considerations of justice and become an instrument of justice.

This relates to the discussions about the choice to develop our theory of climate justice for it to be theoretically sound within a theory of global justice, or do it so as to make it relevant for political decision-making (Zellentin 2015: 12). This chapter showed that a bottom-up approach contributes to the global justice debate and finds ways to better understand the problems that must be addressed to make justice progress. Again, I say that understanding how to make it relevant for political decision-making can contribute to making it theoretically sound. Importantly, this was done without an exact determination of the distribution of 'first-order' responsibilities of climate justice.

Hopefully nations will pursue their own robust initiatives in order to reduce GHG emissions. Hopefully developed nations will contribute to the Green Climate Fund. And all nations could pursue objectives 1 and 2. Many nations already have plans to reduce emissions. From 2020 onwards, we expect most nations to be engaged in considerable emissions reductions strategies following the Paris Agreement. On the practical aspect, this chapter wanted to point out that by designing mechanisms we understand how to implement justice, in ways that considers the input from other disciplines so that action-guiding principles can be formulated.

CONCLUSION

This thesis set out to show how non-ideal theory responds to the 'paradox' of ideal theory, according to which ideal theory is incapable of guiding action under non-ideal circumstances. Indeed, this thesis aimed to show that action-guidance is the province of non-ideal theory. Non-ideal theorising is the response to the apparent paradox of ideal theory. More precisely, the thesis explored how principles of justice are formulated in light of non-ideal circumstances and empirical features of the world, why a non-ideal theory methodology is well-suited to address problems of justice, and what are the elements that should structure the content of a non-ideal theory. These contributions were exemplified in the contexts of tax competition, climate justice and carbon pricing. I do not intend to summarize each of one of the chapters this thesis. This has already been done at the end of each chapter. This conclusion will thus concentrate on some of the central insights that emerge from the thesis as a whole.

By exploring the main elements that any theory that wishes to guide action should consider, this thesis showed that a non-ideal theory can respond to demanding action guidance considerations, much more demanding than the minimal considerations that ideal theory can address. How non-ideal theory responds to the paradox of ideal theory is one core aspect of the ideal and non-ideal theory debate, which has been underexplored. This thesis highlighted the contribution of a non-ideal approach to our theorising about politics and explored the action guiding potential of non-ideal theory by identifying its defining features. I showed that non-ideal theorising provided not only non-normative supplementations to political theory (in terms of elements that counted as empirical and social scientific input), but also some of its core normative components. The problem it set out to fix was centrally a problem with the normative incompleteness of ideal theory. This normative completion and the supplementation of political theory by facts and social science is accomplished at the level non-ideal theory.

I believe it is clear now that the ideal/non-ideal theory debate captures a number of fundamental disputes in political theory, notably about: the methodology of theories of justice (should we proceed from high principles or ongoing social problems?); the unitary aspect of theories of justice (can bottom-up approaches contributes to debate about justice?); the function of political theorising (should political theories be action guiding and, if so, how?); and the division of labour in humanities and social sciences (with philosophers in charge of formulating principles and identifying values, and social scientist in charge of implementing them).

I will answer these questions in turn and succinctly. There is value in proceeding from real-world problems especially in what past and present injustices are given the normative weight they deserve. Secondly, bottom-up approaches contribute to debates about justice in that they favour the in-depth examination of practices so that the nature of the problem (in terms of agency and collective action, for instance) is revealed and the response to it can be modelled accordingly. Thirdly, given that the philosopher can reach conclusions about real world cases by seeking to bring about change, it seems that aiming to formulate action-guiding theories of justice is part of her or his work in non-ideal theory. Finally, the division of labour between political philosophy and disciplines such as social sciences and economics is not impermeable: the political philosopher should be capable of integrating work from these disciplines in order to complete her or his normative work. The importance of the ideal and non-ideal theory debate is clearly observable in that it involves the relation between political action and the judgements political philosophers should be in a position to make.

This thesis promised to present the structure of a non-ideal theory of justice and highlight its contributions to political theorising. The argumentative structure of this thesis followed a pattern. This pattern was to argue for two theses. The two theses comprised a meta-theoretical component and an empirical reflection. The meta-theoretical claims were peculiar in nature precisely because sufficient empirical input was necessary to support them, which was accomplished with three case studies. These

two central theses aimed to provide a framework that integrates theoretical and empirical material to arrive at action-guiding bottom-up contributions to justice.

