A HOBESIAN THEORY OF JUSTICE

Dissertation zur Erlangung des
akademischen Grades des Doktors der
PHILOSOPHIE

vorgelegt von
Da Rosa Ribeiro, André Gustavo

an der

Universität Konstanz

Geisteswissenschaftliche Sektion
Fachbereich Philosophie

Tag der mündlichen Prüfung: 27.03.2019
Referent: Peter Stemmer
Referent: Jacob Rosenthal

Konstanz, 2019
To Adalberto, Angela, Daisy, and Érica
## Summary

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter One: Contractarianism</td>
<td>7</td>
</tr>
<tr>
<td>1. The Social Contract and Its Tradition</td>
<td>7</td>
</tr>
<tr>
<td>1.1 Normativity</td>
<td>8</td>
</tr>
<tr>
<td>1.2 Justification</td>
<td>11</td>
</tr>
<tr>
<td>1.3 The Social Contract</td>
<td>13</td>
</tr>
<tr>
<td>2. Contractualism and Contractarianism</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Different Perspectives</td>
<td>17</td>
</tr>
<tr>
<td>2.2 The Differences in Hobbes and Locke</td>
<td>19</td>
</tr>
<tr>
<td>3. Contractarianism Thoroughly Presented</td>
<td>22</td>
</tr>
<tr>
<td>3.1 Instrumental View of Practical Rationality</td>
<td>22</td>
</tr>
<tr>
<td>3.2 Morality as the Object of Justification</td>
<td>28</td>
</tr>
<tr>
<td>4. Conclusion of the Chapter</td>
<td>38</td>
</tr>
<tr>
<td>Chapter Two: Distributive Justice</td>
<td>40</td>
</tr>
<tr>
<td>1. The Problem of Justice</td>
<td>41</td>
</tr>
<tr>
<td>1.1 Distributive Justice and Its Problem</td>
<td>41</td>
</tr>
<tr>
<td>1.2 The Problem of Justice in the Contractarian Context</td>
<td>45</td>
</tr>
<tr>
<td>2. Contractarian Distributive Justice</td>
<td>49</td>
</tr>
<tr>
<td>2.1 Private Property and Distributive Justice</td>
<td>50</td>
</tr>
<tr>
<td>2.2 A Contractarian Conception of Justice: Justice as Natural Liberty</td>
<td>52</td>
</tr>
<tr>
<td>2.3 Is Justice as Natural Liberty Sufficient?</td>
<td>59</td>
</tr>
<tr>
<td>3. David Gauthier's Lockean Proviso</td>
<td>62</td>
</tr>
<tr>
<td>3.1 Gauthier's Argument for the Proviso</td>
<td>62</td>
</tr>
<tr>
<td>3.2 Gauthier's Proviso Considered</td>
<td>69</td>
</tr>
<tr>
<td>4. Conclusion of the Chapter</td>
<td>75</td>
</tr>
<tr>
<td>Chapter Three: The Circumstances of Moral Justification</td>
<td>76</td>
</tr>
<tr>
<td>1. The Circumstances of Justice</td>
<td>77</td>
</tr>
<tr>
<td>1.1 David Hume's View</td>
<td>77</td>
</tr>
<tr>
<td>1.2 H.L.A. Hart's and John Rawls' Views</td>
<td>81</td>
</tr>
<tr>
<td>1.3 Thomas Hobbes' View</td>
<td>83</td>
</tr>
<tr>
<td>2. The Circumstances of Moral Cooperation</td>
<td>85</td>
</tr>
</tbody>
</table>
2.1 Unsocial Sociability .................................................................................................................. 85
2.2 The Social Balance of Power .................................................................................................. 90
  2.2.1 Equality of Powers ............................................................................................................. 90
  2.2.2 The Concept of Power ....................................................................................................... 93
  2.2.3 The Problem with the Equality of Powers View ................................................................. 96
  2.2.4 A Balance of Power .......................................................................................................... 106
2.3 Absence of Widespread Poverty ............................................................................................. 110
  2.3.1 Extreme Scarcity and Extreme Abundance ...................................................................... 110
  2.3.2 Poverty .............................................................................................................................. 113
2.4 The Three Circumstances of Moral Cooperation ................................................................. 116
3 Critiques ...................................................................................................................................... 117
  3.1 D. Clayton Hubin's Critique and the Justificational Interpretation ....................................... 117
  3.2 Peter Vanderschraff's Critique and the Difference between Justification and Compliance ... 123
4 Conclusion of the Chapter ......................................................................................................... 129
Chapter Four: The Rational Conception of Stability .................................................................... 131
1 Stability ...................................................................................................................................... 132
  1.1 The Fool, the Sensible Knave, and Compliance .................................................................... 132
  1.2 Rawls' Two Senses of the Idea of Stability .......................................................................... 133
  1.3 The Rational Conception of Stability .................................................................................... 139
2 Stability and Rebellion .............................................................................................................. 144
  2.1 Stability as a Public Good ........................................................................................................ 144
  2.2 The Rebel ............................................................................................................................... 153
3 Rebellion and the Circumstances of Moral Justification ......................................................... 158
  3.1 Renegotiation Expectations ................................................................................................... 159
  3.2 The Social Balance of Power ................................................................................................ 161
  3.3 Absence of Widespread Poverty .......................................................................................... 166
4 Conclusion of the Chapter ......................................................................................................... 169
Chapter Five: The Instability of Justice as Natural Liberty ....................................................... 171
1 Preliminary Remarks ................................................................................................................ 172
2 Justice as Natural Liberty and the Social Balance of Power .................................................. 176
  2.1 A Material Account of Power ................................................................................................ 176
  2.2 Economic Power ................................................................................................................... 177
    2.2.1 What is Economic Power ............................................................................................... 177
    2.2.2 Shifts in the Distribution of Economic Power ............................................................... 179
The central goal of this work is to suggest a Hobbesian approach to distributive justice. We employ the contractarian method of moral justification developed by philosophers such as Thomas Hobbes, David P. Gauthier, and Peter Stemmer to justify two principles of distributive justice that create substantive redistributive demands. The goal of one the principles is to correct inequalities in power, whereas that of the other principle is to protect individuals from poverty. The justification we provide for the principles draws heavily on a conception of stability that we develop from views espoused by John Rawls, Gauthier, and James M. Buchanan. We argue that without principles of distributive justice that create substantive redistributive demands, no moral order can be considered stable. Moreover, we also argue that the stability of the moral order is something in which rational agents are strongly interested. Referring to such interest, we manage to justify the two principles with the Hobbesian method. We argue that rational agents would be willing to cede some of their wealth to sustain redistributive efforts, provided that such cession of wealth is a requirement to maintain the stability of the moral order.

Keywords: Hobbesian Contractualism, Distributive Justice, Stability, Redistribution of Wealth.

Stichwörter: Hobbescher Kontraktualismus, Verteilungsgerechtigkeit, Stabilität, Vermögensredistribution.
ACKNOWLEDGMENTS

The Ph.D. that comes to an end with this doctoral dissertation would not have been possible without the contributions of several people, all of whom I would like to thank.

In the first place, I am grateful to my supervisor Peter Stemmer, who was always supportive and helpful during my research, and whose insightful commentaries and critiques helped the development of the research.

In the second place I want to thank Marcelo de Araujo, my friend and teacher who supported and guided me all the way from my undergraduate course to my Ph.D.

The whole Ph.D. was made possible by a grant given to me by a cooperation between the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), from Brazil, and the Deutsche Akademischer Austauschdienst (DAAD), which funded my Ph.D. for four years. I am grateful to the Brazilian and German governments, as well as the CNPq and the DAAD for the funding they provided me.

I am also grateful to all the friends I made during my stay in Germany, who have helped me with insightful discussions and emotional support, and who I will hold as very dear friends throughout my life, Alexandra Diehl, Almir Mehmedovic, Fabian Stöhr, Henrique Neves, Keiichi Satoh, Ludmila O. Feitosa, Luís F. F. André, Mizuki Satoh, Pedro F. M. de Andrade, Takayuki Kurihara, Tim Ho, Yuchan Pang, and a few others.

Lastly, I want to thanks Adalberto, Angela, Daisy, Érica, and Fernando for all the support they have offered me not only during the Ph.D., but during my life as a whole.
This work pertains to two long philosophical traditions. The first is the social contract tradition and the second is the tradition of distributive justice. Our central goal here is to discuss distributive justice from the perspective of the social contract tradition.

Although our discussion is done from the viewpoint of the social contract tradition, it is only part of the tradition that pertains to our discussion. The outlook that informs our discussion is characterized by the social contract theories developed by Thomas Hobbes and his followers. That is the branch of the social contract tradition known as Hobbesian contractualism or contractarianism. It is from the perspective of Hobbes and his followers that we tackle the topic of distributive justice. Our final goal is to suggest a Hobbesian theory of distributive justice.

Hobbesian contractualism is a theory of moral justification. It is about demonstrating how moral rules are normative for those under their authority. In this respect, approaching the field of distributive justice from the Hobbesian perspective refers to tackling the issue of justifying conceptions of distributive justice. As we discuss ahead, Hobbesian theories have had considerable difficulties to fulfill this task. Such difficulties are more significant when the conceptions examined entail strong normative requirements regarding redistribution of goods. The goal we set to ourselves in this work is precisely to overcome such hardship that the Hobbesian tradition faces. We want to bring the Hobbesian method to bear on the justification of a conception of distributive justice that entails substantive redistributive obligations.

The structure of the work is divided as follows. The work is divided into three broad parts. Each part contains two chapters. Chapters one and two contain the presentation of the conceptual and methodological core that inform the whole discussion. In it, we present two central ideas for our argument, Hobbesian contractualism and distributive justice. In chapters three and four, we advance our own views regarding two ideas that are vital for our argument. Those are the idea of the
circumstances of moral justification and the idea of stability. The third and last part of the work contains our central argument. It is in this part that we bring to bear the conceptual apparatus that we have been discussing since chapter one on the matter of justifying a demanding conception of distributive justice. We advance here a negative argument for the justification and positive argument one. We argue that every moral order must include a minimally demanding conception of distributive justice and we also detail in some length how this conception must look like.

In chapter one we present the method of moral justification that informs our whole discussion, Hobbesian contractualism. The chapter has three parts. The first part contains an exposition of the social contract tradition as a whole. In this section, we discuss three critical concepts that appear throughout the tradition, and which indicate us what the social contract is about. The concepts are: normativity, justification, and agreement. In the second section, we discuss the main divide in the social contract tradition, namely the split between contractarians and contractualists. We consider the differences between the two segments of the tradition to stress the particularities of contractarianism. In section three, the last section of the chapter, we discuss contractarianism in particular. The purpose of the discussion is to present Hobbesian contractualism as a whole, stressing its peculiarities as a social contract theory. For this end, we refer to the theories proposed by Hobbes, Gauthier, Stemmer, and the economist James M. Buchanan.

Chapter two contains the presentation of the topic of distributive justice. In this chapter, we draw on specific discussions from particular authors to present what distributive justice is about and to contextualize it in the Hobbesian tradition. In the first section, we pursue two goals. The first goal is the presentation of the formal concept of distributive justice through Aristotle's discussion of the subject. The second goal is to contextualize the idea of distributive justice in the Hobbesian tradition. In order to achieve this goal, we refer to Hobbes' and Buchanan's discussions of the subject. The idea is to understand how a contractarian can approach the field of distributive justice. At the end of the first part of the chapter, we have a clear view of what a contractarian philosopher must fulfill to suggest a conception of distributive justice. The second section of the chapter contains the presentation of a particular conception of justice that the contractarian method can justify. We call this conception justice as natural liberty (JNL). Such conception is not the one whose justification we pursue in our work. JNL is an extremely minimalistic conception of justice regarding redistributive duties. It is a conception similar to Robert Nozick's Libertarian conception of justice. The role that JNL plays in our work is that it is a representation of the difficulty that contractarianism has in dealing with conceptions of distributive justice that create substantive redistributive demands. Still in section two, we discuss in which sense JNL is an insufficient conception of justice and why that is problematic. We examine here what is it in the Hobbesian
theory that creates problems for it to approach the matter distributive justice. We discuss, in short, the problem that the Hobbesian tradition has with duties with beneficence. Lastly, we close the section with a detailed discussion of Gauthier's idea of a Lockean proviso. The proviso is a rule of distributive justice that Gauthier attempts to justify with the Hobbesian method. We consider Gauthier's argument for the rule and use it as an example to examine the difficulty the method faces to justify such rules.

Chapter three is dedicated to one particular goal, advancing an interpretation of an idea suggested by David Hume. The idea we consider is that of the circumstances of justice. Hume suggests the idea in the context of his own practical philosophy, which has not been traditionally regarded as contractarian. Despite that, the idea can play a vital role in the Hobbesian method, if appropriately adapted to it. In the first section of the chapter, we present Hume's idea and elaborations of the idea done by later philosophers, John Rawls in particular. Moreover, we suggest that the idea was already present, although implicitly, in Hobbes' practical philosophy. We consider Hobbes' view at the end of the section, and it indicates us how the circumstances of justice can be interpreted from a contractarian outlook. The second section contains the presentation of our interpretation, which we refer to as the circumstances of moral cooperation. In this section, we discuss three circumstances in particular, namely unsocial sociability, the social balance of power, and the absence of widespread poverty. In essence, the three circumstances are but reformulations of those present in Hume's idea. It is mainly in the case of the second circumstance, the social balance of power, that we distance ourselves from Hume's original idea. The third and last section presents and addresses critical remarks made by two philosophers, D. Clayton Hubin and Peter Vanderschraff, concerning the idea of the circumstances of justice. Considering each philosopher's criticisms of the circumstances help us to elaborate our interpretation further. From Hubin's views, we suggest that the circumstances should be understood as the circumstances of moral justification. From Vanderschraff's view, we stress the difference between justification and compliance. Such difference leads us to the fourth chapter, where we discuss the subject of stability.

Chapter four is, as we indicated, about stability. In it, we discuss in-depth the idea of stability connected with moral cooperation. We start the chapter recapitulating the difference between the justification of moral rules and compliance with them that we discussed at the end of chapter three. We refer here to figures suggested by Hobbes and Hume to stress the difference, namely Hobbes' Fool and Hume's Sensible Knave. The figures indicate what the matter of stability is about. Stability is about compliance with moral rules. With this view in mind, we proceed to consider discussions about stability conducted by John Rawls and Gauthier. With Rawls, we identify two different conceptions of stability, moral and rational stability, and decide to endorse
one of them, rational stability, and reject the other. We then move to Gauthier's analysis of the same subject. We verify that Gauthier and the Hobbesian tradition in general tackle the issue of stability in the sense of rational stability, which we discussed previously. Nonetheless, we conclude that the Hobbesian view on the topic is incomplete. The analysis ignores a fundamental factor that affects the stability of moral cooperation, which is what we term the rationale for rebellion. The second section of the chapter contains the discussion of rebellion. The rationale for rebellion is a rationale for non-compliance with moral principles or a moral order. In this section, we draw on James Buchanan's idea of a renegotiation of the terms of cooperation. The rebel is, in essence, an individual who has his request for renegotiation of the moral rules denied by the others. As a consequence, he becomes non-compliant with the moral order and acts actively to undermine it and reduce its level of stability. We argue that the rebel offers a threat to the stability of the moral order that the contractarian tradition has overlooked. It is in the context of rebellion that the circumstances of moral justification reveal their importance for our argument. It is through the circumstances that we can understand how the rationale for rebellion appears and why it offers a dire threat to stability. The rebel emerges from alterations in the circumstances of moral justification, for he is an individual that regards the moral order as unjustified. Part three of the chapter explores how alterations in the circumstances of moral justification give rise to rebels in the social system.

Chapter five is perhaps the most complex chapter of the work. Essentially, it contains the analysis of the stability of JNL, which we developed in chapter two. The whole chapter can be understood as the elaboration and development of a view quickly espoused by Hobbes in Leviathan (1651), namely that when the state cannot interfere with individuals' private property, the social order is bound to be weakened and even dissolved. We elaborate this view through an analysis of JNL, which is a conception of justice that forbids the state of interfering with the distribution of wealth in society. To demonstrate the instability of JNL, we propose to examine how its normative requirements affect the circumstances of moral justification that ensure that JNL is justified for individuals. We argue that JNL stimulates changes in the circumstances that destabilize the whole cooperative framework and that it is also incapable of solving the problems of stability that such changes provoke. To propose such an argument, we must suggest how the normative requirements of JNL affect and change the circumstances of moral justification. For this purpose, we introduce a more material view of the concept of power. In chapter three and four, we argue that power is a critical factor in the matters of justification and stability. However, in both chapters, our discussion involves the concept of power only from a formal perspective. In such discussions, we do not indicate what constitutes power materially. Our goal in chapter five cannot be attained if we maintain this formal view of power. The reason for that is that the normative requirements of JNL
refer to material things, external goods. JNL is a conception of distributive justice. It determines how external goods ought to be distributed. In this sense, if we want to examine how JNL affects the circumstances of moral justification, we must see how distributions of external goods are connected with the circumstances. To provide such a connection, we suggest understanding power in the forms of economic power and political power. Part two of the chapter revolves around presenting these two types of power and investigating in how far JNL can maintain the social balance of power or restoring the balance in the case it is shattered. Part three considers the relation of JNL with alterations in the third circumstance of moral justification, the absence of widespread poverty. Our conclusion in the chapter is that JNL is entirely unable to prevent changes in the circumstances of moral justification that create rebels and also that it has no capacity of containing the rebels after they emerge. The conclusion is, in short, that Hobbes' view is right, namely that whenever the moral order cannot interfere with individuals properties instability ensues.

Chapter six is the last chapter of the work. In chapter five, we conclude that a conception as minimalistic regarding redistributive requirements as JNL destabilizes the moral order. The first part of chapter six contains a discussion of the value of stability for rational agents. We argue that the stability of the moral order is something that rational agents have strong reason to be concerned with. We do so by demonstrating that stability is intrinsically connected with the enjoyment of the advantages that social cooperation affords individuals. It is because individuals want to enjoy such benefits that they must be concerned with stability. From the value of stability, it follows that rational agents would reject JNL as a suitable conception of justice to govern their society. JNL's failure, demonstrated in chapter five, indicates that rational agents require a conception of justice with more substantive redistributive requirements. In part two of the section, we formulate the idea of an amendment to the social contract. The idea is that of an alteration of JNL in particular aspects to ensure its capability of assuring the stability of the moral order. We formulate such alterations in the form of an amendment. The image of the amendments serves us for two purposes. The first is to indicate the rationale that leads individuals to agree with the alteration of JNL. The second is to differentiate our argument from Gauthier's. We contrapose the idea of the amendment to that of the proviso, and we argue that the idea of the amendment avoids a series of problems that the proviso must face and whose solution is not clearly successful. The third section of the chapter includes the suggestion of revised form of JNL. We present two different principles in this section. Each principle aims at solving specific shortcomings that JNL has in ensuring stability. They are the principle of balance and the principle of charity. Both principles alter JNL to allow for substantive redistributive demands. The principle of balance is introduced to ground redistributive demands regarding wealth inequality. The principle of charity is introduced to ground redistribution regarding
poverty.
CHAPTER ONE: CONTRACTARIANISM

OVERVIEW OF THE CHAPTER

Our goal in this chapter is to present the methodology that guides our discussion throughout this work. The methodology we have in mind is the method of moral justification that we employ in the chapters ahead, namely Hobbesian contractualism or contractarianism. The presentation is done in three steps. In the first step, we cover the social contract as a general idea. With references to the social contract tradition, we discuss three key elements of the idea of the social contract, namely normativity, justification, and the idea of agreement. After the formulation of a general outlook of what the social contract is about, we move to the second step. The second step concerns the divide in the tradition between contractualists and contractarians. We discuss the main differences between each side of the tradition. Here, we indicate what the elements that characterize the contractarian part of the tradition in opposition to contractualism are. Lastly, in the third step, we discuss in some detail two aspects of the contractarian tradition that we stress in the second part. The first element is the model of practical rationality employed by contractarian philosophers. The second element is both the object of the contractarian method, morality, and its purpose, moral justification.

1 THE SOCIAL CONTRACT AND ITS TRADITION

In Ancient Greece, we can find early formulations of the idea of a social contract. In the works of philosophers such as Plato and Epicurus, there are clear and explicit formulations of what came to be known later in the history of philosophy as the social contract.¹ But although the idea had already

¹ For a historical account of the social contract tradition, including formulations of the idea in the ancient world, see Mulgan (1979).
been elaborated in ancient times, the most influential formulations of the idea appeared during the Enlightenment. The idea of a social contract appears in the works of Hugo Grotius, Thomas Hobbes, Samuel von Pufendorf, John Locke, Jean-Jacques Rousseau, Immanuel Kant, and others. The idea has remained influential throughout the history of philosophy until our days. Recently, the idea of the social contract appears in the works of thinkers such as James M. Buchanan, John Rawls, David Gauthier, Jan Narveson, Thomas Scanlon, Peter Stemmer, and Malcolm Murray.

Philosophers have employed the idea of the social contract in wildly different manners, with different philosophical problems in sight, and from different philosophical perspectives. Still, despite the differences, there is a common thread in their discussions which allows us to speak of a social contract tradition. All these philosophers have engaged in the task of justifying some sort of normativity by reference to an agreement. What is at the core of the social contract tradition is an appeal to the image of an agreement as a tool for the justification of some normative entity. This appeal appears in the works of all philosophers mentioned above that we have identified with the social contract tradition. In order to have a better grasp of what is the social contract, we explore this common thread that is at the core of the tradition. We do so by examining the three key concepts that pertain this thread: normativity, justification, and agreement.

1.1 NORMATIVITY

Although a concept that is always present in our daily lives, normativity is a difficult concept to grasp in virtue of its broadness and generality. From our daily exchanges in bakeries and shops to our pursuits of good grades in school, normativity is present.² Derek Parfit has argued that one can think of normativity or normative entities in five different senses.³ Referring to these senses, Parfit suggests that there are five different conceptions of normativity. The first is what he calls the rule-involving conception. According to this conception, normativity refers to rules that distinguish between what is allowed and disallowed or what is correct and incorrect. The second conception is the reason-implying conception. This conception connects normativity with reasons. Parfit argues that a normative claim under this conception is a claim that implies that one or more individuals have, or might have, a reason or apparent reason to do something. The third conception is connected with actual or possible motivation to do something. Parfit calls this conception the motivational

² Our purpose here is not to discuss normativity, but the social contract. For this reason, we cover the theme very briefly, with a focus on the aspects that concern the social contract tradition. For an in-depth discussion of the subject of normativity, see Stemmer (2008).

conception. Under this conception, a normative state of affairs can only be considered normative if it can motivate an agent to act under specific requirements. The fourth conception is the attitudinal conception. For this conception, normativity involves some forms of attitudes toward actions performed by oneself or other individuals, such as, for example, the attitudes of approval and disapproval. The fifth conception is called the imperatival conception. In it, normativity is about commands.

The idea of the social contract can be connected with all five conceptions of normativity. Nonetheless, three conceptions seem to have some protagonism in the context of the social contract. Those are the rule-involving, the imperatival, and the reason-implying conceptions. The first two are connected with the object of justification, i.e., to what is justified by the social contract. The third conception relates to what is justification itself. We examine justification in the next section, and, for this reason, we do not discuss the relation between the social contract and the reason-implying conception of normativity now. Our focus here is how the rule-involving and the imperatival conceptions appear in the context of the social contract. Let us rapidly consider two authors inserted in the tradition, Locke and Rawls, to see how both conceptions appear in their works.

In a passage of the Second Treatise on Government, Locke defines what the social contract, or the Original Compact, as he calls it, is. It is the act by which persons constitute one Body Politick under one government, and by which they submit themselves to the determination of the majority.\(^4\) The social contract makes the Body Politick under one government the holder of political power. In another passage, Locke defines political power. He identifies it with the right of legislating, establishing private property, and mobilizing the community for the execution of laws and defense of the community against foreign threats, all that done for the public good.\(^5\) With the social contract and political power so defined, it is possible to see how normativity and the social contract are connected in Locke's approach. Such a connection is twofold. On the one side, the social contract seals the submission of individuals to the determination of the majority. The submission that Locke writes about refers to the acceptance of judgments of the majority, or, in other words, the commands of the majority. Submission indicates that the imperatival conception of normativity is in the background of Locke's discussion. It is the commands of the majority over the individual that Locke is attempting to justify with the social contract. On the other side, when one stresses to what the political power of the majority refers – legislating – one can notice that another conception of normativity is involved in the argument. That is the rule-involving conception. This conception

---

appears indirectly in the discussion. The social contract justifies the authority of the majority to create laws to govern the behavior of individuals. In this sense, the particular rules that differentiate legal from illegal behavior are justified by reference to their source, the political authority, which is itself justified by the social contract.

The importance of the two conceptions of normativity can also be seen in John Rawls' social contract theory. Rawls attempts to justify two principles of justice by reference to a social contract idea. Rawls defines a principle of justice as a principle to define the basic structure of society and determine the distribution of rights, duties, and social and economic advantages. The principles are rules, rules of distributive justice, more specifically. They are standards to differentiate just from unjust distributions of rights and duties, and economic and social advantages. We can observe, then, that the central conception of normativity underlying Rawls' contractualism is the rule-involving conception, although the imperatival conception also appears in Rawls' discussion. The principles that Rawls attempts to justify are principles of justice, i.e., moral principles. They are principles that create duties for those under their authority. In this sense, if the principles indicate that an existing distribution is unjust, then the principle also creates an obligation to correct such an injustice. The principle also issues what we may call 'commands' to perform specific actions, such as correcting unjust distributions. Hence, insofar as Rawls' view of the social contract attempts to justify principles of justice, it is connected with the rule-involving conception. Moreover, since the rules are sources of commands for individuals, the view is also connected with the imperatival conception.

In the social contract tradition, what philosophers have attempted to justify are normative entities. Since the two conceptions of normativity associated with the objects whose justification has been attempted by the tradition are the rule-involving and imperatival conceptions, we may easily guess what the objects that philosophers attempted to justify are. The objects of justification in the tradition have been either rules or authorities with imperative powers. That was exemplified above. Locke, on the one side, attempted to justify the authority of the state over the citizen, whereas Rawls, on the other hand, attempts to justify the normative force of principles of justice over individuals.

---

7 The other two conceptions of normativity, the motivational conception and the attitudinal conception, could also be associated with the idea of the social contract. One can claim, for example, that the legislative activity of the Body Politick is the taproot for our attitudes of approval and disapproval of persons' behaviors. On the other side, one can affirm that it is only through the social contract that we can ensure that individuals will be motivated to follow the determinations of the Body Politick, for they would feel unmotivated to do so otherwise. Nonetheless, it seems to us that these other two conceptions, in the context of the social contract, always appear as a derivative form of either the rule-involving or the imperatival conception. For this reason, we do not explore them further here.
8 It is not entirely clear whether the rule-involving and the imperatival conceptions are independent of one another. The concept of a rule necessarily involves some idea of authority, and thus of command. The rule is different from a
Now that we have a clearer view of what is the normativity and the normative entities to which the social contract tradition refers, we can proceed to examine the next essential factor of the tradition. That is the notion of justification.

**1.2 Justification**

We come now to the second key element of the social contract tradition, justification. The goal of the social contract is to justify some normative entity. Justification is intrinsically connected with normativity. In its practical sense justification is about giving reasons for conforming with a requirement or a demand. David Gauthier, for example, writes at the beginning of *Morals by Agreement* (1986) that his purpose in the book is to develop a justificatory framework for moral behavior and principles. Moreover, he classifies this task as that of developing normative theory. In other words, he equates the development of a justificatory framework with the development of normative theory.

In its practical sense, justification is about giving reasons for conforming with a requirement or a demand. If someone tells you to avoid visiting Northern Brazil during the summer, and you ask this person to justify why you should behave as she says you should, you are asking the person for a reason to do as she says. What you are doing is merely asking in return why you should not travel to Northern Brazil in the summer. When the person answers you that you should not go there in this season because the temperature can surpass forty-five degrees Celsius, she is giving you a reason to conform your behavior with what she tells you to do. Justification is intrinsically connected with normativity in Parfit's reason-implying sense. To justify a request to someone is to show that the person has a reason to fulfill the request.

Justification can also be understood in an epistemological sense. This sense appears when one demands an individual justify a statement such as 'all men are mortal'. The demand for justification in this case amounts to a demand for evidence that the statement is true. The formal sense, however, is not important for our discussion here and we will not examine it.

---

9 Justification can also be understood in an epistemological sense. This sense appears when one demands an individual justify a statement such as 'all men are mortal'. The demand for justification in this case amounts to a demand for evidence that the statement is true. The formal sense, however, is not important for our discussion here and we will not examine it.


11 'Justification' can be translated in German as both 'Begründung' and 'Rechtfertigung'. Stemmer differentiates the two concepts regarding their usage in connection with norms and actions. He argues that although both concepts are translated to English as 'justification', they mean different things. In Stemmer's view, 'Begründung' refers to providing a reason for something, whereas 'Rechtfertigung' refers to verifying whether something is in accordance with a norm. Social contract theories refer to justification in both senses. On the one side, they attempt to show that
What the social contract attempts to do, then, is to show that individuals have a reason to accept the normative authority of a particular normative entity. When Locke refers to the social contract to justify the authority of the state over its citizens, the purpose of the procedure is to show that the citizens have a reason to accept the constraints the state imposes on them. However, the reason required for the justification of a request, command, or rule is not any sort of reason. It must be a reason that is sufficient to elicit voluntary compliance or acceptance to the request, command, or rule in question. It must be a sufficient reason. A sufficient reason involves two factors. The first factor is that it must be strong enough to elicit the acceptance of the normative requirement. Consider the following case. In a football match, the striker of a team approaches the adversary goalkeeper and asks him to allow him to score a goal. The striker attempts to justify his request by arguing that if he scores in this match, he will undoubtedly win the golden boot of the league this season. The striker's appeal is indeed a reason for the goalkeeper to allow the striker to score one time. However, it is not a sufficient reason for the goalkeeper to accept the striker's request. The striker winning the golden boot is not something that the goalkeeper is interested in. He is interested in winning the match, and allowing the opposite striker to score one time decreases the chances that the goalkeeper's team has of winning the game. The request of the striker, then, cannot be justified to the goalkeeper.

The second factor involved in justification is that the acceptance of the normative requirement must be voluntary. The reason to accept the demand must not be a product of coercion or imposition. Consider the case of the robber in the woods suggested by Rousseau. If a robber approaches you with a pistol and commands you to hand over your wallet, you have a reason to act as the robber commands. You do not want to be hurt or killed by the robber. Consequently, the robber's command has normative force over you. The robber's pistol gives you a sufficient reason to act as the robber wants. Nonetheless, the command is an unjustified normative requirement. The command is a sheer exercise of power from the robber. You accept the command not because you want to, but because you are forced to do so. A normative requirement can only be deemed justified if an individual would voluntarily, i.e., without any exercise of coercion, accept it.

More recently, philosophers such as Rawls, Gauthier, and Stemmer have emphasized the role of rationality in the contractual justificatory procedure. In their approach, to justify a principle...
is to show that accepting the principle with its normative requirements is rational. The emphasis in rationality does not affect the prominence of the reason-implying conception of normativity regarding justification. A rational agent is an agent that acts for reasons. To argue that it is rational to act in a particular manner is the same thing as claiming that there is a sufficient reason to act in such a way.

We can now summarize the idea of justification that pertains to the social contract tradition. For social contract theorists, to justify a normative entity for someone means giving this person a reason that is sufficient to elicit her voluntary acceptance of the entity and its normative requirements. Justification shows that the person accepts to be constrained by normativity whose justification is being attempted. Notice, nonetheless, that normativity is not created through the justification procedure necessarily. That is the case of the suggestion to avoid Northern Brazil during the summer. If you accept the recommendation, it becomes normative to you to not go to this region of Brazil during summer. However, there is also the figure of unjustified normativity. A person may reject specific normative requirements but be forced to act under its constraints. Since the person's 'acceptance' is not voluntary, the normative requirement cannot be justified. That is the case of the authority of a tyrant over its subjects, for example. In no plausible sense, it can be said that the subjects accept the authority of the tyrant. Still, the commands of the tyrant have normative force for his subjects due to the coercion that the tyrant exerts upon them. In that case, justification does not create normativity.

1.3 The Social Contract

The next element we have to consider is the image of the agreement. It is a crucial aspect of the social contract tradition that the justification of normative entities refers to some form of agreement. A normative entity, such as a rule, for example, is deemed justified when all individuals over whom the entity is supposed to have normative force can agree on an agreement whose object is the attribution of normative force to the entity. One could imagine that what the social contract accomplishes is the transformation of an advice or a recommendation into a rule or a norm through an agreement. It is only if the normative entity is capable of eliciting the acceptance of every individual for whom it is supposed to be normative that, in the social contract view, it is justified.

Although it is not completely clear in the work of some philosophers whether the agreement

---

is factual or counterfactual, Kant's view on the matter has become a hallmark of the tradition.\footnote{Hobbes seems to consider that there are societies that lived in a state of nature as he depicts it. See Hobbes (1998), p. 85. In the same page, Hobbes expresses a view that is somewhat common in the social contract tradition. That is the view that the international field is a state of nature. Locke expresses the same view, see Locke (2003), p. 280.} Kant argued that the social contract is but an idea of reason.\footnote{See Kant (1793), p. 297.} What Kant's view conveys is that the justification sought by the social contract is not a historical or sociological justification. The idea is not that at some point in history a contract established some rules or some authority, and therefore the normativity connected with such rules or authority is justified. The idea is, instead, that the justification is hypothetical. The social contract is, in this sense, a thought-experiment. The manner in which the thought-experiment proceeds is somewhat simple and is divided into two steps.

The first step can be called the \textit{state of nature step}. A state of nature scenario is an imaginative exercise where the philosopher formulates how individuals lives and interactions with one another would be in a situation in which the normative entity whose justifiability is under examination does not exist. In the Enlightenment's tradition of the social contract, philosophers were concerned mainly with the justification of the political and legal authority of the state. For this reason, their formulations of the state of nature always involved a scenario in which there is no legal or civil authority to constrain individuals' interactions with one another, i.e., a condition in which there is no state. This first step involves the description of several different factors. Among these factors, three are of major importance. The first factor is the provision of an account of individuals' interests. In the Enlightenment part of the tradition, providing such factors amounted to advancing a view of human nature, whereas more recently it means suggesting a model of practical rationality.\footnote{The social contract method is directed primarily at individuals. When the procedure concludes that a certain normative entity is justified, the conclusion is that it is justified to some individuals. The procedure refers to groups insofar as a group's decision or intention can be reduced to the individual level. It is possible to affirm that the social contract tradition is committed in some length with what has been called methodological individualism. That is the belief that social phenomena, such as, for example, morality, should be explained through references to individual actions. See Uhden (2002), pp. 480-484 for a discussion of methodological individualism in the contractarian tradition. For a more general discussion on methodological individualism, see Miller (1978).} The second factor is a description of the environment in which individuals find themselves.\footnote{An important remark is necessary here. The contractarian tradition is focused mainly on examining normative entities in a limited societal context. When examining whether a given rule is justified, contractarians have in mind a rule that governs a single community. Traditionally, they do not take into account rules to regulate inter-communal, or international, relationships. Such a tendency has changed recently with the introduction of Rawlsian contractualism in the field of global justice. Authors such as John Rawls, Charles Beitz, Thomas Pogge, Darrell Moellendorf, and Gillian Brock have attempted to formulate social contract theories to justify international principles of justice. See Rawls (1999)(1), Beitz (1979), part three, Pogge (1989), ch. 5 and 6, Moellendorf (2002), ch. 2, and Brock (2009), ch. 2 and 3.} The third factor is a description of how individuals interact with one another based on their preferences.
and the environment in which they find themselves. Hobbes' description of the *Bellum omnium contra omnes* or Locke's description of the peaceful but disorganized and chaotic natural society are examples here. With the description of all three factors concluded, the model of the state of nature is completed and with it the first step.