The first thesis is what I called the reflective integration thesis. The reflective integration thesis states that: if we wish to formulate principles of justice that can guide action in non-ideal circumstances, we need to integrate ideal and non-ideal theory, and the way to integrate ideal and non-ideal theory is by seeking reflective equilibrium between these levels. I argued that a theory of justice can only become action guiding if it integrates the appropriate elements from ideal and non-ideal theory. In the climate justice case study, I showed that non-ideal circumstances and empirical features of the world constrained the interpretation of principles. For instance, the empirical analysis of Chapter 6 corroborated with the moral assessment that the 'equal emissions per capita' principle leads to considerable injustices and showed that the 'principle of capacity' should not be applied to emissions reduction targets but rather to mechanisms of compensation. The way to effect the distribution of burdens in climate justice should be constrained by inputs from other disciplines, so that we can make sure to achieve the required emissions reduction to avoid catastrophic climate change in a way that is fair. Also, the consideration for mitigation efficiency revealed that integrating empirical research allows developing a position that includes an argument to lower the overall cost, or sacrifice, of mitigation, by pointing to cost-effective mitigation opportunities. It determined as well the feasible paths to the necessary emissions reduction objectives and how these determinations impacted on the formulation of principles. Distributive justice thought in reflective integration with these empirical constraints resulted in the moral constraints being understood in the context of climate politics, and perhaps facilitating the moral constraints to become a more informative part of that process.

In the carbon pricing case study, I argued that considerations of justice contributed to the design of climate policy, specifically market-based instruments for climate change mitigation. This chapter went beyond the objective of proving the meta-theoretical claims of this thesis in that it aimed at offering an account of a bottom-up theory of climate justice as it applied to carbon pricing. Justice-based considerations allowed me to put forward a design of MBIs so that these mechanisms became

simultaneously effective climate change mitigation tools and instruments of justice. This allowed to relax some of the tension between 'mitigation efficiency' and fairness in this debate. Moreover, data about the economic and health-related co-benefits of climate change mitigation investments informed us about why the collective action aspect of the problem should not inhibit national or sub-global initiatives to mitigate climate change. These data helped balance between options that allow us to pursue the greatest possible emissions reduction, while being as sensitive as possible to vulnerable and poorer populations, within countries and at the global level.

The second thesis is what I call the 'Non-Ideal Theory Thesis'. This thesis was composed of four claims: (i) ideal theory is not analytically prior to non-ideal theory, (ii) theories of justice – notably theories of global justice and climate justice – are best advanced as pluralist theories composed of bottom-up contributions, meaning that different sets of principles regulate different practices, (iii) determining our duties of justice in a non-ideal world is not a question of non-compliance with ideal principles, but rather an inquiry about compliance with non-ideal principles and (iv) whether there is path-dependence between ideal principles and non-ideal reforms can only be determined through non-ideal theorising. Claims (i) and (ii) defined the *methodology* of non-ideal theory. Claim (iii) related to the *content* of non-ideal theory. Claims (iv) focused on the transition between the two theories. I explored these four claims by distinguishing, in a non-ideal theory of institutional design, between four components of action-guidance: compliance, feasibility, fact-sensitivity and path-dependence. In the tax competition case study, I showed that no ideal theory of global justice was required to tackle tax competition. On the contrary, I showed that ideal theories of global justice run the risk of being insensitive to the particularities of the individual practices that influence the formulation of principles of justice. Through the examination of specific practices, in this case tax competition, I have highlighted the elements that action-guiding principles had to address. Notably, I showed that understanding the problem as a case of an 'asymmetric prisoner's dilemma' underscored that the action-guiding principles had to include specific compliance mechanisms. Whether our ideal duties of justice changed in light of non-compliance is a wrong-headed question. Favouring

compliance is what the work in non-ideal theory is about. In the chapter about carbon pricing, we have seen that the institutional design of climate policy allowed to define second-order duties in a way that contributed to fulfilling first-order duties, without settling on an exact definition of first order duties.

Also, importantly, in the case of tax competition we saw the path-dependence between ideal and non-ideal theory was heavily informed by equating feasible institutional proposals with an account of the nature of the problem given by an examination of the practice. The normative work of the philosopher was informed by the real world understanding of the practice. In the case of climate justice, we saw how the feasibility constraints could be expanded given the institutional decisions of today, which therefore could allow for more ideal settings to be pursued.

The three case studies allowed me to highlight the action-guiding capacity of non-ideal theory, by showing that it can indeed respond to desiderata that should compose action-guidance, such as: whether non-ideal theory enables the assessment of conflicts between values and between courses of action; whether it points to circumstances that influence the compliance of agents and how that informs the formulation of regulations; whether it clarifies which feasibility constraints can relax in the future; and can we choose between different institutional settings in non-ideal theory so as to know how we can achieve more ideal settings in the future. The non-ideal level analysis allowed to identify the institutional proposals that we have 'dynamic duties' to implement, in order to move towards more desirable institutional contexts in the future.

At the end of Part I, I suggested that the role for the political philosopher at the gates of social sciences is neither the role of commander nor the role of night-watchman. This role is more collaborative and interactive than this. This will not supplant the role of philosophers as truth seekers, who follow the argument where it leads. But there is way to understand the contribution of the philosopher as non-ideal theorist, so as to realise that the philosopher will not be doing 'something else than philosophy' when doing so. This role as non-ideal theorist is dedicated to answering questions which could occupy a more central function in political philosophy, notably the relation between

incentives and compliance, how to think about path-dependence between second best options and ideal institutional settings, the trade-offs between efficiency and fairness, and the pivotal role about the choice of facts that compose a conception of justice. This indicates that there is still some work to be done in helping understand the practical role of theories of justice. There is a way to bridge the gap between ideal and non-ideal theory, but the questions that need answering to do so have not received sufficient attention.

The philosopher has something to say about practical issues but she or he has also to find ways to frame their research so as to make it more clearly accessible to social scientists. The interest for the philosopher is notably to find ethical issues in policy and institutional design. There are central ethical questions that arise in the process of regulating individual practices and designing specific policy. It is the competence of the philosopher to cast light onto those elements, to expose the moral problems and to look at identifying moral actions in all things considered judgements that involve economics, history, natural sciences and social sciences.

Problems such as tax competition and climate change, and policies like emissions trading systems, require the collaboration of researchers across a range of disciplines in order to help decision-makers meet these challenges.

References

Abizadeh, A., 2007. Cooperation, pervasive impact, and coercion: On the scope (not site) of distributive justice. *Philosophy and Public Affairs,* 35 (4), 318-358.

Adam, D., 2011. Climate change in court. *Nature Climate Change,* 1 (3), 127-130.

Aldy, J.E. & Pizer, W.A., 2014. Comparability of effort in international climate policy architecture *Harvard Project on Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School,.*

Aldy, J.E. & Pizer, W.A., 2015. The competitiveness impacts of climate change mitigation. *Harvard Project on Climate Agreements, Belfer Center for Science and International Affairs.*

Aldy, J.E. & Stavins, R.N. eds. 2007. *Architectures for agreement. Addressing global climate change in the post-kyoto world,* Cambridge: Cambridge University Press.

Aldy, J.E. & Stavins, R.N., 2011. The promise and problems of pricing carbon: Theory and experience. *Harvard Project on Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School,.*

Anderson, E., 2010. *The imperative of integration* Princeton: Princeton University Press.

Armstrong, C., 2012. *Global distributive justice* Cambridge: Cambridge University Press.

Arneson, R., 1989. Equality and equal opportunity for welfare. *Philosophical Studies,* 56, 77-93.

Athanasiou, T. & Baer, P., 2002. *Dead heat: Global justice and global warming* New York: Seven Stories Press.

Avi-Yonah, R.S., 2000. Globalization, tax competition, and the fiscal crisis of the welfare state. *Harvard Law Review,* 113, 1573-1676.

Avi-Yonah, R.S., 2007. *International tax as international law : An analysis of the international tax regime* Cambridge: Cambridge University Press.

Avi-Yonah, R.S., 2016. A proposal for unitary taxation and formulary apportionment (ut+fa) to tax multinational enterprises. *In* Dietsch, P. & Rixen, T. eds. *Global tax governance. What is wrong with it and how to fix it.* . Colchester: ECPR Press.

Beitz, C.R., 1979. *Political theory and international relations* Princeton, N.J: Princeton University Press.

Bell, D., 2008. Carbon justice? The case against a universal right to equal carbon emissions. *In* S, W. ed. *Seeking environmental justice* Amsterdam: Rodolphi, 239–57.

Bessell, S., 2015. Individual deprivation measure, engendering poverty measurement. *Global Justice Post-2015.* Yale University

Blake, M., 2002. Distributive justice, state coercion, and autonomy. *Philosophy and Public Affairs,* 30 (3), 257-296.

Bowen, A., 2011. *The case for carbon pricing.* London and Leeds, UK.

Brock, G., 2008. Taxation and global justice: Closing the gap between theory and practice. *Journal of Social Philosophy,* 39 (2), 161-184.