The second step concerns introducing the possibility of an agreement on the rule or principle that the philosopher desires to justify for the individuals in the imagined state of nature. It is the *contracting step*. What it is carried out in this step is the examination of whether the presence of the normative entity in question is something that can elicit the acceptance of every individual in the state of nature, and thus if it can be the object of a contract between them. This is the step in which the view of justification we considered previously is employed. Here, the philosopher refers to a formulation of the state of nature to examine whether individuals have a sufficient reason to endorse the normative entity in question. More recently, this step is carried out utilizing a comparison. Here, the philosopher compares two different outcomes. The first is the natural outcome, i.e., the situation as it is in the state of nature. The second is the cooperative outcome, i.e., the situation in which the examined normative authority is in place. If every individual prefers the cooperative outcome over the natural outcome, it follows that every individual is willing to take part in the contract, and thus the normativity of the entity in question is justified. Otherwise, if individuals believe that they are better off in the natural outcome rather than in the cooperative outcome, then the normative entity fails the procedure and can be considered unjustified. In essence, what is behind the contracting step is the utilization of a *Unanimity Rule*.22 To inquire whether a given normative entity can be the object of the contract is the same thing as verifying whether every individual in a specific system accepts the normative constraints of the entity under examination. When a single individual rejects such restrictions, then the contract is not possible, and thus the entity cannot be justified.

In sum, we may conclude this section with a more complete and precise formulation of the idea of the social contract. We saw that the social contract is composed of three key elements: normativity, justification, and the idea of agreement. The social contract is a tool for justification. It serves to analyze whether a given normative entity, a rule or an authority, is justified for those over whom it has normative force. In this context, justification refers to whether individuals have a sufficient reason to accept the entity and its normative requirements voluntarily. The justification,

---

22 See Stemmer (2013), p. 78 and also Murray (2017), p. 5. Murray argues that the basic rule of contractarianism is what he calls a *Consent Principle*. The principle states that 'an act is morally permissible if an only if all parties (or their proxies) are competent and suitably informed, and voluntarily, or would voluntarily, consent.' The consent principle is but a more articulated formulation of the unanimity rule. It is interesting to note that if we depict the contractarian procedure as a procedure that verifies the adequacy of a normative entity with a rule of unanimity, the procedure becomes a justificatory procedure in the sense of 'Rechtfertigung', and not of 'Begründung', as discussed by Stemmer. See note 11 above.
however, is not actual in the sense that we evaluate what actual individuals in the world think of the rule. It is, instead, a thought-experiment. The experiment revolves around the idea that if a group of individuals would accept to take part, together, in a contract whose object is to confer normative force to the rule or authority under examination, then the entity under examination is justified. What provides the connection between the thought-experiment and the actual people over whom the examined entity has normative force is the descriptions of the individuals and the environment included in the experiment. It is expected that such descriptions, although not entirely in accordance with empirical reality, capture the elements of reality that pertain individuals and their environment that are most relevant for convincing individuals in the real world of the results of the experiment.

The central goal of the social contract doctrine can be formulated as follows: The justification of a normative entity – such as a rule or a command – employing a counterfactual agreement between individuals over whom the normative entity is supposed to have power. Now that we have a general view of what the social contract tradition is about, we proceed to the examination of the two main branches of the tradition, contractualism and contractarianism.

2 CONTRACTUALISM AND CONTRACTARIANISM

The social contract tradition divides itself into two branches: contractualism and contractarianism. This division already appears in the Enlightenment's social contract tradition, but it has become sharper recently. On the contractualist side, we find Locke, Rousseau, Kant, and Rawls. As for the contractarians, among them are Hobbes, Buchanan, Gauthier, and Stemmer. The exact name of each branch in the tradition varies. Contractualism is also known as Kantian contractualism or political contractarianism. On the other hand, contractarianism is also known as Hobbesian contractualism or moral contractualism. In our discussion, we use these terms interchangeably. In this sense, contractualism, Kantian contractualism, and political contractarianism should be understood as the same thing, and the same is valid for contractarianism, Hobbesian contractualism, and moral contractualism.

What differentiates the two branches from one another are specific details in how they employ the social contract. As argued before, the justification performed by the social contract

---

23 We follow here the overview of the social contract tradition suggested by Stephen Darwall, and also present in Lisa Herzog's overview of the modern social contract tradition. See Darwall (2003), pp 1-7, and Herzog (2013) for the exposition of the divide between contractualism and contractarianism. There are other forms to divide the tradition. David Boucher and Paul Kelly, for example, stress that there are three varieties of social contract theories, moral contractarianism, civil contractarianism, and constitutional contractarianism. See Boucher and Kelly (1994), pp. 1-16, for such other overview of the social contract tradition.
works through a thought-experiment divided into two phases, the state of nature phase and the contracting phase. There are differences between both sections of the tradition in both phases, but they concentrate on the first phase. The most significant differences lie in the first part of the state of nature phase, the conception of rationality employed, and in the objects that each line of the tradition seeks to justify. Let us consider such differences now.

2.1 Different Perspectives

One of the vital parts of the social contract approach is to formulate a description of a state of nature. Such a description involves providing an account of the forces that govern individuals interests and preferences. During the Enlightenment age, accounting for these forces amounted to suggesting a conception of human nature, whereas nowadays philosophers refer to models of practical rationality when they approach the task of determining individuals interests and preferences. With regards to this task, both contractualism and contractarianism agree on an element. They both affirm that individuals are generally self-interested. Individuals act typically based on what they regard as important. The difference appears regarding a particular factor that determines individuals' interests in a state of nature. Contractualists, in general, think that moral elements figure among these interests, whilst contractarians refuse to include morality among the factors that drive individuals to act in the state of nature. Contractarians believe that the deliberations done by individuals in the state of nature phase are always a calculus of advantages and disadvantages. In contractualist views, on the other side, morality already appears in the state of nature as an impartial constraint on the pursuit of individuals' interests which cannot be reduced to the individuals' self-interest. Contractarians reject to include such type of constraints in their state of nature formulations. Hence, in a state of nature situation, the perspective of the contractarian is that if an individual desires some good, the only thing that can hinder this individual from acting to acquire the good is if such an action ends up giving the individual more disadvantages than advantages. On the other side, contractualists usually admit that individuals in the state of nature already have some form of moral constraint on the pursuit of their interests.

The description of the environment in which individuals find themselves is an element that does not change much in the two branches of the tradition. Hume has referred to this part of the thought-experiment as the description of the outward circumstances in which a convention of justice or a social contract come into being. Later, Rawls has referred to a conjunction of these
circumstances and the description of individuals as the *circumstances of justice*. We do not discuss further the circumstances of justice here since we do so in chapter three, which is dedicated to such discussion exclusively.

In virtue of the differences in how they conceive the state of nature phase, different descriptions of the contracting phase appear between contractarians and contractualists. The idea in the background is the same, a comparison between how advantageous or disadvantageous for oneself is accepting the normative entity under examination. However, the comparison has a different tone in each branch. The difference arises from the different perspectives on the presence of morality in the state of nature. In the contractarian side, the comparison is made solely through a calculus of advantages and disadvantages. If accepting the normative entity is more advantageous than refusing it, then the individual is willing to take part in the social contract, and thus the entity under consideration is justified for him. Contractualists, on the other hand, admit the presence of moral rules and principles in the state of nature. Although they often assume individuals to be self-interest, the presence of morality gives a different tone to the comparison. The comparison is in the contractualist perspective not only a calculus of advantages and disadvantages, but also an evaluation of whether the participation in the social contract accords with the set of moral rules and principles that inform individuals' choices. The contractualist contracting phase is more complex than the contractarian because the evaluation at its center involves additional factors, namely moral principles. The contractualist analysis, then, also involves an examination of the adequacy of the normative entity under scrutiny regarding certain moral principles.

Apart from the contractual procedure itself, the two branches differ in another important aspect. It is often the case that contractualists and contractarians have in mind the justification of normative entities of different natures. Contractualists have been traditionally concerned with the justification of *political normativity* whereas contractarians were mainly concerned with *moral normativity*. Whereas authors like Locke, Rousseau, Kant, and Rawls employed the social contract to justify the political authority of the state, and the normativity of political principles of justice in the case of Rawls, authors such as Hobbes, Gauthier, and Stemmer used the contractual procedure to justify moral rules.²⁴ To give some materiality to the comparison between the two branches, we go over a rapid overview of two philosophers, each belonging to a segment of the tradition, Locke and Hobbes.

---

²⁴ Hobbes is an exception to this case. Hobbes' social contract theory has both a moral and a political dimension. It involves an attempt to justify certain moral rules and also an attempt to justify the state.
Hobbes has provided what perhaps the most influential view of the state of nature is. The *Bellum omnium contra omnes*, the war of all of against all.²⁵ Hobbes starts his depiction with a view of how individuals are. In this respect, he argues that individuals are fundamentally self-interested, that they put themselves ahead of all the others, and that they have roughly equal physical and mental attributes. These are the two factors that mainly characterize individuals in the Hobbesian thought-experiment. Hobbes adds that individuals are locked in a condition where the goods they desire are scarce. Scarcity is, in his view, the taproot of diffidence between persons. When two individuals desire the same good, and they cannot have it at the same time due to the good's scarcity, diffidence between them arises. From diffidence, a war between individuals emerges, i.e., they gain a reason to attack one another. An additional reason for conflict is that in the state of nature, individuals cannot be sure about the intentions of the others. They do not know whether other persons want to hurt them or not. In this situation of ignorance, the best option to avoid the others' aggression is to anticipate the aggression of another and thus be the aggressor first. Since that is a perception that every single individual has, the anticipation logic leads to the installment of a general state of conflict, a state of war. For Hobbes, the state of war is the state of nature. Without a common authority to put in check each individuals' intentions of encroaching in other individuals' interests, the anticipation logic drives everyone to a general conflict, a state of war.²⁶ Hobbes argues explicitly that during the state of war, the differences between just and unjust or good and evil are meaningless and empty.²⁷ Albeit Hobbes writes about a law of nature and a right of nature, both concepts have no a moral meaning. The Hobbesian law of nature is a precept of reason which states that each person ought not to act in a manner that is destructive for her own life. It involves no prohibition of encroaching upon other individuals' interests or life. The right of nature in the Hobbesian view is but the liberty each has of using his power to preserve oneself. Again, it involves no restriction on the use of one's power against other individuals.²⁸ A direct consequence of this state of war is that the lives of individuals are assumed to be terrible, nasty, brutish, and short. It is due to the misery of the state of nature that individuals are led to take part in the social contract. For Hobbes, the agreement establishes the rules to terminate the state of war, the rules that give

²⁸ See Hobbes (1998), p. 86. Gauthier has discussed comprehensively the notions of 'right of nature', 'law of nature', and 'obligation' in the context of Hobbes' philosophy. For such discussion, see Gauthier (1969), part II. See also Kavka (1986), ch. 8 and 9, for a discussion of the same concepts.
meaning to just and unjust and correct and incorrect. It is an agreement that establishes morality between individuals. Such rules are justified because they can be an object of a social contract, i.e., in a comparison between the state of nature, where the rules have no normative force, and a state in which they have normative force, individuals find more advantageous the latter than the former. Hobbes' argument also has a political aspect. We do not examine such aspect here. Instead, we leave this aspect for later chapters, where the examination best suits our discussion. Let us see now how Locke's argument proceeds.

Locke characterizes the state of nature as the state of absolute freedom, but he refuses to call it also a state of license. He argues that although individuals are uncontrollably free in the state of nature, it does not follow that they do not have any obligations. Locke claims that there is a law of nature to govern the state of nature. This law of nature obliges individuals not to harm one another in their life, health, liberty, and possessions.\(^ {29}\) The reason behind this obligation is the claim that individuals are but property of their Maker, and that their bodies and those of the others are, therefore, not their own to dispose of as they will. From the claim that every individual is but the property of their Maker, Locke also derives a mandate, a right, that every individual has to administer justice to offenders of the law of nature. In the Lockean state of nature, every individual has executive power over all the others to execute the law of nature.\(^ {30}\) Although violations of the law of nature might occur in the state of nature, Locke does not think, like Hobbes does, that the state of nature is a state of war. He argues that what defines the state of nature are individuals living a communal life under reason – the law of nature – but without any common authority or superior among them. The state of war is, in its turn, a particular state that may arise in the state of nature. The state of war is installed by the use of force or a declaration that one will use force to encroach in the life, health, liberty, and possessions of another.\(^ {31}\) The installment of a state of war gives every individual in the community a right to punish the offender who started it. As it is in Hobbes' view, the social contract is for Locke the tool that individuals use to exit the state of nature. However, the reason that Locke gives individuals for preferring to take part in the contract rather than not doing so is slightly different from the one given by Hobbes. In the same manner that Hobbes does, Locke argues that the state of nature offers an inconvenience for the lives of individuals. This inconvenience is related, as it is in Hobbes' argument, to the disadvantages that the state of war brings to individuals.\(^ {32}\) It is also related to the difficulty of properly executing the law of nature in the state of nature, since every individual has a right of doing so.\(^ {33}\) Lastly, the moral component of

---

\(^ {29}\) See Locke (2003), p. 102.
\(^ {30}\) See Locke (2003), p. 102.
Locke's state of nature is also present in his view of the contracting phase. The state that is justified through the social contract is a state whose power over its subjects has certain restrictions. It cannot, for example, confiscate part of an individual's property or subject him to a particular form of political authority without the individual's consent.\footnote{See Locke (2003), p. 161.} It is only when the social contract includes provisos for such restrictions that individuals show a willingness to take part in it. Otherwise, the normative authority of the state cannot be justified.

We may now consider the differences that appear between both branches of the social contract tradition as they appear in the views of Hobbes and Locke. The first difference relates to the presence of morality in the thought-experiment. Although both Hobbes and Locke write about a law of nature, it is only in Locke's case that the law of nature can be understood as a form of moral law. Hobbes explicitly writes that there is no morality or justice in the state of nature, whereas Locke is also explicit to write the contrary, and affirm that individuals do have moral obligations in the state of nature. In the Hobbesian state of nature, a person can attack another, and she cannot be morally reprehended for doing so. In the Lockean natural condition, on the other side, such an action can be morally reprehended by reference to the law of nature. A second clear difference concerns the depiction of the state of nature. Whereas Hobbes describes it as the ruthless place where people's lives are terrible, Locke refers to a community of individuals living under the law of nature. The Hobbesian natural state cannot be described as 'community' in any meaningful way. The existence of a community requires a minimum of cohesion and harmony between its members, which are two things that cannot develop between persons in the Hobbesian state of nature. The anticipation logic hinders the emergence of any sort of community. It is true that the Lockean natural community is vulnerable to the inconveniences caused by a state of war. Nonetheless, the natural condition is not a condition of permanent war between individuals in Locke's view as it is in Hobbes'. In the contracting phase, we notice a third difference. In Hobbes' the sole factor that drives the move out of the state of nature is the advantage afforded by the cessation of the state of war. However, for Locke, the consideration of a better and more efficient application of the law of nature is a factor for the superiority of the civil state over the natural state. A fourth difference concerns the object that each philosopher attempts to justify. What Hobbes attempts to justify is the installment of rules – moral rules – between individuals that will terminate the state of war between them.\footnote{Hobbes' argument also involves the justification of the state, the Leviathan. Nonetheless, the state is introduced in the context of the justification of moral rules. The state is the sanctioning mechanism that assures that compliance with the rules is always rational. In chapter four we discuss this matter in detail.} The civil authority is but an accessory in this respect. It is an authority to enforce the rules established. Locke, on the other side, is concerned with justifying the civil authority itself. In the Lockean social
contract, the moral rules, the law of nature in his view, are not determined nor defined by the contract. The contract is but a tool to assure that the already valid and normative laws of nature be properly applied and enforced by an authority mutually acceptable by every individual. In other words, Locke is more concerned with justifying the political authority that will be responsible for enforcing moral rules that are valid and normative even in the absence of such authority. A fifth and final difference is the presence of moral considerations in the state of nature. In Hobbes', the thought-experiment starts in a scenario without moral principles, rules or convictions. Individuals have no moral duties or obligations. In Locke's case, it is different. There are moral obligations already in the state of nature. Such moral elements also influence the contracting phase. The principles restrict the type of political body that can be justified.

Now that we covered the split in the social contract tradition and the main differences between the two branches, we proceed to the last section of the chapter. In the section, we analyze and present one of the two branches in particular. That is the branch of contractarianism, which is the moral theory that will inform our whole discussion throughout this work.

3 CONTRACTARIANISM THOROUGHLY PRESENTED

Contractarianism is the branch of the social contract tradition that descends from Hobbes' philosophy. For this reason, it became also known as Hobbesian contractualism. Among its contemporary proponents, we find James Buchanan, David Gauthier, Peter Stemmer, and Malcolm Murray. As presented in the last section, contractarianism differentiates itself from the other section of social contract theory for the two key factors. First, contractarians assume that individuals' preferences are assumed to be determined exclusively by their interests. Moral principles or factors play no role in determining individuals' preferences in the state of nature. Second, what contractarians attempt to justify with the social contract is the normativity of moral rules.