Brock, G. & Pogge, T., 2014. Global tax justice and global justice. *Moral Philosophy and Politics* 1(1), 1-15.

Broome, J., 2012. *Climate matters: Ethics in a warming world* London: W. W. Norton & Company.

Brown, E., 2011. Plato's ethics and politics in the republic. *The Stanford Encyclopedia of Philosophy,* Winter Edition.

Caney, S., 2002. Cosmopolitanism and the law of peoples. *Journal of Political Philosophy,* 10, 95-123.

Caney, S., 2005. *Justice beyond borders* Oxford: Oxford University Press.

Caney, S., 2011. Climate change, energy rights, and equality. *In* Arnold, D.G. ed. *The ethics of global climate change.* Cambridge: Cambridge University Press.

Caney, S., 2012. Just emissions. *Philosophy and Public Affairs,* 40 (4), 255-300.

Caney, S., 2014. Two kinds of climate justice: Avoiding harm and sharing burdens. *The Journal of Political Philosophy,* 22 (2), 125–149.

Caney, S., 2015. Responding to non-compliance. *In* Heyward, C. & Roser, D. eds. *Climate justice in a non-ideal world.* Oxford: Oxford University Press.

Caney, S. & Hepburn, C., 2011. Carbon trading: Unethical, unjust and ineffective? *Royal Institute of Philosophy,* 69.

Chung, R., 2013. Ideal and non-ideal theory in international relations. *Academic Foresights,* 8, 1-9.

Clausing, K.A., 2016. The nature and practice of tax competition. *In* Dietsch, P. & Rixen, T. eds. *Global tax governance. What is wrong with it and how to fix it.* Colchester: ECPR Press.

Cohen, G.A., 1997. Where the action is: On the site of distributive justice. *Philosophy and Public Affairs,* 26 (1), 3-30.

Cohen, G.A., 2003. Facts and principles *Philosophy and Public Affairs,* 31 (3), 211-245.

Cohen, G.A., 2008. *Rescuing justice and equality* Cambridge, Mass. ; London: Harvard University Press.

Cohen, G.A., 2009. *Why not socialism?* Princeton: Princeton University Press.

Cohen, J. & Sabel, C., 2006. Extra rempublicam nulla justicia? *Philosophy and Public Affairs,* 34 (2), 147-175.

Cullen, S. 2011. Research finds carbon tax compensation unjustified *ABS news.*

Dessler, A.E., 2012. *Introduction to modern climate change* Cambridge: Cambridge University Press.

Dietsch, P., 2015. *Catching capital: The ethics of tax competition* Oxford: Oxford University Press.

Dietsch, P., 2016. Whose tax base? The ethics of global tax governance. *In* Dietsch, P. & Rixen, T. eds. *Global tax governance what is wrong with it and how to fix it.* Colchester: ECPR Press.

Dietsch, P. & Rixen, T., 2012. Tax competition and global background justice. *The Journal of Political Philosophy,* 22 (2), 1-28.

Dietsch, P. & Rixen, T., 2014. Redistribution, globalisation, and multi-level governance. *Moral Philosophy and Politics,* 1 (1), 61-18.

Dietsch, P. & Rixen, T., 2016. Global tax governance: What it is and why it matters. *In* Dietsch, P. & Rixen, T. eds. *Global tax governance. What is wrong with it and how to fix it?* Colchester: ECPR Press.

Dunn, J., 1990. *Interpreting political responsibility* Cambridge: Cambridge University Press.

Dworkin, R., 1981. What is equality? *Philosophy and Public Affairs,* (v.10, p.185-246), p. 185-246.

Dworkin, R., 2000. *Sovereign virtue : The theory and practice of equality* Cambridge, Mass. ; London: Harvard University Press.

Estlund, D., 2014. Utopophobia. *Philosophy & Public Affairs,* 42 (2), 113-134 Available from: http://dx.doi.org/10.1111/papa.12031.

Farber, D., 2013. Carbon leakage versus policy diffusion: The perils and promise of subglobal climate action. *Chicago Journal of International Law,* 13 (2), 359-379.

Farrelly, C., 2007. Justice in ideal theory: A refutation. *Political Studies* 55 (4), 844-864.

Feinberg, J., 1973. Duty and obligation in the non-ideal world. *The Journal of Philosophy,* 70 (9), 263-275.

Gajevic Sayegh, A., 2016. Justice in a non-ideal world: The case of climate change. *Critical Review of International Social and Political Philosophy*, 1-26 Available from: http://dx.doi.org/10.1080/13698230.2016.1144367.