3.1 INSTRUMENTAL VIEW OF PRACTICAL RATIONALITY

In the state of nature, Hobbes argued, there is no justice or distinction between just and unjust. There, theft and fraud are virtues because they are proper instruments for the pursuit of one's interests. In the contractarian picture, individuals are depicted as self-interested rational agents. In
this picture, what is rational and irrational is defined by the interests and desires of the individual. An action is rational insofar as it accords with an interest or desire of the agent. Gauthier, for example, identifies practical rationality with utility-maximization. He depicts individuals in the state of nature as rational agents who act toward the maximization of their utility. In this sense, given several possibilities of action, the action whose performance is rational for the agent is the one that affords him the higher amount of utility possible. Stemmer adopts a similar view concerning practical rationality. He emphasizes that an action can be deemed to be rational only by reference to someone's desires. In his opinion, as it is the case with contractarians in general, practical rationality – the rationality connected with acting – is but an instrumental form of rationality. To act rationally is merely to act following one's desires and interests. This view of practical rationality has been referred to as the Humean conception of practical rationality, the utility-maximizing conception, or simply the instrumental conception of practical rationality.

Consider the following example of the picture of practical rationality that informs contractarianism. Imagine that someone asks you whether you find it rational to read a full biography of Klemens W. L. von Metternich. You can sit down and reflect on several reasons for learning about Metternich's life. Metternich was a paragon of the European aristocracy in an age where such aristocracy was facing its downfall. Metternich was the first state chancellor of the Austro-Hungarian Empire. The work he performed as a statesman was fundamental for the establishment of the so-called Concert of Europe that maintained peace in Europe until the First World War. There are countless reasons to read Metternich's biography. Despite the plurality of reasons, what defines whether reading it is rational or not is whether one has an interest or desire that is promoted by reading the book. In this respect, if you are, for example, interested in the history of the Austro-Hungarian Empire, reading about Metternich's life is rational for you. The same thing is valid if you are interested in how the life of the German nobility was during the XIX Century. In case you have such interests, to read the book affords you more utility than not doing so, and therefore it is rational for you to read it.

The instrumental view of rationality does not contain any substantive assumptions about individuals' preferences. It states merely that rational agents always act with the intent of maximizing their preferences. The factors responsible for the determination of the preferences are not part of the model, these must be supplied independently of it. Contractarian philosophers often provide these factors in their theories. Hobbes, for example, writes about passions such as the fear

37 The basic model of rationality adopted by contractarian philosophers resembles that used in economic models, the so-called homo oeconomicus. For an in-depth analysis of the homo oeconomicus model, see Kirchgässner (2008).
39 That is the view famously suggested by Hume. See T 2.3.3.4, SBN, pp. 414-415.
of death and the desire of things that are necessary to living comfortably. Stemmer, in his turn, provides a more detailed account of such factors. Stemmer writes about what he calls fundamental general interests. Such interests have two essential characteristics. They are very basic, and every rational agent has them. The examples he provides are the interest in surviving, in health and physical integrity, interest in happiness, and the interest in not being humiliated or socially diminished. Although somewhat too broad and abstract, fundamental general interests serve the contractarian philosopher to give some materiality to the thought-experiment of the state of nature. By reference to these interests, contractarian philosophers can examine whether a given norm or principle is acceptable to rational agents. The interests give substance to the instrumental view of practical rationality and allow the contractarian philosopher to have a more material view of whether the acceptance a particular normative entity is rational for individuals.

Even though an account of the interests and desires that inform the picture of practical rationality is necessary for the contractarian endeavor, contractarians avoid committing themselves with far-reaching and ambitious lists of desires and interests. The contractarian justificational procedure refers directly to individuals' interests and desires when it is used to evaluate normative entities. One should not forget that although the social contract procedure involves a simplification of human nature in the form of a model of practical rationality, its objective is to show to real persons in the real world that a certain rule or principle is justified. The model is supposed to reflect conditions that real persons would accept as valid to guide the evaluation of principles and norms. In this regard, the inclusion of self-interest and fundamental general interests in the model is surely non-problematic, since persons are self-interested and they all have fundamental general interests. When the model includes just these two factors, its likely that the scope of the justification will be vast and general regarding the individuals to whom the justification is directed, since the factors that sustain the justification are also very broad and general. However, when additional factors are included in the model, it is likely that the scope of justification will be reduced. The more substantive and rich is the constellation of interests that informs the model of rationality, the smaller is the sphere of individuals for whom the justification can be done. Stemmer's fundamental general interests are interests that every individual is supposed to have. If such interests are combined with the model of instrumental rationality in the contractarian endeavor, the consequence is that

42 Rawls makes his point very well. He considers the question of why one should any interest in the outcome of the social contract thought-experiment. Rawls' answer is that the experiment is but the elaboration of conditions that we accept as valid and adequate to evaluate the acceptability of norms and principles. In other words, the procedure is relevant for real people because it embodies what people believe to be the appropriate conditions under which normative authority is justifiable. See Rawls (1999), p. 19.

24
whichever rule the procedure manages to justify, it will be justified for every individual in general. On the other hand, when we suppose that more substantive factors are included in the model of practical rationality, the scope of the justification is likely to be affected. John Rawls' theory offers a good example of such a case.

Albeit a contractualist, and not a contractarian, Rawls attempts to justify two principles of justice with the same instrumental view of practical rationality that permeates contractarianism. Despite the similarity, Rawls adds an element to his justificational procedure that contractarians would refuse to include in their theories. Rawls includes a conception of fairness among the factors that determine the individuals' preferences. This conception involves a series of moral convictions that are typical of liberal democratic societies such as the convictions that all individuals are free and equal, and that some liberties that individuals have may not be violated for any purpose, not even to promote a greater good or the public good. Introducing this conception of fairness constrains the contractual justification procedure. A principle is justified if it fulfills at the same time two criteria. Firstly, it must fulfill the criterion of instrumental practical rationality, i.e., individuals must be better off under the principle than they would be without it. Secondly, the principle must not conflict with the conception of fairness that Rawls assumes rational agents to endorse. In Rawls' argument, the constraint to which the conception of fairness subjects the procedure is represented by what he calls the veil of ignorance. Individuals in Rawls' description of the state of nature – the original position – are behind the veil. The veil makes them ignorant about a series of empirical factors about themselves and their position society – factors such as their social class, gender, skin color, religion, a conception of the good, and other factors. The idea is that if everyone is ignorant about such factors, the deliberation process for a social contract will then be a fair deliberation, i.e., a deliberation in accordance with the conception of fairness assumed to be endorsed by every individual. To argue that rational agents would accept to deliberate behind a veil of ignorance requires assuming that these agents somehow accept Rawls' view of fairness. Such acceptance, it is important to stress, is done in virtue moral convictions that these agents are assumed to have. They accept the constraints of the veil because they accept the moral constraints that emerge from the conception of fairness.

The problem with the model of rationality behind Rawls' procedure is that it limits the justificatory scope of the procedure. The procedure can justify norms and principles only for those

44 See Rawls (1999), pp. 3 and 12.
45 See Rawls (1999), pp. 118-123.
46 Rawls' idea is that shaping certain public institutions in accordance with such criteria is something morally unacceptable. The veil of ignorance is introduced in the contractual procedure to ensure that the result of the contract is in accordance with such moral beliefs.
that have the moral convictions that Rawls includes in the model of practical rationality. In other words, it is a procedure appropriate only to address citizens of liberal democratic societies with the moral beliefs that Rawls ascribes to such individuals.\(^{47}\) When directed against individuals who do not have such beliefs, the procedure fails. In order to avoid such problem, contractarians maintain a minimalistic posture regarding practical rationality. They do not go much beyond the desires and interests that figure in Stemmer's constellation of fundamental general interests. A more straightforward and economic model of practical rationality gives the justificational procedure a broader scope and a more robust character, since it always appeals to interests that are supposed to be very basic and strong.

The preference for a broader and stronger justificational procedure leads contractarians to reject the inclusion of any sort *altruistic interests* in their model of rationality. Following Stemmer, an altruistic interest is an interest in the well-being of another disconnected from any form of self-interest.\(^{48}\) That is the case of the interest that parents have, for example, on the well-being of their children. Another example is the individual that has his interests defined by a religious doctrine that inspires him to work for the well-being of the poor. According to the instrumental view of practical rationality, when a poor father abstains from dinner to give the food to his children, the father is acting rationally. In the same manner, when the religious person uses his money to help the poor instead of buying a new book that she desires, she is also acting rationally. In both cases, individuals are acting entirely under their interests, altruistic interests, more precisely. Contractarians such as Gauthier and Stemmer refuse to introduce such altruistic factors in their procedure of justification. Gauthier starts his investigation from what he calls a 'morally free zone', an environment where the constraints of morality have no place.\(^{49}\) Stemmer, on the other side, argues that the task of justifying a moral rule or moral principle is the task of convincing a moral skeptic to act following the moral rule or principle.\(^{50}\) And the moral skeptic is an individual without any altruistic preferences or altruistic ideals. The skeptic, in Stemmer's view, acts solely for his advantage and benefit. Contractarians have a good reason, we believe, to leave altruism outside of their projects. The inclusion of altruism in the justification procedure faces a practical problem.

The practical problem concerns the actual presence of altruism in individuals and its consequence for the justification of moral principles. If altruism is connected with the justifiability of a moral principle, this connection must have one of two forms. Either the moral principle is

\(^{47}\) What Rawls' procedure demands is much more than citizenship in a liberal democracy. It is a conviction in certain moral beliefs that are typical of liberal democratic societies, which citizens of liberal democracies may not necessarily have.


\(^{50}\) See Stemmer (2000), p. 17.
dependent on altruism to be justified or it is not. By 'dependent' it is meant here that altruism is a crucial element for the justification of the principle. If accepting the principle is possible in virtue of altruism, then the principle can be justified only to altruistic individuals. However, it is empirically verifiable that individuals are not universally nor generally altruistic. Some persons are whilst others are not. It is a contingent matter whether one is or not altruistic, for altruistic interests are not fundamental general interests. However, since what is in question here are principles that are supposed to be normative for everyone, since they are moral principles, altruism fails to be a proper ground for their justification. Any moral approach that considers altruism a vital part of the justifiability of moral principles faces this difficulty. If a moral principle is justified, then it must be justified for everyone. If its justifiability is dependent on altruism, then altruism must be general as well. However, altruism is not a general trait among humans. As a consequence, one of two things must follow. On the one side, one can maintain that altruism is vital for the justifiability of moral principles, and as a result, one must deny that moral principles are justifiable for everyone. Alternatively, one can argue that moral principles must be justified for everyone, and if there is a justifiable moral principle, then this principle cannot be justified through altruistic reasons. One must either argue that morality is a system of rules justifiable only to altruistic persons, which is an absurd claim, or one must commit oneself with the view that persons are altruistic in general, which is a false statement.

The inclusion of altruistic reasons in the model of practical rationality is likely to reduce the scope of the justificatory procedure. Nonetheless, it is important to stress that it is not the case that contractarians will not allow altruism to play any role in their justificational endeavors. Instead, what contractarians refuse is to give altruism a fundamental role in the procedure. But altruistic interests may be included in the procedure insofar as they play a non-fundamental role in it. They may be employed to increase the justificatory power of the procedure, for example. If it can be shown that a certain moral rule is justified by reference to a fundamental general interest, but that the rule also promotes a particular altruistic interest, then the justificatory power of the procedure is even stronger for a group of individuals who possess such an altruistic interest.

In sum, contractarians refer to the model of practical rationality that connects rationality with the promotion of interests and desires of the agent. When choosing between two outcomes, the rational choice is always to pick the outcome in which the agent's desires and interests are best promoted. By itself, such conception of practical rationality is materially empty. It says nothing about which interests and desires rational agents have. In this respect, contractarian philosophers

51 Stemmer conducts an interesting analysis of altruistic actions in the context of contractarianism. See Stemmer (2000), part III.
have combined the conception of practical rationality with constellations of interests and desires. The common thread in the contractarian approach to such constellations is that they avoid any interests or desire beyond what can be described as fundamental general interests, i.e., interests that are very basic and which every individual possesses. Combining the instrumental conception of practical rationality with an account of such fundamental interests, contractarians elaborate a model of rationality that allows them to evaluate whether certain norms or principles can be justified.

3.2 Morality as the Object of Justification

Now we move to examine the second key aspect of contractarianism, the object it attempts to justify. Contractarianism is a method of moral justification. It is a method to address what Christine Korsgaard has called the 'normative question'. When we state to someone that she ought to act in a certain way because doing so is morally required, the person will expect us to show why acting in such a way is morally mandatory. That is the normative question. A method of moral justification is a method to handle the normative question. To answer it is the same thing as to justify the normativity of a moral demand. The contractarian answer involves the thought-experiment we discussed above, the social contract thought-experiment. The philosopher examines whether the moral rule under examination could be the object of a contract between the individuals for whom the rule is supposed to have normative force. What determines whether an individual finds it rational or not to participate in such a contract is the model of practical rationality that informs the thought-experiment.

An assumption that underlies the contractarian method is that moral normativity is a product of hypothetical imperatives, and never of categorical imperatives. Immanuel Kant advanced the distinction between categorical and hypothetical imperatives. An imperative is a source of practical necessity. A hypothetical imperative demands the performance of an action as a means to a goal. You ought to read Metternich's biography if you desire to learn more about the history of the Austro-Hungarian Empire. The practical necessity of reading the biography appears only for those who have an objective whose fulfillment requires reading it. For those that have no such objectives, there is no practical necessity of reading Metternich's biography. Categorical imperatives, on the other hand, command the performance of particular actions for themselves, without any appeal to particular goals or ends. The intuition behind a categorical imperative can be understood with the

53 See Kant (1785), pp. 413-414.
idea of the promise. We often consider a promise the source of an obligation that we must fulfill regardless of the circumstances in which we find ourselves. Our intuition is that if we promise something to someone, we ought to fulfill that promise whether we want to do it or not, whether it promotes our interests or not. The normative force of the promise, so our intuitions go, is independent of any interest or desire we have, it is supposed to remain normative even if fulfilling it is something prejudicial for oneself. Contractarians reject the possibility of a categorical imperative. In their view, desires and interests are the sources of every type normativity, moral and non-moral. Consequently, the normative question is, in the contractarian perspective, always a question about interests and desires.

The fact that contractarianism refers to interests to answer the normative question has an interesting consequence for how the theory depicts morality. In the contractarian view, morality and moral rules are but instruments. They help individuals to promote particular interests they have. The normative question is answered when it is demonstrated that the normative entity under examination makes individuals better off regarding the fulfillment of interests and desires they have. The idea is that in the state of nature, i.e., in the absence of moral rules, individuals find themselves in a suboptimal position regarding the promotion of their own interests. Gauthier, for example, explicitly argues that morality arises from market failures.54 His view is that individuals fail to maximize their utility in the state of nature. It is precisely because there is such suboptimality that morality and its norms can be justified. Morality is an instrument to achieve an optimal level of interest-satisfaction. If the state of nature were a condition of optimality regarding the promotion of interests, it would follow that morality could not be justified to anyone. The paradigmatic depiction of the state of nature's suboptimality is Hobbes' description of the state of nature. The lack of constraints over the pursuit of self-interest leads individuals into a situation of social conflict and chaos that, in the end, is prejudicial for everyone.

Contractarians have argued that the suboptimality of the state of nature in Hobbes' view is due to the fact that, in such a state, individuals find themselves in what is called a prisoner's dilemma.55 A prisoner's dilemma is an influential mathematical game from the field of Game Theory.56 A situation with a prisoner's dilemma-like structure is one in which the utility-maximization view of rationality fails to lead individuals to the optimal outcome. We suggest below an example of a prisoner's dilemma.57

Consider that two farmers, Jack and James, find themselves in a state of nature. There are no

54 See Gauthier (1986), p. 84.
56 For a comprehensive game-theoretic discussion of the prisoner's dilemma, see Luce and Raiffa (1985), pp. 94-102.
57 The example we suggest is based on an illustration suggested by Hume known as the Farmer's Dilemma, see Hume T 3.2.5.8, SBN, pp. 519-520.
moral rules that have normative force over them. Their mutual interactions are governed solely by utility-maximization. In this sense, Jack interacts with James only when doing so is advantageous for him. The same is true for James. Given this setting, imagine that the time for the harvest has arrived for Jack's crops, but not for James'. James' harvesting season begins only in a few weeks. James could offer Jack to help him with the harvest in exchange for Jack's future help when the harvesting season starts at his own farm. Together, they can do the harvest much more efficiently and rapidly than they would do alone. James knows, however, that he and Jack are not friends nor relatives. They interact with each other solely on the basis of utility-maximization. James' offer of help would be certainly welcome by Jack since it helps him fulfilling his desire of harvesting his crops. Despite that, James has all the reasons to expect that when the harvesting season starts in his own farm, Jack will refuse to help him, even if they agreed earlier on a pact of mutual help. As soon as Jack's farm is harvested, it becomes rational for him to defect from the pact of mutual help. In this case, defection spares Jack all the sweat and work he would have along James in James' farm, and it does not cost him anything since his crops have been harvested already. After Jack and James work at Jack's farm, it becomes rational for Jack to double-cross James. James is aware of that. For this reason, he does not approach Jack to offer cooperation, and they both harvest their crops alone, without any help from the other.