Gardiner, S., 2011. *A perfect moral storm* Oxford: Oxford University Press.

Genschel, P., 2002. Globalization, tax competition, and the welfare state. *Politics & society,* 30 (2), 245-275.

Geuss, R., 2005. *Outside ethics* Princeton ; Oxford: Princeton University Press.

Geuss, R., 2008. *Philosophy and real politics* Princeton, N.J. ; Oxford: Princeton University Press.

Gilabert, P., 2012a. Comparative assessments of justice, political feasibility, and ideal theory. *Ethical Theory and Moral Practice,* 15 (1), 39-56.

Gilabert, P., 2012b. *From global poverty to global equality : A philosophical exploration,* 1st ed. Oxford: Oxford University Press.

Goodin, R., 1995a. Political ideals and political practice. *British Journal of Political Science,* 25, 37-56.

Goodin, R.E., 1995b. Political ideals and political practice. *British Journal of Political Science,* 25 (01), 37-56.

Goodin, R.E., 1998. *The theory of institutional design,* 1st pbk. ed. Cambridge [England] ; New York, NY, USA: Cambridge University Press.

Gosseries, A., 2005. Cosmopolitan luck egalitarianism and the greenhouse effect. *In* Weinstock, D. ed. *Global justice, global institutions,.* Calgary: Canadian Journal of Philosophy.

Green, F., 2015. *Nationally self-interested climate change mitigation: A unified conceptual framework.* London.

Green, F. & Stern, N., 2016. China's changing economy: Implications for its carbon dioxide emissions. *Climate Policy*, 1-15 Available from: http://dx.doi.org/10.1080/14693062.2016.1156515.

Griswold, C., 1999. *Adam smith and the virtues of enlightenment* Cambridge: Cambridge University Press.

Hamlin, A.P. & Stemplowska, Z., 2012. Theory, ideal theory and the theory of ideals. *Political Studies Review,* 10, 48-62.

Holtman, S.W., 1999. Kant, ideal theory, and the justice of exclusionary zoning. *Ethics* 110, 32-58.

Hunter, D., 2007. Negligence in the air: The duty of care in climate change litigation. *University of Pennsylvania law review,* 155 (6), 1741-1794.

Icescr, 1976. International covenant on economic, social and cultural rights. New York: UN General Assembly.

Ipcc, 2001. *A report of working group iii of the intergovernmental panel on climate change.* Accra, Ghana.

Ipcc, 2007. International panel on climate change fourth assessment report: Climate change. Cambridge University Press.

Ipcc, 2013a. *Fifth assessment report (ar5).* Cambridge, UK: Press, C.U.

Ipcc, 2013b. International panel on climate change fifth assessment report: Climate change. Cambridge University Press.

James, A., 2005. Constructing justice for existing practice: Rawls and the status quo. *Philosophy and Public Affairs,* 33 (3), 281-316.

James, A., 2012. *Fairness in practice : A social contract for a global economy* New York: Oxford University Press.

Jamieson, D., 2005. Adaptation, mitigation, and justice. *Perspectives on Climate Change: Science, Economics, Politics, Ethics,* 5, 217-248.

Kant, I., 1970. *Kant's political writings* Cambridge: Cambridge University Press.

Kaswan, A., 2014. Reconciling justice and efficiency: Integrating environmental justice into domestic cap-and-trade programs for controlling greenhouse gases. *In* Arnold, D.G. ed. *The ethics of global climate change.* Cambridge: Cambridge University Press.

Kraut, R., 1992. The defense of justice in plato's republic *In* Kraut, R. ed. *The cambridge companion to plato.* Cambridge University Press.

Krugman, P. 2010. Building a green economy *New York Times*, April 7.

Lawford-Smith, H., 2009. Debate: Ideal theory - a reply to valentini. *The journal of political philosophy,* 18 (3), 357-368.

Levy, J.T., 2014. There's no such thing as ideal theory. *In* University, M. ed. Social Sciences Research Network.

Liebowitz, S.J. & Margolis, S.E., 2000. Path dependence. *Encyclopedia of Law and Economics, Volume I. The History and Methodology of Law and Economics*, 981-998.

Lipsey, R.G. & Lancaster, K., 1956. The general theory of second best. *The Review of Economic Studies,* 24 (1), 11-32 Available from: http://www.jstor.org/stable/2296233.

Macintyre, A., 1981. *After virtue* London: Bloomsbury

Maltais, A., 2014. Failing international climate politics and the fairness of going first. *Political studies,* 62 (3), 618-633.

Mason, A., 2012. What is the point of justice? *Utilitas,* 24 (4), 525-547.