In the case above, the suboptimality lies in the fact that the model of utility-maximization leads James and Jack to work separately, while cooperating is a much more efficient outcome for both of them. Such a problem has two dimensions. One concerns exploitation and the other fear of exploitation. Both of them contribute to the temptation to defect from cooperation. James does not know whether Jack will double-cross him after they finish harvesting the crops on Jack's farm. If Jack does so, it is a terrible result for James since he will have worked twice as much as he would have, in case he had not to help Jack before. It might be the case that Jack does not intend to exploit James, but James will never know until it comes the time for Jack to fulfill his part of the bargain. From Jack's perspective, the risk is too high. Given the loss in utility that Jack's defection imposes on him, it is rational for James to avoid cooperation with Jack. The other dimension of the problem regards an intend to exploit. James refrains from cooperating because there is the possibility that Jack will exploit him. Such a possibility arises from the fact that the model of rationality depicts exploitation as the rational action for Jack. After the work on Jack's farm is concluded, the rational thing for Jack is to double-cross James and spare himself from all the work in James' farm. In order for Jack to cooperate with James, Jack must refuse to maximize his utility, i.e., he must act irrationally. James' fear of exploitation and Jack's incentive to exploitation annul the possibility of cooperation between them. The dilemma faced by Jack and James is the same that individuals in a
Hobbesian state of nature face. They either fear being exploited or they desire to exploit the other, and, in both cases, they are driven to act with the logic of anticipation that Hobbes writes about, and thus they lock themselves in a suboptimal outcome.

Prisoner's dilemmas such as the one faced by Jack and James involve a problem of cooperation. For both individuals, cooperation affords a result that is more advantageous than non-cooperation. Yet, they fail to cooperate. What contractarian theories attempt to do is to provide an explanatory framework to demonstrate that cooperation is the rational choice instead of mutual defection. The solution to the problem is, in the contractarian perspective, the introduction of morality in the system. Morality is, in the contractarian view, a self-imposed constraint on the pursuit of one's interests that can be reduced to the individuals' interests.\(^\text{58}\) The imposition of the constraint finds its justification in the very utility-maximizing model that creates the problem of cooperation. Adopting the constraint is rational because it maximizes the individuals' utility in comparison to the expected outcome, i.e., mutual defection. The maximization produced by the constraint refers to shifting the result of the interaction from mutual defection to cooperation. Such a shift is beneficial for both individuals. Assured that the constraint successfully moves them to cooperation, its adoption can be said, then, to be rational for both persons. The image of the moral agreement appears to demonstrate that the introduction of the constraint is rational for everyone in the system.

The contractarian method of justification is focused on the suboptimality factor. To show that a particular moral rule or principle is justified, contractarians develop a state of nature scenario to argue that the absence of the rule or the principle leads individuals to suboptimal outcomes. That is how the contractarian method works. Consider, for example, the case of the moral rule that prohibits murder. To justify such a rule, the contractarian procedure refers to a state of nature where there are no constraints on the use of murder as a tool for the promotion of interests. In such a state, individuals will murder each other whenever doing so maximizes one's utility. With such a state of nature elaborated, the contractarian must compare it with a situation in which the rule has normative force over individuals. The rule is justified if individuals prefer to be under the authority of the rule rather than without it. In the case of the rule on murder, it is the case that individuals have a fundamental general interest in surviving. From such an interest, it emerges a derivative interest in not being murdered. In a state of nature, the interest in not being murdered is severely impaired. Although individuals might profit from the use of murder to promote their interests, they might also be victims of other individuals promoting their interests. The best outcome for each particular

\(^{58}\) See Gauthier (1986), p. 3.
individual is a scenario in which he is not constrained to refrain from using murder, but everyone else is under such restraint. Such an outcome, however, is unattainable. It is an outcome that strongly disfavors everyone but the individual who ought not to be murdered. The only option individuals have to protect themselves from murder is to introduce a norm that prohibits murder in general. Even though such a norm implies a restriction in freedom, which we can simply assume to be something undesirable, the benefits it produces outweigh the harm that this loss of freedom causes. The benefits of the norm spring from the interest in surviving. The introduction of a norm forbidding murder is a requirement for the appropriate promotion of such an interest.

We presented contractarianism so far as a moral theory that attempts to reduce moral normativity to the normativity of desires and interests. And although interests do play a fundamental role in the procedure, two other factors are of significant importance for it, power and rationality. Contractarianism can justify rules for persons only insofar as they have similar levels of power. Consider again the case of the moral rule that forbids murder. Individuals accept the rule because it helps them to promote their fundamental general interest in surviving. But rule serves this purpose only given the assumption that individuals can murder each other, i.e., if individuals are powerful enough to murder one another. Consider two individuals, John and Joe. Joe is extremely weak in comparison with John. Joe's weakness is such that he is entirely unable to murder John, should he wish to do so. In such a situation, the massive inequality in power between John and Joe hinders the possibility of any agreement on a rule prohibiting murder between them. Because Joe is no threat to John, a rule that prohibits murder between them promotes no interest of John whatsoever. What the rule supposedly affords John – safety from being murdered by Joe – is something he already has in virtue of the inequality in power between them. Consequently, the agreement is irrational for John, and thus the norm cannot be justified for him.

Every moral rule commands individuals either to perform specific actions or to abstain from certain practices. In order to justify a particular rule, the contractarian method demands that the powers to either perform or abstain from the action connected with the rule be roughly equal between those over whom the rule is supposed to have normative force. Recalling the case of the two

59 A loss of freedom can be considered something bad because it implies a reduction in one's power to promote one's interests. The idea is that when a moral rule is introduced, an individual loses a certain freedom, such as the freedom to use murder against the other, or the freedom to steal from the others, for example. Such a loss of freedom implies that the individual has less means than he had before to promote his interests. In some sense, the less freedom one has, then the less able one is to promote his own interests. For this reason, curtailing the means available for the promotion of one's interests can be described as worsening one's condition. The contractarian method of moral justification can be then, also described as a method to show that abdicating from a certain freedom is rational for individuals.

farmers, Jack and James. A pact of mutual help is rational for them only insofar as Jack can help James and vice-versa. If we assume that Jack is unable to harvest crops for some reason – if Jack is physically disabled, for example – it becomes immediately irrational for James to enter the pact with Jack, since he can draw no benefit from it, even though Jack would still benefit from James' help. The contractarian method is firmly anchored on an idea of reciprocity. Individuals cooperate morally because doing so is profitable for everyone taking part in the cooperative effort. And it is profitable for everybody because there is a degree of equality in powers with regards to the object over which individuals cooperate.\footnote{In chapter three, we discuss such requirement in detail. There, we connect it with David Hume's idea of the circumstances of justice and suggest a different understanding of the requirement.}

The third key factor for the contractarian method, along with interests and equality of powers, is a rather self-evident factor, rationality. As such a third factor, rationality refers to the capacity of making agreements.\footnote{See Hobbes (1998), p. 97, and Stemmer (2000), pp. 256-257.} More specifically, the factors refer to the capacity for coordinated self-restraint, i.e., moral cooperation. Although a human and a lion both have a strong desire to survive and have similar levels of power to kill each other, cooperation between them is impossible because the lion is not rational.

In sum, although the central factor in the contractarian method of justification is an appeal to interest, the appeal assumes two factors. It assumes that the individuals for whom the method is applied have equal powers and that they are rational. If such conditions do not obtain, it follows the procedure does not work correctly, since some individuals, the powerful, will not be interested in cooperating with the others, the weak.

Before closing the chapter, there is a subject connected with the contractarian method that we want to consider briefly. That is the matter of compliance with moral rules. We consider this matter briefly now and return to discuss it in detail in chapter four. Contractarianism is a theory of moral justification. It includes an elaboration of the conditions and factors that are key to justify moral rules and moral principles. We described these factors above. But the contractarian justification must always come accompanied by a theory of compliance. The suboptimality of state of nature situations is based on the fact that such situations have a prisoner's dilemma-like structure. For overcoming such a situation, the introduction of moral rules is recommended. But although moral rules serve the purpose of eliminating the suboptimality, they create a new problem of cooperation, the problem of compliance.

In the contractarian view, individuals are utility-maximizers. They always act in the way that is most advantageous for themselves. Consequently, whenever an individual confronts a situation in which violating a moral rule is advantageous for himself, such an individual has a strong reason not
to comply with what the rule requires of him. Although the introduction of rules results in an improvement of individuals' situations, it is not sufficient to ensure robust and stable cooperative behavior between individuals. Consider the case of the two farmers again. Jack and James can indeed come together and agree to introduce a moral rule that will constrain them to help each other in the harvesting season. Given that without such rule they would be worse off than they would with the rule, the constraints it introduces are clearly acceptable from a rational standpoint. Nonetheless, when the harvesting season starts in James' farm, and Jack's crops have already been harvested, Jack still has a strong reason to refuse to help James, i.e., to abstain from complying with the moral rule. The moral rule, by itself, does give individuals any reason to act as it requires. It is because the moral rule itself does not affect individuals' intent of acting morally that the problem of compliance arises. And that is why the contractarian moral theory must be accompanied by a theory of compliance.

Notice, however, the rationale that Jack has not to comply is strongly diluted if we assume that the dilemma he faces will repeat itself in some form in the future. If Jack expects other harvesting seasons to come in the future, and therefore that he will need James' help in the future, he becomes fully motivated to comply with the moral rule. Otherwise, in case James defects from the agreement, he cannot expect Jack to help him in the future again, i.e., the defection brings a cost to Jack. Such alteration in the dilemma amounts to transform what was a one-shot prisoner's dilemma into an iterated version of the same dilemma, a dilemma that repeats itself and whose previous choices affect how individuals act in the future. In the case of the farmers, if James knows that Jack defected from cooperation in the last harvesting season, he will not be willing to cooperate with Jack again, since he will expect the same result. In such a scenario, James' defection eliminates the chance of cooperation between the two farmers. Consequently, Jack acquires a considerable incentive to fulfill his part in the agreement and help James in his harvest.

Robert Axelrod work indicates that the prospects for cooperation in an iterated version of the dilemma are more promising than one might expect. He shows that rational agents would be better off in adopting a strategy more friendly toward cooperation rather than simply always choosing to defect from cooperation.\(^{63}\) The strategy that Axelrod favors is a variation of the so-called 'Tit for Tat' strategy.\(^{64}\) A player using 'Tit for Tat' always cooperate in his first move and afterward he mimics

---

\(^{63}\) See Axelrod, (1984), part 2. Axelrod's argument is based on a tournament between strategies for prisoner's dilemma that he simulated with computers.

\(^{64}\) See Axelrod (1997), p. 34. The variations are the so-called 'generous' TIT FOR TAT and 'contrite' TIT FOR TAT. Both variations were developed as solutions to the problem of 'noise', when such a factor is introduced in an iterated version of the prisoner's dilemma. 'Noise' refers to a small chance of one of the players misunderstanding what the other player did on his previous move. It was observed that the performance of TIT FOR TAT is considerably affected when 'noise' is introduced in the game. To cope with the decrease in performance, the two variations were introduced. In the 'generous' variation, the player using TIT FOR TAT cooperates 10% of the time that he would
what the opposing player did in the previous round. In the circumstances of an iterated prisoner's
dilemma, Axelrod observed that the 'Tit for Tat' and some variations of 'Tit for Tat' were the most
successful strategies in successive rounds of the game. The success of 'Tit for Tat' and its variations
indicates that rational agents trapped in an iterated version of a prisoners dilemma may prefer to
cooperate rather than defect from cooperation.

The rationale for compliance in an iterated version of prisoner’s dilemma may be strong
enough to sustain compliance in a considerable number of cases. However, it suffers from certain
limitations. James Buchanan observes that in a two-person setting, the rationale for cooperation is
powerful, but as the number of individuals in the setting increases, the strength of the rationale
decreases for every single individual. The more individuals are added to the system, the less
impact each individuals' actions have over the choice of the others. Buchanan thinks that after a
certain number of individuals is reached in the system, the effect that each single choice has over
the choices over the others becomes virtually insignificant. In such circumstances, every single
decision to cooperate or defect that individuals' face starts to resemble more a one-shot prisoner's
dilemma rather than its iterated version. Consequently, one should expect that a considerable
number of individuals will opt to defect from cooperation. The problem of compliance resurfaces
and with it the need for a theory of compliance.

Contractarians diverge on the adequate theory of compliance. We consider here briefly two
options to solve the problem. One is the traditional approach that we find in contractarianism,
namely the approach of sanctions. Hobbes, Buchanan, and Stemmer are proponents of such an
approach. The other approach relies on dispositions to solve the problem of compliance, and it is
suggested by Gauthier. Both approaches target the same problem, namely demonstrating that
compliance with moral rules is rational.

The sanction-based approach is that the introduction of moral rules must be accompanied
with the introductions of a system of sanctions. Such a system is a constraining mechanism
designed to provide individuals with 'artificial' incentives to engage in cooperative behavior and to
avoid defecting from cooperation. Stemmer differentiates between a pre-moral condition 1 and a
pre-moral condition 2. The pre-moral state 1 is the Hobbesian state of nature – the war of everyone
against everyone. The pre-moral state 2 is a condition in which individuals come together and

\[\text{defect, i.e., the player allows the uncooperativeness of another player to go unpunished sometimes. In the 'contrite'
version. This version of the strategy has three versions, contrite, content, and provoked. It starts cooperating and
remains cooperative until it suffers a unilateral defection from the part of another player. If that happened when it}
\text{was content, the player becomes provoked and will defect until another player makes him become content. If, on}
\text{the other hand, the own player defected when content, it becomes contrite and cooperates. When contrite, the player}
\text{becomes content after it has successfully cooperated.}

65 See Buchanan (1975), p. 85.
declare their intentions of cooperating to cease the conflict between them, i.e., they declare that they find mutually beneficial to cooperate.\textsuperscript{67} Stemmer believes that the mere declaration of an intention to cooperate is not enough to assure the justification of a moral principle. The mere declaration is but an indication that moral cooperation is rational for both individuals, but it establishes no normativity for the principle itself, i.e., individuals are not put under an obligation to cooperate. To move from the pre-moral state 2 to the moral state itself, i.e., to justify the normativity of a moral principle, Stemmer believes that individuals must agree to institute \textit{sanctions} that will work to provide artificial incentives to act following the principle.

Contractarians diverge on the exact types of sanctions that must be introduced to solve the problem of compliance. Hobbes and Buchanan refer mainly to legal or juridical sanctions. In Hobbes' argument, the moral agreement involves the creation of the Leviathan, the civil state. It is the power of the sovereign that commands the civil state that ensures the end of the state of nature and solves the problem of compliance. Stemmer, on the other side, differentiates between moral and juridical sanctions.\textsuperscript{68} Juridical sanctions are those administered by the state. Moral sanctions, in their turn, are informal social sanctions. Stemmer argues that the sanctions which pertain moral rules are those that are directly applied by society itself, without the need for a state. More specifically, Stemmer connects moral sanctions with social exclusion carried on by members of the moral community.\textsuperscript{69} The moral sanctions that Stemmer has in mind are those implicit in Axelrod's argument that leads individuals to cooperate. In an iterated prisoner's dilemma, when an individual defects from cooperation, the other parties in the system will refuse to cooperate with him in the future. They close the prospects of cooperation for the defector. He is excluded from the social order.

Gauthier's solution to the problem of compliance involves what he calls a self-imposed change in the maximizing behavior of rational agents.\textsuperscript{70} He differentiates between what he calls a straightforward maximizer and a constrained maximizer. The constrained maximizer evaluates whether the cooperative outcome under examination is nearly fair and optimal in comparison with the outcome where no one chooses to cooperate.\textsuperscript{71} If the cooperative outcome is more profitable than the other, the constrained maximizer cooperates, provided he expects everyone else to cooperate. A straightforward maximizer, on the other hand, evaluates which is the best outcome available in general, including cases which involve exploitation. He is disposed to maximize his

\textsuperscript{67} See Stemmer (2000), p. 94.
\textsuperscript{68} See Stemmer (2000), p. 104-105
\textsuperscript{69} See Stemmer (2000), p.152
\textsuperscript{70} For Gauthier's full-blown treatment of the problem of compliance, see Gauthier (1986), ch VI.
\textsuperscript{71} Gauthier often refers to 'fair outcomes' in his contractarian argument. We ignore this matter for now, but in the next chapter we discuss it in some detail.
utility in every particular choice he makes. The disposition of constrained maximization allows individuals to reach the cooperative outcome even in a one-shot prisoner's dilemma. Gauthier emphasizes that constrained maximization is not a more effective form of straightforward maximization. Instead, it is a different form of reasoning about cooperation. A constrained maximizer does not look to future benefits of cooperation to seek reasons to comply in the present. Such an attitude is an attitude of straightforward maximizers. The constrained maximizer considers what is the rational thing to do based on two factors. Firstly, a comparison that involves only the cooperative outcome and the universal defection outcome. Secondly, the likelihood of other individuals cooperating. In a different manner from the straightforward maximizer, the constrained maximizer does include in the calculus the benefits he could get by double-crossing the other parties. Moreover, constrained maximizers are willing to cooperate only with constrained maximizers. Whenever they are confronted with straightforward maximizers, they behave as straightforward maximizers to avoid being exploited by them.

Gauthier thinks that constrained maximization enables individuals to attain cooperative profits and benefits unattainable for straightforward maximizers. The first reason for that is that it is much harder for straightforward maximizers to engage in cooperative activities. Neither a constrained maximizer nor a straightforward maximizer is willing to cooperate with a straightforward maximizer due to the risk of exploitation such interaction affords. The second reason is that even if we introduce Axelrod's logic of cooperation for straightforward maximizers, the rationale to cooperate works only in the case of an iterated prisoner's dilemma. In Gauthier's view, cooperative benefits in a one-shot version of the prisoner's dilemma are always unattainable for straightforward maximizers, since not cooperating is, for them, always the rational option. We disagree with Gauthier's conclusion that cooperative benefits are unattainable for constrained maximizers in a one-shot prisoner's dilemma. The sanction-approach is also suited to solve the dilemma without any appeal to constrained maximization. Recognizing that the likely outcome is mutual defection from cooperation, individuals can agree to establish sanctions to alter the structure of the dilemma they face. The introduction of sanctions can turn cooperation into the rational choice for both individuals if it attaches severe penalties to uncooperative behavior. There is no reason to believe that such a method could not be used in a one-shot prisoner's dilemma as well.