Mckinnon, C., 2011. Climate change justice: Getting motivated in the last chance saloon. *Critical review of international social and political philosophy,* 14 (2), 195-213.

Mckinnon, C., 2015. Climate justice in a carbon budget. *Climatic Change,* 133 (3), 375-384.

Mill, J.S., 1869. *The subjection of women*: British Library.

Miller, D., 1995. *On nationality* Oxford: Clarendon Press.

Miller, D., 2001. Distributing responsibilities. *The Journal of Political Philosophy,* 9, 453-471.

Miller, D., 2007. *National responsibility and global justice* Oxford: Oxford University Press.

Miller, D., 2009. Global justice and climate change: How should responsibilities be distrbuted? *The Tanner Lectures on Human Values,* 28, 119-56.

Miller, D., 2012. Political philosophy for earthlings. *Justice for earthlings: Essays in political philosophy.* Cambridge: Cambridge University Press, 16-39.

Mills, C., 2005. 'Ideal theory' as ideology. *Hypatia* 20, 165-184.

Murphy, L.B., 2003. *Moral demands in nonideal theory* Oxford: Oxford University Press.

Nagel, T., 2005. The problem of global justice. *Philosophy and Public Affairs,* 33 (2), 113-147.

Nelson, R.R. & Winter, S.G., 1982. *An evolutionary theory of economic change* Cambridge, Massachusetts: The Belknap Press Harvard University Press.

Neumayer, E., 2000. In defence of historical accountability for greenhouse gas emissions. *Ecological Economics,* 33 (2).

Nickel, J., 2009. *Making sense of human rights* Oxford: Blackwell Publishing

Nili, S., 2014. Global taxation, global reform, and collective action. *Moral Philosophy and Politics,* 1 (1), 83-103.

Nordhaus, W., 2009. Economic issues in a designing a global agreement on global warming. *Keynote Address Prepared for Climate Change: Global Risks, Challenges, and Decisions.*

Nozick, R., 1974. *Anarchy, state and utopia* New York: Basic Books.

O'neill, O., 1996. *Towards justice and virtue : Constructive account of practical reasoning*: Camb.U.P.

Oecd, 1998. *Harmful tax competition. An emerging global issue.* Paris, France.

Oecd, 2011. *Transfer pricing legislation - a suggested approach.* Paris, France.

Oecd, 2014. *Tax transparancy. Report on progress.* Paris, France.

Okin, S.M., 1989. *Justice, gender and the family* New York: Basic Books.

Olmstead, S. & Stavins, R.N., 2010. *Three key elements of post-2012 international climate policy architecture.* Cambridge, Mass.

Parfit, D., 1984. *Reasons and persons,* Repr. with corrections. ed. Oxford: Clarendon Press.

Peet, J. 2004. The luck of the irish. A survey of ireland *The Economist*

Phillips, M., 1985. Reflections on the transition from ideal to non-ideal theory. *Noûs,* 19 (4), 551-70.

Philp, M., 2008. Political theory and the evaluation of political conduct. *Social Theory and Practice,* 34 (3), 389-410 Available from: http://www.jstor.org/stable/23558714.

Plato, *The republic of plato* New York, NY: Harper Collins Basic Books.

Pogge, T., 1989. *Realizing rawls* Ithaca: Cornell University Press.

Pogge, T., 2004. "Assisting" the global poor. *In* Chatterjee, D. ed. *The ethics of assistance: Morality and the distant needy.*

Pogge, T.W., 2000. On the site of distributive justice: Reflections on cohen and murphy. *Philosophy & Public Affairs,* 29 (2), 137-169 Available from: http://dx.doi.org/10.1111/j.1088-4963.2000.00137.x.

Pogge, T.W.M., 1992. Cosmopolitanism and sovereignty. *Ethics,* 103, 48-75.

Pogge, T.W.M., 2008. *World poverty and human rights,* 2nd ed. Cambridge: Polity.

Posner, E.A. & Sunstein, C.R., 2008. Justice and climate change. *Harvard Project on International Climate Agreements, Belfer Center for Science and International Affairs, Harvard Kennedy School.*

Posner, E.A. & Weishbach, D., 2010. *Climate change justice* Princeton: Princeton University Press.

Rawls, J., 1993. *Justice as fairness* New York: Irvington.

Rawls, J., 1996. *Political liberalism,* Pbk. . ed. New York ; Chichester: Columbia University Press.

Rawls, J., 1999a. *The law of peoples* Cambridge, Mass. ; London: Harvard University Press.

Rawls, J., 1999b. *A theory of justice,* Rev. ed. Cambridge, Mass.: Belknap Press of Harvard Univeristy Press.

Rawls, J., 2001. *Justice as fairness : A restatement* Cambridge, Mass. ; London: Harvard University Press.

Rawls, J. & Freeman, S.R., 1999. *John rawls : Collected papers* Cambridge, Mass. ; London: Harvard University Press.

Rezai, A., 2016. Who pays to avoid climate change, and how much does it cost? *ESRC Seminar Series: Climate Ethics and Climate Economics.* University of Nottingham

Risse, M., 2012. *Global political philosophy* Basingstoke: Palgrave Macmillan.

Rixen, T., 2011a. From double tax avoidance to tax competition: Explaining the institutional trajectory of international tax governance. *Review of international political economy,* 18 (2), 197-227.

Rixen, T., 2011b. Tax competition and inequality: The case for global tax governance. *Global Governance: A Review of Multilateralism and International Institutions,* 17 (4), 447-467.

Rixen, T., 2016. Institutional reform of global tax governance: A proposal. *In* Dietsch, P. & Rixen, T. eds. *Global tax governance. What is wrong with it and how to fix it.* Colchester: ECPR Press.

Robeyns, I., 2008. Ideal theory in theory and practice *Social Theory and Practice,* 34 (3), 341-362.

Ronzoni, M., 2009. The global order: A case of background injustice? A practice-dependent account. *Philosophy and Public Affairs,* 37 (3), 229-256.

Ronzoni, M., 2014. Global tax governance: The bullets internationalists must bite – and those they must not. *Moral Philosophy and Politics,* 1 (1), 37-59.

Rorty, R., 1998. *Truth and progress* Cambridge: Cambridge University Press.

Rosen, M., Wolff, J. & Mckinnon, C., 1999. *Political thought* Oxford ; New York: Oxford University Press.

Rousseau, J.-J., 1762. *The social contract.*

Sangiovanni, A., 2007. Global justice, reciprocity, and the state. *Philosophy and Public Affairs,* 35 (1), 4-39.

Sangiovanni, A., 2008. Justice and the priority of politics to morality. *The Journal of Political Philosophy,* 16 (2), 137-164.

Scanlon, T., 2000. *What we owe to each other,* 1st Harvard University Press pbk. ed. Cambridge, Mass.: Belknap Press of Harvard University Press.

Scanlon, T., 2003. Rawls on justification. *In* Freeman, S. ed. *The cambridge companion to rawls.* Cambridge: Cambridge University Press.

Schelling, T., 1992. Some economics of global warming. *American Economic Review,* 82 (1).

Sen, A., 2006. What do we want from a theory of justice? *The Journal of Philosophy,* 103 (5), 215-238.

Sen, A., 2009. *The idea of justice* Cambridge, Massachusetts Harvard University Press.

Shue, H., 1994. After you: May action by the rich be contingent upon action by the poor? *Indiana Journal of Global Legal Studies,* 1 (2), 343-366.

Shue, H., 1995. Ethics, the environment and the changing international order. *International affairs,* 71 (3), 453.

Shue, H., 2009. Global environment and international inequality *International affairs* 75 (5), 531-545.

Shue, H., 2014. *Climate justice* Oxford: Oxford University Press.

Simmons, A.J., 2010. Ideal and non-ideal theory. *Philosophy and Public Affairs,* 38 (1), 5-36.

Singer, P., 2004. *One world : The ethics of globalization* New Haven, Conn., London: Yale University Press.

Smith, A., 1759. *The theory of moral sentiments* Oxford: Oxford University Press.

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A., Folke, C., Gerten, D., Heinke, J.,

Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B. & Sörlin, S., 2015. Planetary boundaries: Guiding human development on a changing planet. *Science,* Available from: http://www.sciencemag.org/content/early/2015/01/14/science.1259855.abstract.

Stemplowska, Z., 2008. What's ideal about ideal theory? *Social Theory and Practice,* 34 (3), 319-340.

Stemplowska, Z. & Swift, A., 2012. Ideal and nonideal theory. *In* Estlund, D. ed. *The oxford handbook of political philosophy.* Oxford: 'Oxford University Press'.

Stern, N., 2007. *The economics of climate change : Stern review on the economics of climate change* Cambridge: Cambridge University Press.

Stern, N., 2016. Current climate models are grossly misleading. *Nature,* 530 (7591), 407-409.

Stiglitz, J., 2006. A new agenda for global warming. *The Economists' Voice,* 3 (7).