Either in the sanction-based approach or the disposition-based approach, contractarians solve the problem of compliance by altering the structure of the problem. Those that favor the sanction approach introduce a mechanism to alter the profits and losses involved in the dilemma of

cooperation. The purpose is to connect defection from cooperation with penalties in order to diminish the benefits that individuals can expect to acquire by defecting. In this approach, compliance with moral rules becomes rational because one does not want to be penalized. The disposition approach, on the other hand, introduces an alteration of the model of rationality that informs individuals choices. It introduces the idea of constrained maximization, which is a different form of practical reason than pure utility-maximization. It alters the way in which individuals perceive the dilemma to demonstrate that compliance is rational for everybody. In any case, in the discussion that follows, the theory of compliance that we refer to is mainly the sanction-based approach in the line of Hobbes and Buchanan. In chapter four, we return to the theme of compliance and discuss it in more detail. But we can already make it clear that the discussion includes the state as part of the sanctioning mechanism. Such inclusion is one of the pillars in chapter five.

4 CONCLUSION OF THE CHAPTER

We may conclude the chapter now. We covered here the essential methodological elements that will inform our discussion throughout this chapter. These elements are given by the social contract tradition, and more specifically by the contractarian branch of the tradition. Contractarianism is a method of moral justification. It attempts to justify the normative demands of norms or principles to those for whom the norm or the principle is supposed to have normative force. The justification procedure is done by means of a thought-experiment. The philosopher elaborates a hypothetical scenario in which the normative entity under examination has no normative force over individuals. That is the state of nature scenario. The elaboration of the state of nature involves the suggestion of a model of practical rationality that informs how individuals act and a description of how individuals' interactions in the state of nature are. Given the hypothetical state of nature, the philosopher has to compare it with the moral state, i.e., the state in which individuals are under the authority of the normative entity under examination. If it is found that individuals are better off under the rule rather than without it, the reason for that is that in the state of nature individuals are in a suboptimal condition concerning the promotion of their interests. If it can be shown that such suboptimality affects everyone in the state of nature, i.e., that participation in an agreement to confer normative force to the principle or norm in question is rational for everybody, then the principle or norm can be justified. In contractarianism, the model of practical rationality involves a combination of the utility-maximization conception with the idea of fundamental general interests.
The conception serves to define what is to act rationally, whereas the other factor provides the substance to evaluate which actions are rational and which are not. In this sense, the contractarian method is about evaluating whether a norm helps individuals to overcome a situation of suboptimality with regards to the promotion of some fundamental general interest.

Connected with the theory of moral justification is, in the contractarian view, a theory of compliance. We touched the matter of compliance briefly in this chapter. We indicated two approaches to the problem of compliance suggested by the tradition, the sanction approach and the disposition approach. We indicated that the sanction approach will figure in our discussion. In any case, we return to the topic of compliance in chapter four. Compliance is, as it is argued ahead, intrinsically connected with the problem of stability, which, in its turn, constitutes one of the critical problems of our discussion here. In the next chapter, we examine the problem of distributive justice that contractarianism faces. We present why it is the case that contractarian doctrines have trouble to justify rules of distributive justice.
CHAPTER TWO: DISTRIBUTIVE JUSTICE

OVERVIEW OF THE CHAPTER

In the last chapter, we introduced the methodology that will underlie the discussion throughout this work, contractarianism. In this chapter, we present the second fundamental methodological aspect that pertains to our discussion in the next chapters, distributive justice. We present the idea of distributive justice, but the central goal of the chapter is to formulate the problem that we want to tackle in our work, which is the difficulty that contractarianism faces to justify principles of distributive justice that create substantive redistributive obligations. We start the chapter with a presentation of what distributive justice concerns. For this purpose, we refer to Aristotle's discussion of the matter and present his view of justice as proportionality. From Aristotle's concept, we suggest what we understand to be the problem of justice and then proceed to contextualize this problem from the contractarian perspective. The contextualization is performed by an analysis of Thomas Hobbes' and James M. Buchanan's views on the topic of distributive justice. The second part contains the presentation of a contractarian approach to distributive justice based on Buchanan's ideas. Although Buchanan himself has not explicitly advanced a conception of distributive justice, we argue that one of such conceptions entails from his contractarian view. The second section also contains a discussion of whether Buchanan's principle is sufficient to solve the problem of justice. This discussion leads to the third section, where we consider another contractarian discussion of distributive justice, namely Gauthier's discussion of the Lockean Proviso. The principle he suggests, the Lockean Proviso, is more substantive than the principle we find in Buchanan's discussion. The third section is dedicated to present and critically examine Gauthier's principle. The examination of Gauthier's principle exemplifies the problems that the Hobbesian method faces when approaching the matter of distributive justice.
1 THE PROBLEM OF JUSTICE

We start now with the presentation of the problem of justice. For this purpose, we first refer to Aristotle's discussion of justice, and from it, we conceptualize what we believe to be the central problem pertaining principles of distributive justice. When the problem is conceptualized, we move to contextualize the problem from a contractarian perspective. The contextualization is done with reference the work of two eminent contractarians, Hobbes and Buchanan.

1.1 DISTRIBUTIVE JUSTICE AND ITS PROBLEM

The idea of distributive justice has been in the philosophical tradition since Ancient Greece. Albeit its long history, the concept of distributive justice articulated by Aristotle is still the same that informs our intuitions and daily judgments about justice. The concept can be found in book V of the *Nicomachean Ethics*. There, Aristotle discusses the subject of justice as a whole. Aristotle's discussion in book V is, we must notice, carried out from a *formal* perspective. The theme of the discussion is not what is *substantially* or *materially* just or fair.\(^{74}\) Aristotle does not discuss which particular distributions of goods are just and which are not. Instead, his purpose is to examine the *form* that justice has. The direction of his investigation is to conceptualize what is it that we talk about when we speak about justice. More appropriately, Aristotle's discussion is about the *concept of justice*, and not about *conceptions* of justice.\(^{75}\)

Aristotle divides justice into two different kinds. There is, he argues, universal justice and particular justice.\(^{76}\) Universal justice consists of the disposition to comply with the law. Particular justice, on the other hand, refers to more specific subjects. Aristotle divides particular justice into two separate groups. Particular justice contains corrective justice and distributive justice. The subject of corrective justice is the rectification of wrongs committed in the past. Distributive justice,

---

\(^{74}\) Throughout our work, we refer to justice and fairness interchangeably.

\(^{75}\) See Rawls (1999), p. 9. In *A Theory of Justice*, Rawls comments on the difference between the concept and conceptions of justice. He argues that the concept of justice refers to the 'role of its principles in assigning rights and duties and in defining the appropriate division of social advantages.' A conception, on the other hand, is an interpretation of the concept. In his sense, whereas Aristotle's discussion concerns the concept of justice, Rawls' discussion concerns a conception of justice, namely justice as fairness.

\(^{76}\) See Aristotle (2004), p. 84, (Arist. EN, 1130b).
on the other hand, refers to how goods should be distributed between individuals. Universal justice and corrective justice are not crucial for our discussion in this work. For this reason, we do not further consider both concepts here but focus our examination on Aristotle's discussion of distributive justice.\footnote{From this point onward, whenever we refer to 'justice' we are speaking of 'distributive justice', unless otherwise indicated.}

Distributive justice is, in Aristotle's view, about how to distribute certain goods.\footnote{See Aristotle (2004), p. 85-87, (Arist. EN, 1131a,1131b).} The goods he has in mind are goods that have to be shared among members of a single political community. The subject of justice, then, is always a distributional pattern of goods. A pattern is just in Aristotle's view if it contains a particular form of equality. Equality is a factor that requires some form of identity between two different terms. A distributional pattern necessarily involves at least four terms. It must include at least two persons and two shares of a good, one share held by each person. The equality connected with justice is not, as one could expect, equality of shares. Instead, Aristotle argues that a just distribution is a distribution in which there is the same level of equality between persons as there is between their shares of the good. If the same level of equality appears between the persons' shares and themselves, then the proportion between the shares is equal to the proportion between the individuals. A just distribution of goods for two equal persons, then, is a distribution that leaves them with equal shares. On the other side, a just distribution of goods for two unequal persons is one that leaves them with unequal shares. In both cases, there is equality between the proportion to each individuals' share and the proportion between each individual himself.\footnote{It is from Aristotle's concept of justice that the saying 'One ought to treat equals equally and unequals unequally' springs from.}

Regarding the proportion between individuals, Aristotle claims that such proportion is that between each individuals' \textit{merit}. Merit is the factor through which persons can be considered equal or unequal to each other. A given distribution is just, then, if the proportion between each individuals' share is equal to the proportion between each individuals' merit. The concept of justice advanced by Aristotle defines justice as a \textit{proportion}, a \textit{ratio}.\footnote{See Polansky (2014), p. 161 and Young (2006), p. 185. Both discuss Aristotle's view of distributive justice, and his views of justice more generally considered, in some depth.} It is often referred to as \textit{proportional justice}. Justice is determined by the equality of two ratios, namely the ratio between each individuals' share and that between each individuals' merit.

A key factor, then, to determine whether a given distribution is just or unjust is the distribution of merit between persons. We must know each persons' merit in order to judge the justice of their shares. Aristotle, however, does not discuss what the standard that defines how merit is distributed among individuals is. He affirms that although everyone accepts that a just
distribution is that done in observance of merit, people have different views of what merit is. Aristotle mentions as an example the different views on the subject that democrats, oligarchs, and aristocrats have regarding the distribution of political offices. They all agree that the offices should belong to those with merit. Nonetheless, democrats believe that merit is defined by being a free citizen, oligarchs that wealth or nobility defines it, and aristocrats that virtue is the determining element of merit.\footnote{See Aristotle (2004), p. 86, (Arist. EN, 1131a).}

Merit is a fundamental aspect of Aristotle's concept of justice. It is the factor that determines how much of a good the individual is entitled to have. Merit is an entitlement to some parcel or share of some good. However, the word 'merit' seems to be out of place when we have in mind particular distributive cases. When we speak about the distribution of trophies and medals in a swimming competition, for example, the term 'merit' seems appropriate. However, when we have in mind a case of a distribution of food between a wealthy and a poor person, the word 'merit' seems semantically dislocated. Suppose that we accept that the fair distribution is that in which the largest share of the food is given to the poor person. We give the larger share to the poor because we believe that the poor person is entitled to the food. However, whereas we can safely speak of the person being entitled to the food because she is poor, it would be odd to affirm that the person merits the food because she is poor. We associate merit with achievement and effort involuntarily. The concept carries with itself a semantic load that connects it with the ideas of achievement and effort. In virtue of this semantic load, we prefer to abandon the word 'merit' in the context of distributive justice and to refer to 'entitlement' in its place. We believe that the change does not cause any consequences or alterations to Aristotle's concept of proportional justice. It only makes it more suitable to address a larger range of topics, such as the case of distributing food between rich and poor that we considered above. With the substitution of 'merit' for 'entitlement', a just distribution in Aristotle's sense becomes that distribution in which the ratio or proportion between the individuals' shares of a good is equal to the ratio between their entitlement to the good.

Now, regarding the matter of determining how entitlements themselves are distributed, Aristotle is right. We disagree about which standards of entitlement are appropriate to particular situations. In some cases, it is easier to see what constitutes entitlement. Consider the example of a swimming competition again. The case is about distributing medals to swimmers. In this case, entitlement is defined by the physical prowess to swim faster than all the other competitors over a certain distance in a pool and under some rules. The first to arrive is entitled to the gold medal, the second to the silver medal, and the third to the bronze medal. Any other distributional pattern is unjust because there is no equality between the proportion of medals distributed and the proportion
of each swimmer's entitlement to medals. Although in the case of the swimmers it is easy to see what entitlement is about, that is not the case always, as Aristotle observes. If we shift the object of distribution from medals in a competition to other goods such as, for example, income, wealth, food, political rights, and political obligations, it becomes harder to define what is the appropriate standard of entitlement. Consider the case of income. A stereotypical libertarian can argue that what defines entitlement to income is the capacity to acquire income in a free-market society. A Marxist certainly disagrees and would appeal to the Marxist principle 'From each according to his ability, to each according to this need'. Egalitarians, in their turn, would disagree with both the libertarian and the Marxist and argue that every person is equally entitled to income merely because she is human.

The problem of justice is the problem of finding a suitable standard of entitlement to substantiate Aristotle's concept of distributive justice. Aristotle's analysis of distributive justice helps us to understand that the core problem of distributive justice is not the principle by which we ought to divide goods. This principle is clear. The just distribution is that in which the proportion between the shares is equal to the proportion between entitlements to the good that individuals have. Justice is about proportional equality. What is unclear is the standard that defines entitlement. Due to the diversity of opinions on this topic, the conceptual problem that pertains to distributive justice is the determination and justification of a standard that defines entitlement. Our objective in this work is to perform this task.

There is an aspect of distributive justice that we want to stress, and that does not appear in Aristotle's discussion. That is the matter of redistribution. Justice is a moral concept, and, as a moral concept, it is connected with normativity. The claim that a given distribution is unjust is not a merely descriptive claim, but also a claim that involves a normative demand. In the same manner that the claim 'murder is morally wrong' involves the normative demand that one ought not to murder, the claim that 'the distribution is unjust' involves the demand that one ought either not to reproduce the unjust pattern or that one ought to redistribute goods to eliminate the injustice. Consider the case of the Marxist principle that we mentioned above. The principle determines that needs define entitlement to goods. If we observe a distribution in which the larger shares belong to individuals that already have their needs fulfilled, whereas deprived individuals have small shares, we may conclude, if we accept the principle, that such distribution is unjust. The injustice of the distribution creates a normative demand to undo such injustice. A moral obligation to redistribute resources in order to increase the share of the needy at the expenses of the non-needy emerges. One can discuss, of course, about who is the bearer of such an obligation. It may be that the wealthy are

obligated, or that the state is under such obligation, or even that a third party has such duty, or also
that all of them have the responsibility. This discussion, however, is a different discussion and does
not pertain our topic here. The sole point that we want to stress is that justice, as a moral concept, is
a concept that necessarily entails normative demands to perform specific actions. The normative
demands that justice creates are of two types. The first demand is to distribute goods justly and the
second demand is to correct unjust distributions so that they become just.

One last aspect of distributive justice must be considered before we proceed to see how
contractarianism tackles the problem of justice. That is the case that the standard that defines
individuals' entitlements changes when the goods whose distribution is in question also changes.
The standard that defines entitlement to medals in a swimming contest is different from the standard
that defines entitlement to slices of cakes in a birthday party. To approach the problem of justice,
one must first specify what the goods whose distribution one is investigating are. Our discussion is
focused on the broad category of goods that we shall call *scarce external goods*. These are external
and material goods that individuals require to fulfill their desires and life plans. Moreover, because
they are scarce, it follows that individuals are likely to conflict over them, since there is not enough
available for everyone. Individuals need, then, common rules to define how such goods should be
divided.

1.2 The Problem of Justice in the Contractarian Context

We move now to contextualize the problem of justice in light of contractarianism. However, before
doing so, there is a topic we want to address. That is the topic of whether we can import Aristotle's
concept of justice into a contractarian framework without causing problems for the framework or
requiring drastic alterations to the concept. Concerning this matter, we think that the contractarian
framework faces no difficulties to import Aristotle's concept. To see that, remember that Aristotle's
concept of distributive justice is merely formal. It is a concept to determine what the object of
justice is. It gives us no information about what are the actual requirements of justice. In this sense,
the actual normative demands of justice in Aristotle's sense are entirely dependent on the standard
of entitlement that informs it and the actual goods whose distribution the standard refers to. As a
method of moral justification, contractarianism does not conflict with the formal idea of
proportional justice. The concept itself cannot be a matter of justification, since the concept, by
itself, contains no actual normative demands for individuals. The discussion about justification is
only suitable when the concept includes a standard of entitlement and a distribution of goods. When loaded with both elements – the standard and a distribution of goods – justice starts to have normative requirements for individuals. Such requirements can be subjected to the contractarian analysis. We can discuss whether they can be justified from the contractarian perspective. Otherwise, Aristotle's concept of justice is but a concept to define what justice is about and what it is not about, and as a formal concept, it offers no problem to be incorporated into a contractarian framework. Having concluded that we may import Aristotle's concept into the Hobbesian view, the question that follows is: How should a contractarian approach the problem of justice? Two representatives of the contractarian tradition, Hobbes and Buchanan, offer us some guidance in this respect.

In *Leviathan*, Hobbes briefly discusses what is distributive justice. He rejects part of Aristotle's terminology on the matter. Hobbes argues that the distribution of equal benefit to persons of equal merit is not distributive justice as Aristotle affirms, but *equity* instead. Distributive justice, he argues, is the justice of an arbitrator. Merit – entitlement to goods – is, for Hobbes, awarded either by covenant – when a party fulfills his obligations in a covenant, this party merits the performance of the other party – or by the grace of an arbitrator. When awarded by covenant, merit is a matter of commutative justice, and thus it is not connected with distributive justice. When awarded by an arbitrator, it can only be, in Hobbes' view, improperly called distributive justice. The task of the arbitrator is exactly to define what is just and what is unjust.