Swift, A., 2008. The value of philosophy in nonideal circumstances. *Social Theory and Practice,* 34 (3), 363-387.

Tan, K.-C., 2004. *Justice without borders : Cosmopolitanism, nationalism, and patriotism* Cambridge: Cambridge University Press.

Taylor, C. 2015. Ireland the main beneficiary of us foreign direct investment *The Irish Times*

Tjn, 2005. *Tax us if you can: The true story of a global failure.* London.

Unfccc, 2014. [online]. United Nations Available from: http://unfccc.int/essential_background/items/6031.php [Accessed Access Date 2015].

Valentini, L., 2009. On the apparent paradox of ideal theory. *The Journal of Political Philosophy,* 17 (3), 2009.

Valentini, L., 2010. Global justice and practice-dependence: Conventionalism, institutionalism, functionalism. *The Journal of Political Philosophy,* 19 (4), 399-418.

Valentini, L., 2011a. *Justice in a globalized world: A normative framework* Oxford: Oxford University Press.

Valentini, L., 2011b. A paradigm shift in theorizing about justice? A critique of sen. *Economics and Philosophy,* 27 (3), 297-315.

Valentini, L., 2012. Ideal vs non-ideal theory: A conceptual map. *Philosophy Compass,* 7 (9), 585-668.

Van Parijs, P., 2004. Basic income: A simple and powerful idea for the twenty-first century. *Politics & Society,* 32 (1), 7-39 Available from: http://pas.sagepub.com/content/32/1/7.abstract.

Vanderheiden, S., 2008. *Atmospheric justice : A political theory of climate change* Oxford: Oxford University Press.

Waldron, J., 1999. *Law and disagreement* Oxford: Oxford University Press.

Walzer, M., 1987. *Interpretation and social criticism* Cambridge: Harvard University Press.

Weinstock, D., 2006. The real world of (global) democracy. *Journal of Social Philosophy,* 37 (1), 6-20.

Weinstock, D., 2009. Motivating the global demos *Metaphilosophy,* 40 (1), 92-108.

Weitzman, M., 2013. Can negotiating a uniform carbon price help to internalize the global warming externality? *Journal of the Association of Environmental and Resource Economists,* 1 (1/2), 29–49.

Wiens, D., 2015. Against ideal guidance. *The Journal of Politics,* 77 (2), 433-446.

Wiens, D., 2016. Assessing ideal theories: Lessons from the theory of second best. *Politics, Philosophy & Economics,* 15 (2), 132-149 Available from: http://ppe.sagepub.com/content/15/2/132.abstract.

Williams, B., 1981. *Moral luck : Philosophical papers 1973-1980* Cambridge: Cambridge University Press.

Wilson, J., 2014. Embracing complexity: Theory, cases and the future of bioethics. *Monash Bioethics Review,* 32, 3-21.

Wisor, S., 2015. Individual deprivation measure, poverty measurement. *Global Justice Post-2015.* Yale University.

Wolff, J., 2009. Global justice and norms of co-operation: The 'layers of justice' view. *In* De Wijze, S., Kramer, M.H. & Carter, I. eds. *Hillel steiner and the anatomy of justice: Themes and challenges.* Routledge, 16--34.

Wolff, J. & De-Shalit, A., 2007. *Disadvantage* Oxford: Oxford University Press.

World Resources Institute, C., 2015. Climate analysis indicator tool.

Wto, 1994. Agreement on trade-related aspects of intellectual property rights (trips. Uruguay.

Young, I.M., 1997. *Dilemmas of gender, political philosophy, and policy* Princeton, N.J. ; Chichester: Princeton University Press.

Ypi, L., 2010. On the confusion between ideal and non-ideal in recent debates on global justice. *Political Studies Review,* 58 (3), 536-555.

Ypi, L., 2013. What's wrong with colonialism. *Philosophy and Public Affairs,* 41 (2), 158-191.

Ypi, L., Goodin, R. & Barry, C., 2009. Associative duties, global justice, and the colonies. *Philosophy and Public Affairs,* 37 (2), 103-135.

Zellentin, A., 2015. How to do climate justice. *In* Brooks, T. ed. *Current controversies in political philosophy* London: Routledge

Zucman, G., 2013. The missing wealth of nations: Are europe and the u.S. Net debtors or net creditors? *The Quarterly Journal of Economics,* 128 (3), 1321.

Zucman, G., 2014. Taxing across borders: Tracking personal wealth and corporate profits. *The Journal of Economic Perspectives,* 28 (4), 121-148.