Hobbes' comparison of distributive justice with the justice of an arbitrator is important. It indicates us how a contractarian must approach the problem of distributive justice. Hobbes notices that the authority of the arbitrator emerges from the *trust* that the parties involved in a distributive conflict have on him. It seems to be the case, then, that the normative authority of the arbitrator is reducible to the parties' trust in his judgment. In other words, the arbitrator can determine what is just for the parties because they want him to do so and they trust him to do so. Should it be the case that the parties have no trust in the arbitrator, then his judgment on what is just for them would have no binding normative force. If the authority of an arbitrator is established through two or more individuals' interests in solving a distributive conflict, it seems that the contractarian method of justification underlies the normative force of the arbitrator's authority. What Hobbes' discussion of distributive justice indicates is that the standard or criterion to define what persons' entitlements to goods, i.e., the answer to the problem of justice, can be defined by the thought-experiment of the agreement.

We believe, however, that the figure of the arbitrator is dispensable to a contractarian view.

---

of distributive justice in a certain sense. The arbitrator is superfluous as a source of the standard of entitlements. To supply the standard alone, individuals do not need an arbitrator. In the same manner that they constitute an arbitrator by reasons of interest and trust, they can convene on a standard under the same reasons, without any need of an arbitrator. To see why that is true, consider the actual role of the arbitrator. When the arbitrator decides how the distribution of goods should be performed, he is deciding it in observance to a standard of entitlement. If the parties are ready to accept the arbitrator's decision, that means that they are also prepared to accept the principle that he utilizes to define what a just distribution is. Hence, as a source of the standard of entitlement, the arbitrator is a superfluous figure. The standard itself can be the object of an agreement between individuals. However, the arbitrator is not unnecessary as a constraining mechanism that assures that the standard that both individuals accept will be employed. In this respect, the arbitrator has the same role that the *Leviathan* itself has regarding moral rules. We explore this role better in chapter four. For now, what is relevant to keep in mind is that the Leviathan, and the arbitrator in our case, is not a constitutive element of the justification process of a moral rule.

Hobbes' view gives us a decisive lead. The problem of justice is the problem of determining the standard that defines individuals' entitlements. With the contractarian method, we can address this problem in the following manner. Once Aristotle's concept is substantiated with a standard of entitlement, it will produce normative claims over individuals. If we substantiate it with Marx's principle, for example, it will produce an obligation of the talented and rich to transfer part of their goods to the poor. With the contractarian method, we can verify whether such normative demands can be justified by using the contractarian thought-experiment to evaluate whether the principle and its normative demands can be the object of a moral agreement. So if we find out that they can be justified, we may conclude that we found a justifiable standard of entitlement for the distribution of goods. When individuals accept the normative requirements that proportional justice produces when substantiated by a given standard, it seems clear that the standard itself is acceptable for them. From the individuals' acceptance, it follows that we justified a *conception of justice*. To overcome the problem of justice, then, the contractarian method must find a standard of entitlement whose normative requirements can be justified through the contract experiment.

James M. Buchanan offers a more recent case of a contractarian approach to the matter of distributive justice. The view is presented in a paper entitled *Rules for a Fair Game*. In this paper, Buchanan tackles some issues of distributive justice, and one of them is what justice in a contractarian viewpoint is. To define fairness, he examines what makes a game rule fair. His answer to the question is quite simple. A rule is fair if the players agreed with it before they started.

---

84 See Buchanan (1986), pp. 123-139.
the game. If the rule was put in place before the game started, i.e., before individuals acquired
determinate positions inside the game, then it is just. What defines the justice of a rule is agreement
over the normativity of the rule. He stresses that the fairness of a rule is not the reason why
individuals agree to it. Instead, it is because individuals consent with the rule that it is deemed to be
fair. Buchanan's view summarizes our reading of Hobbes' discussion of distributive justice. As
Hobbes seems to do, Buchanan also argues that distributive justice is a product of agreement.

Peter Stemmer has argued against Buchanan's contractarian approach to distributive
justice. Stemmer has argued that Buchanan's characterization that a rule or agreement is just
merely because individuals agree to it is a revisionary concept of justice. This view of fairness, so
argues Stemmer, is different from that suggested by Aristotle, which informs our traditional views
and intuitions about fairness. As Buchanan characterizes justice and fairness, both concepts seem to
have no connection with the proportional idea that Aristotle attributes to the formal concept of
justice. In this sense, the discussion of fairness that Buchanan conducts has nothing to do with what
we consider to be fairness and justice. For this reason, Stemmer argues that the use of the words
'just' or 'fair' to refer to a rule or distribution in Buchanan's sense is inappropriate. The revisionary
character of Buchanan's view of fairness can be stressed if we notice that his discussion is about fair
and just *rules*. Rules are not the object of justice in the Aristotelian concept of justice. Instead, the
central object is the distribution of some good. In this sense, justice and fairness are attributes of
distributions, and not of rules, if we remain loyal to Aristotle's concept.

Although we agree in part with Stemmer's critique against Buchanan, we believe that
Buchanan's general conclusion that justice is defined by the agreement is correct. Buchanan's
discussion seems indeed to consider a different subject from Aristotle's. The concepts of merit,
entitlement, distributional patterns, proportionality, which are all fundamental in Aristotle's idea of
distributive justice, do not figure in Buchanan's discussion. Instead of such concepts, Buchanan is
concerned with what he calls a 'fair rule'. As we mentioned above, fairness and justice are not
attributes of rules in the Aristotelian concept of justice. Regarding the object of Buchanan's
discussion, Stemmer is right. What Buchanan discusses is not what we traditionally understand to

---

86 One could, of course, attempt to connect the idea of a rule with that of a distribution in order to connect the idea of
a fair rule with that of a fair distribution. In this sense, a fair rule could be considered that which commands the
realization of a fair distribution, which is, in its turn defined by Aristotle's concept of justice. Suppose that we
accept the Marxist principle and that we are confronted with a distribution in which the talented hold the largest
shares in the system, whilst the needy have the smallest. Suppose also that we are presented with a rule that
commands us to increase the shares of the talented at the expenses of the needy. In light of the Marxist principle,
we can consider such a rule unfair. In this respect, we would be able to judge whether rules are fair or unfair by
examining the harmony between their normative requirements with those of a specific conception of justice. When
there is no conflict between them, the rule is fair. However, when they conflict, the rule is unfair. In such a sense, it
would be possible to speak about fair and unfair rules.
be distributive justice. Nonetheless, despite the rectitude of Stemmer's critique, we still think that Buchanan's conclusion that justice is defined by the agreement, is correct. As we argued at the start of this section, the Aristotelian view of proportional justice can be incorporated into contractarianism without any problem for the concept of justice or contractarianism. A contractarian has no reason to reject from the outset the idea that the justice of a distribution consists of the equality of the ratio between individuals' shares and the ratio between individuals' entitlements. What is problematic for contractarianism to incorporate – or justify – are standards of entitlement. Concerning the justification of such standards, contractarianism must perform it through the thought-experiment of the moral agreement. If rational agents are ready to take part in an agreement whose objects are the normative demands of proportional justice informed by a particular standard of entitlement, the whole package of justice – the proportional view, the standard, and the normative consequences of the conjunction of both – can be incorporated to contractarianism. Despite the revisionary character of Buchanan's discussion, he is right in affirming, then, that agreement defines justice.

We may now summarize what the contractarian task in the context of justice is. The problem of justice is the problem of finding and justifying a standard of entitlement to govern the distribution of goods. A contractarian can approach this problem by investigating whether the normative demands that follow from a given standard and thus from justice itself can be justified through the contractarian method. To do so amounts to justify a conception of justice. The conception of justice, in the contractarian perspective, can be selected through an analysis of whether the normative claims that its standard if entitlement produces can be justified. If the justification can be done, then the problem of justice can be tackled.

2 **Contractarian Distributive Justice**

We come now to examine a contractarian conception of justice. That is the conception of *justice as natural liberty*.87 Such a conception is based on a standard of entitlement that we can induce from the contractarian justification of private property. In this section, we examine James Buchanan's justification of private property and suggest that a standard of entitlement, which we call the principle of natural liberty, is embedded in the discussion. JNL is the conception of justice with which we occupy ourselves for most of our work. We argue that although such a conception can be

87 From now onwards, we refer to justice as natural liberty as JNL.
justified by the contractarian method, we have reasons to refuse it in order to adopt another conception of justice that involves two other standards of entitlement.

2.1 Private Property and Distributive Justice

As indicated before, our concern here is with the distribution of external goods. The problem of justice we have before us is, then, that of determining the standard of entitlement for the distribution of such goods. In the contractarian context, the question over the standard of entitlement for external goods is intrinsically connected with the justification of one of the most basic moral rules: private property. An agreement on private property is an agreement to solve the following question: How are goods to be distributed in society? This question has two different aspects. One aspect concerns how individuals can acquire, exchange, and transfer goods. This matter involves the justification of the institutions of acquisition, contract, donation, loan, promise, and so on. The second matter is more abstract and fundamental. It concerns how we should divide property between individuals. It is about which distributions of goods are morally acceptable – just – and which are not. To see how this question is intrinsically connected with that of the justification of private property, consider the example below.

Consider a group in individuals in the state of nature. One of the elements of the natural condition is that there is between individuals an intense and ruthless competition for the goods available. In the state of nature, there is no private property, but mere possession of goods. If an individual desires a good under the possession of another, and the individual can acquire the good either by sheer power or by cunning, he will take the good for himself. Since individuals are in a state of nature, such form of acquisition cannot be considered immoral nor unjust. For this reason, possession of goods is highly insecure and unstable for every individual in this state. It is reasonable to assume, then, that individuals would find it rational to establish a moral rule to assure the stability of possessions, i.e., a moral rule to establish private property. The cessation of the instability of possession is, we can reasonably expect, beneficial for everyone involved. Participation in an agreement to establish private property, then, is rational for everyone. However, to establish this agreement, a fundamental question must be addressed. The agreement transfers individuals from the state of nature to the moral state. With the establishment of private property, the goods that individuals had possession over become their property. In this sense, private property is created on top of a distributional baseline of goods, in which each's share is determined by the goods that the individual managed to keep under his possession in the state of nature. Before
closing the moral agreement, persons must specify which is the distributional baseline for moral cooperation. A moral agreement for the establishment of private property can only be concluded if this question is answered first. If the question is left unanswered, individuals have no distributive basis of external goods upon which they can establish the institution of property. In sum, the problem of justice must be necessarily solved before the establishment of private property.

The discussion that determines the baseline is an important discussion. From each individuals' perspective, the best possible baseline is one in which he possesses all of the goods available while the others have nothing. However, if each presses for the best possible baseline for oneself, a moral agreement will be irrational for everyone, except the owner of all the goods. It is only with a baseline that elicits the acceptance of everyone that they can abandon the state of nature. A solution to the problem of selecting the baseline involves necessarily a suggestion of a conception of justice. The solution determines how goods should be distributed in the moral order.

One should not, however, mistake what is at stake in the solution of this problem. The contractarian approach involves the thought-experiment of the state of nature, but one should not lose oneself in the metaphors of the state of nature and the moral agreement. Although the discussion is focused on hypothetical individuals in a hypothetical scenario discussing hypothetical goods, the conception of justice that such thought-experiment entails is a conception that is supposed to have a use for us in the real world. A conception of justice is, first of all, a normative conception in Derek Parfit's rule-involving conception, which we examined in the first chapter. We can use it to evaluate actual distributions of goods morally. It works as a standard for assessing whether a given distribution is just or unjust. It follows from its normativity that we must act in certain manners given the justice or injustice of the distribution in which we are involved. By investigating the standard that individuals in a state of nature would accept to determine a division of goods among them, we are investigating a standard to judge the actual distributions of goods in our world.

Ahead, we consider Buchanan's discussion of a moral agreement about private property. Buchanan does not touch the subject of distributive justice in the discussion, but we believe that a conception of distributive justice can be extracted from Buchanan's discussion nonetheless. We first present Buchanan's view on the emergence of private property and then proceed to consider a conception of distributive justice that we believe to emerge from his view.
2.2 A CONTRACTARIAN CONCEPTION OF JUSTICE: JUSTICE AS NATURAL LIBERTY

Although we followed Stemmer's critique against Buchanan's view of fairness, Buchanan's discussion of private property involves an answer to the problem of justice regarding the distribution of goods. In *The Limits of Liberty* (1975), Buchanan discusses, among other topics, the conditions that make a moral agreement on private property – or constitutional agreement, in his terminology – rational for utility-maximizers. The starting point for his discussion is the matter of the distributional baseline of goods that must be in place in the social system in order for the emergence of a rationale to participate in the constitutional contract.

Buchanan differentiates two types of distributions, the original distribution and the natural distribution. The original distribution is the distribution of goods in place before individuals start to interact with one another in the state of nature. In Buchanan's discussion, it is represented by a distribution established by goods 'falling down' on individuals. The natural distribution, on the other hand, is the distribution that results from individuals interacting with each other in the natural state. Buchanan describes the movement from an initial to a natural distribution in the following manner.\(^88\)

Consider a two-person-world where every good except a good 'x' is superabundant. The good 'x' is the only good that is scarce in this world. This good is assumed to fall down in fixed proportions over each individual, and the consumption of 'x' requires no production at all. We also assume that both individuals desire to have as much of the good as they can. Given these settings, two consequences follow. Firstly, individuals cannot gain any surplus of utility by trading. Since every good but the good 'x' is superabundant, there is nothing they may offer to each other but units of 'x' in exchange for units of 'x'. Secondly, every time an individual consumes a unit of 'x', the consumption causes a loss of utility for the other person since a unit of 'x' is removed from the system. Since both individuals desire to consume as much of 'x' as possible, they will engage in a competition to protect their shares of 'x' and also to augment it at the expenses of the other individual. The natural distribution is a distribution that results from this competition for larger shares of 'x'. This distribution is established when the shares of 'x' that each individual has correspond to the individuals' capacity of acquiring and protecting units of 'x'. The natural distribution is the distribution in which each share of 'x' perfectly matches the power of acquiring

---

\(^{88}\) Although the idea of natural distribution has a key role in Buchanan's discussion of private property, he is not the original proponent of the concept. The concept is first found in a paper published by Winston Bush. See Bush (1972).
and defending shares of 'x' that each has.\textsuperscript{89}

With the installment of the natural distribution, further investment of effort in attack to increase one's share becomes irrational for every person.\textsuperscript{90} When the initial distribution was still in place, each individuals' share was such that they could invest efforts to augment their share at the expenses of the other individual. 'Effort' is considered to be a bad, a loss of utility. However, even though 'effort' causes a loss of utility when invested to increase one's share of 'x', the increase gives the individual some utility that overweights the loss of utility caused by the investment. In this sense, it is rational for the individual to remain in conflict with the other in order to augment his share of 'x' insofar as the overall result of his attempts of enlarging his share are positive with regard to his expectation of utility. However, as the conflict progresses, the distribution shifts following each individual's capacity to increase and protect his own share. At some point in time, each individuals' share corresponds exactly to that individual's power of acquiring and defending the share he has. If the share does not correspond to his power, two cases might occur. The first is that the individual has a smaller share than his power allows him to defend, which means that he can still increase his share of 'x'. The second case is that the individual has a larger share than his power allows him to defend. In this case, he is not powerful enough to prevent others from taking units of 'x' from him. In the natural distribution, every individual has exactly the share that his power allows him to have. In such distribution, the cost of investing efforts to acquire a larger share of 'x' either equals or outweighs the benefit that the individual expects to reap by increasing his share of 'x'. When individuals find themselves in the natural distribution, investing more effort to increase one's share is irrational because the individual is either unable to increase his share or because he is unable to protect and assure the increase in the share that he will gain. The reason for that is that the in a natural distribution each individuals' share perfectly reflects that individuals' capacity for acquiring and protecting his share of 'x'. In sum, the natural distribution is the distribution of resources in which the power of each individual determines that individual's share.

Buchanan argues that the natural distribution is the distributional baseline for the constitutional agreement in virtue of its stability.\textsuperscript{91} Its stability arises from the fact that no individual has an incentive to attempt to change it. Further attempts of increasing one's share through power are irrational for every person in the system. Since the goods are distributed accordingly with each individual's power, each has the exact share of goods that he can defend from the others. In such distribution, investment of effort in attack becomes irrational because attacking now requires an investment of effort that causes a higher loss of utility than the increase of utility that the

\textsuperscript{89} See Buchanan (1975), pp. 31-33.
\textsuperscript{90} See Buchanan (1975), pp. 33-35.
enlargement of one's share would afford. To invest efforts in defense, however, remains rational for everyone. Cessation of such investment translates itself immediately in an incentive for other individuals to return their investments in attacking. However, the investment in defense, in a natural distribution, becomes wasteful for everyone. Every individual knows that as long as everyone maintains the investment, the distributional pattern will not change. They could, then, agree to maintain the pattern but collectively agree on some rules to make the investment in defense unnecessary, and rid themselves from the utility loss that the investment in effort for defense causes. It is the rationale for a collective cessation of the investment in defense that, in Buchanan's view, drives individuals toward an agreement on private property.

It is important to stress that the natural distribution is a formal concept. It does not refer to a particular distributional scheme. Any distribution of goods can be a natural distribution if it is a stable distribution regarding what we have argued above. It must be a distribution in which further investments in attack are irrational for everyone and, consequently, that taking part in a constitutional agreement is rational for everyone. There is no intrinsic connection between the natural distribution and, for example, equality of shares. What determines the shares are individuals' capacities of acquiring and maintaining under their possession units of the good. When such capacity has shaped the distribution of shares entirely, the natural distribution is installed. What determines how a natural distribution looks like are the dynamics of power in the social scheme. A natural distribution can be an egalitarian distribution if we assume, for example, that power is somewhat equally distributed in the social system, and thus individuals will have somewhat equal shares of goods in the natural distribution. However, if individuals have different capacities, i.e., if one individual is better in acquiring and holding units of the good than the other, then this individual will have a larger share in the natural distribution than the others.

The constitutional agreement, then, emerges from distributional baseline based on power, the natural distribution. When the natural distribution is reached, then a constitutional agreement can be justified for every person in the system. That is Buchanan's view of the justification of private property. We proceed now to examine what is the principle of distributive justice embedded in Buchanan’s view.

The agreement on private property is the agreement that determines how goods are supposed to be distributed in the social scheme. The first important aspect to be stressed is that whichever the manner that the goods are going to be distributed, the transfers involved in the distribution must respect some constraints and restrictions. Transfers ought not to involve, for example, theft, robbery, or fraud. Now, there are two factors that we can extract from Buchanan's justification of private property that can serve as standards of entitlements, interest and power. In Buchanan's, and also
Hobbes', view of the natural state, two conditions must be fulfilled for an individual to acquire a certain good. The first condition is that the individual must be interested in the good. He must want to have it under his possession. The second condition is that the individual must be powerful enough to acquire it. If the good is not under his possession, then he must do something to bring it under his possession. Whether he can successfully acquire the good depends on whether the individual has enough power to do so. The natural distribution emerges when each individuals' share corresponds exactly to that individuals' power to acquire the goods in which he is interested. Since the natural distribution is the baseline accepted for the constitutional agreement, the criteria that determine the distribution of goods, which individuals accept, are interest and power.

We could argue, then, that embedded in Buchanan's discussion is a standard of entitlement that defines entitlement through interest and power. Under this standard, an individual is entitled to a good if he, firstly, is interested in having the good, and, secondly, is powerful enough to acquire it. One should not forget, however, that transfers inside a moral framework, as we noticed before, must obey certain restrictions. Individuals cannot acquire goods by, for example, using violence against other persons. There are means which the rules of private property forbid individuals to employ to obtain goods they desire. For this reason, the factors of interest and power do not determine absolutely the entitlement to goods. The restrictions prohibit that certain forms of power, such as coercion, for example, from being a determining factor to entitlements. 'Power' is a vast concept. In its usage in contractarian theories, it often refers simply to the capacity for an individual to fulfill his interests or of acquiring a certain good he desires. It is not necessarily connected with coercion, violence, or exploitation. However, in the same manner that 'merit' carries with itself a semantic load that may cause confusion when the concept is applied to some circumstances, the word 'power' may also lead to confusion. 92 The case we are discussing is one of such cases. If we argue that individuals agree that power should be one of the determining factors of the standard of entitlement, one can understand us as saying that the capacity for coercion, for example, is an important element of the standard, what is false, since coercion is generally forbidden by private property. To avoid such confusion, we prefer to abandon the word 'power' and speak of talents or skills instead. A talent is an innate capacity an individual has for efficiently and successfully executing a particular task. A skill, on the other hand, is a learned capacity an individual also has of efficiently and successfully performing a particular task. A talented or skilled individual, then, is someone who is particularly good at performing some task. In our analysis, talent and skill both refer to the capacity for acquiring goods.

92 The concept of power is fundamental for our whole discussion in this work. In the next chapter and in chapter five we discuss in some length the concept of power. Such deeper consideration of the concept, however, is not relevant for our discussion in the current chapter. Right now, a mere intuitive idea of power is sufficient for the discussion.
The standard of entitlement is, then, determined by two factors. On the one side, talents and skills to acquire goods. On the other hand, by the restrictions on transfers imposed by the agreement of private property. An individual is then entitled to a good if he was talented or skillful enough to acquire it, and the acquisition was made in observance of the rules of property. The standard of entitlement can be considered, then, to be defined by a principle such as: *To each according to his talent and skill.* In this view, then, a distribution is just if it emerges under the constraints established by private property. The share that each individual has or is supposed to have is determined by that individuals' capacity of acquiring the share. Entitlement is defined by the individual's talents and skills, his capacity for acquiring goods. The system that follows the establishment of the constitutional agreement, and thus the system that incorporates the principle we suggested above, is what John Rawls has called the *system of natural liberty.* In this system, goods, careers, and public offices are open to the individuals' talents and skills. What determines the allocation of a good is simply the individuals' capacity to acquire it in conformity with the rules of private property. For this reason, we may name the conception of justice that emerges from Buchanan's view as *Justice as Natural Liberty* (JNL).

It is important to make a distinction here. We must distinguish between the conception of justice, JNL, and the standard of entitlement that informs such conception, what we may call the *principle of natural liberty.* The principle of natural liberty determines that an individual is entitled to a good when he has acquired such good in conformity with the rules of private property. The conception of justice, on the other hand, serves us to distinguish a just distribution from an unjust distribution. JNL is the outcome of a conjunction of Aristotle's concept of justice with the principle of natural liberty. It is a conception of justice that defines as just the distribution of goods in which the proportion between each share of the good is equal to the proportion of individuals' entitlements, which are set by the principle of natural liberty. As a consequence, JNL states that a distribution is just when every share in the system was obtained observing the rules of private property.

93 The principle we suggest defines entitlement by means of the individuals' capacity of acquiring goods under the rules of private property. The paradigmatic case for such notion of entitlement is that acquisition by labor. If an individual labors to produce or acquire a given good, and the acquisition was in accordance with the rules of private property, he is entitled to the good. However, an objector could argue here that the principle does not account for the role of luck in acquisitions. Consider an individual that completely lacks talents and skills, but that one day, wondering through the woods, finds a chest full of gold. Is such an individual entitled to the chest of gold? Provided the acquisition of the chest does not violate the rules of private property, the conception of justice we suggest would endorse the lucky individual's acquisition of the chest. The objector could argue that he did not expend any effort to find the chest nor was he actively looking for it in the woods. His acquisition of the chest is a mere lucky event and that mere luck is not sufficient to create entitlement. To discuss such topics regarding luck and acquisition would drive us too far from our discussion. But, in short, we do not see any problem in including 'luck' as one of the factors that entitles individuals to goods. In this sense, the principle we suggest here could be understood as: *To each according to his talent, skill, and luck.*

We can achieve a better grasp of JNL if we compare it with Robert Nozick's Libertarian view of distributive justice. JNL is very similar to the view of justice espoused by Nozick in *Anarchy State and Utopia* (1974), which he calls the *entitlement theory*. Nozick argues that a distribution is just if everyone involved in the distribution is entitled to the holdings they possess. Entitlement, in its turn, is defined by three principles of justice, the principle of *acquisition* of holdings, the principle of *transfer* of holdings, and the principle of *rectification* of violations of the two previous principles. The principle of acquisition of holdings determines how individuals may acquire holdings that did not belong to anyone previously. It is a principle that defines the terms for the *original acquisition* of holdings. The principle of transfer of holdings, in its turn, refers to how individuals may transfer and exchange their holdings. Lastly, the principle of rectification is about how violations of the principles of acquisition and transfer must be addressed. In Nozick's view, an individual is entitled to his holdings if he obtained the holdings under the three principles. A distribution of holdings, in its turn, can be considered just if every individual involved in it is entitled to the holdings he possesses.

Nozick's three principles are, in short, the principles that define the institution of private property. They are principles that determine how goods ought to be acquired and transferred, and also how violations of lawfully acquisitions of goods must be addressed. These are the principles that define entitlement, i.e., an individual is entitled to the goods he has if he acquired them in observance of the rules of private property. A just distribution of goods is that in which the goods that every individual possesses have been obtained following the rules of private property. JNL is very close to Nozick's conception of justice. Both views refer to private property to differentiate just from unjust distributions. The conceptions differ, nonetheless, when it comes to their foundations. The normativity of Nozick's entitlement theory refers to an idea of natural rights, whereas JNL is sustained by the Hobbesian method.

Stemmer argues that the natural distribution is the baseline that individuals would agree upon to start moral cooperation. He also maintains that power is the fundamental aspect that constitutes the rationale for moral cooperation regarding private property. However, Stemmer thinks that the standard of power is not a proper standard to define what is a just distribution. What Stemmer calls the standard of power is a standard based on Buchanan's natural distribution.

---

95 Contractarianism and Libertarianism are theories close to each other. Some philosophers have engaged in explicit attempts to ground libertarian views through contractarianism. Jan Narveson's *The Libertarian Idea* contains what is perhaps the most well-known example of such an attempt. See Narveson (2001), ch. 14, in particular. Fabian Wendt discusses in some detail the use of Hobbesian contractualism, the branch of the social contract tradition that informs our discussion here, to ground libertarian political views. See Wendt (2009), pp. 216-242, for such a discussion.


Stemmer believes that the standard is only rationally, or prudentially, correct since it enables the creation of private property. However, such a standard is not morally correct and thus cannot serve to substantiate a conception of justice. He affirms that a distribution is only just when two criteria are fulfilled. The first is that the distribution of goods corresponds proportionally to the entitlements of individuals to the goods as defined by a standard of entitlement. The second is that such a standard must itself be the morally correct standard. The standard of power does not fulfill the second criterion in Stemmer's view.

We believe that Stemmer fails to notice the subtle distinction between the standard of power and the standard of natural liberty. If we read him correctly, Stemmer critique targets the importation of the central aspect of the natural distribution into the moral order as a standard of entitlement. Such importation would result in a so-called 'standard of power'. Such a standard would state that an individual is entitled to a good if he is powerful enough to acquire it. The means employed to obtain the good are irrelevant for the so-called standard of power, as it is the case in the natural distribution. The sole factor that would count to define entitlement would be the individuals' capacity to acquire a good, either utilizing coercive power, or one's talents, or one's skills. It is hard to see how such a standard could be justified by the contractarian method. To define entitlement to goods solely in terms of power necessarily implies rejecting any sort of moral rule that constrains individuals' capacity to acquire the goods, such as, for example, the prohibition of theft. What the standard requires is, in essence, that individuals remain in the state of nature, without any sort of framework to constrain their interactions. The standard or principle of natural liberty that we suggest from Buchanan's idea of natural distribution is not the standard of power that Stemmer has in mind. The principle of natural liberty involves constraints to the individuals' powers. With it, power remains somewhat present in the form of talents and skills, as we argued. However, the whole coercive and violent aspect of power is not contemplated by the standard. Under the principle of natural liberty, entitlement is only constituted in observance to the rules of private property. The restrictions on acquisition that the principle of natural liberty imposes on individuals ensure that it can be justified for everyone, as we argued above. The standard of power, on the other, cannot be justified, since it offers individuals no assistance to overcome the state of nature.

JNL is a conception of justice that can be justified by the contractarian method. The conception is a result of a conjunction of Aristotle's concept of distributive justice with what we have called the principle of natural liberty. The principle of natural liberty is a standard of entitlement. It determines what entitlement is. For the principle, individuals are entitled to goods if

---

two conditions are fulfilled. The first condition is that the individual must acquire the good in question, and the second condition is that the act of acquisition must have been performed in observance to the rules of private property. When both conditions are fulfilled, the individual is entitled to the good. When we fuse this standard of entitlement with the idea of proportional justice, the result is JNL. For JNL, the distribution of a particular good is just if the proportion between each individuals' share is equal to the proportion between each individuals' entitlement to the good, entitlement here being defined by the principle of natural liberty. In shorter terms, a distribution is just if the share of each individual in the system is a product of an acquisition that did not violate the rules of private property.

2.3 Is Justice as Natural Liberty Sufficient?

JNL is the conception of justice with which we occupy ourselves for most of our work. It is not, however, the conception of justice that we endorse. JNL is indeed justifiable by the contractarian method. Nonetheless, as we argue ahead and in future chapters, it involves problems and difficulties that give us good reasons to abandon it.

A problem from which JNL suffers can be quickly observed. JNL is hugely economic when we come to the redistributive requirements it produces. Since what entitles individuals to shares of goods they have is their capacity to acquire the share under the rules of private property, the mere possession of the share is already an indicator that the distribution is just. Redistributive requirements, under this criterion, can only appear in case of violations of the rules of private property during the acquisition process. In this sense, if someone stole a good from another person, the thief stands under the requirement of returning the good to that person. Apart from such requirements, it is hard to conceive how someone might be obliged to give away resources to other persons. The minimalism of JNL is a representation of a particular difficulty that contractarianism has. It is a difficulty of justifying moral rules of beneficence, i.e., moral rules that command the practice of actions from which no immediate benefit for the agent is produced. Malcolm Murray, for example, notes that some consider that the failure of moral contractualism of accounting for a moral principle of beneficence would amount to a *reductio ad absurdum* against contractarianism.

As we considered in chapter one, contractarian justification is dependent on an idea of

---

99 One could even argue that such redistributions are but a part of corrective justice, and not distributive justice. Such comment would entail the consequence that JNL has no redistributive requirements whatsoever.

100 See Murray (2017), p. 200. Murray discusses the difficulty that contractarianism has to approach moral rules of beneficence, see Murray (2017), pp. 19-23, and ch. 10.
reciprocity. It is rational to engage in moral cooperation because moral cooperation is advantageous for every party involved. For this reason, it is not that difficult to consider how a moral rule that prohibits murder or theft can be justified. No person wants to be murdered or robbed, yet every person is capable of murdering and robbing other persons. Consequently, it is rational for them, then, to agree not to murder or rob each other. The argumentative structure of the justification is to appeal first to a fundamental general interest that the parties share – to not be murdered or robbed – and then demonstrate that the effective promotion of the interest requires moral cooperation. If the demonstration is successful, then one can safely argue that the parties will find it rational to engage in moral cooperation. However, such an argumentative structure cannot be applied to rules of beneficence.

Beneficence demands the practice of an action on behalf of another individual without expecting any immediate reward for the action. Cases of beneficence do not have the element of reciprocity that contractarianism requires to justify moral rules. A moral rule of beneficence produces advantages only for one of the parties involved. Consider the case of charity, for example. To give alms to a poor person is an act of charity. The act certainly produces advantages for the poor, since it alleviates the harming effects that poverty has for him. However, the act affords no advantage for the charitable person. The act can even be described as disadvantageous for her, since it reduces the person's overall level of wealth. In the act of beneficence, there is no reciprocity involved, no expectation of mutual benefit. As a consequence, it becomes tough to see how the model of practical rationality that contractarians use – utility-maximization or the moral skeptic – can allow the endorsement of an act of beneficence as rational without appealing to altruistic preferences, as contractarians refuse to do.

JNL is a typical contractarian rule in the sense that it does not demand the performance of any action that could be described as beneficent. Consider, for example, the case of justice and poverty. Poverty is not contemplated in the principle of natural liberty nor in JNL. To be poor does not create any form of entitlement to goods. One could argue to also include poverty as a standard of entitlement, but it is hard to see how that can be done from the contractarian perspective.

We can briefly consider the case of Engels and Marx's principle again to see the hardship that contractarianism faces in order to justify moral rules of beneficence. The principle requires the wealthy – the talented – to transfer part of their resources to the poor – the needy. To justify such a principle, we must demonstrate that the wealthy and the poor would find it rational to conform their

---

101 Another fundamental aspect of the justification procedure involves the dynamics of power between the parties involved. We do not discuss this aspect now, but instead in the next chapter.

102 We can assume here that a reduction in the level of wealth is prejudicial because it implies a reduction in the person's capacity of pursuing her own interests, since she has less wealth after performing charity than she did before it.
behavior to the demands of the principle. We must show that the principle could be the object of a moral agreement between the wealthy and the poor. The poor would accept the principle. A direct consequence of the principle is that the poor will have their shares augmented, what makes them better off. The wealthy, on the other hand, are made worse off by the principle. Their shares will be reduced because of the transfers they will be obliged to perform to the poor. The reduction, then, already indicates that the wealthy will refuse the principle since it would leave them worse off than they would be without it. In order to justify the principle, the contractarian must point out that the principle would make the wealthy somehow better off. Moreover, the contractarian must do so without appealing to altruistic preferences. That is the hardship that Hobbesian contractualism faces to justify principles of distributive more substantive than JNL.

There is another important aspect of JNL that we must stress. The contractarian difficulty to handle beneficence that appears in JNL’s lack of substantive redistributive requirements has two faces. The first is that individuals stand under no obligation of redistributing their goods to other persons. The second is that no one can interfere with another's goods unless the interference is done in observance of the rules of private property. JNL creates a strong right of ownership over goods. Such characteristic of JNL is, we believe, problematic. The reason for that was appointed by James Buchanan and Gordon Tullock in *The Calculus of Consent*. Buchanan and Tullock suggest that there is decisive empirical evidence for the presence of some collectivization of redistribution of income in almost every society. If JNL’s strong ownership rights cannot be found in reality, then the conception must be revised in some sense to be adequate to our empirical reality. Otherwise, the conception remains an abstract idea without much practical relevance.

The central goal of our work is to overcome the difficulty that contractarianism faces to justify rules of beneficence, but with regards to rules of distributive justice. We argue that a conception of distributive justice with more substantive redistributive requirements than JNL can be justified through the contractarian method. In the next chapters, we start the presentation of our argument for such a conception. However, before starting our presentation, we consider one ambitious contractarian attempt to justify a principle of distributive justice. That is David Gauthier's argument for a Lockean proviso. We believe that Gauthier's argument fails in the end. Nonetheless, considering his argument serves to stress the difficulty that contractarianism has to justify substantive principles of distributive justice.

---

103 See Buchanan & Tullock (2004), p. 185.
3 DAVID GAUTHIER'S LOCKEAN PROVISO

We consider now Gauthier's argument for a Lockean proviso in the state of nature. The proviso is a principle of distributive justice, and Gauthier argues for it from a contractarian perspective. This section has two parts. We present Gauthier's proviso first. The focus of the presentation is more the manner in which Gauthier justifies it rather than a detailed exposition of the content of the proviso. After presenting the proviso, we move to the second part, where we consider Gauthier's arguments critically.

3.1 GAUTHIER'S ARGUMENT FOR THE PROVISO

The proviso is an essential part of Gauthier's contractarian endeavor performed in *Morals by Agreement*. Right at the start of the book, Gauthier lists the proviso along four other conceptions that he deems to be central to the theory developed in the book. The importance of the proviso, Gauthier argues, is that its observance is a necessary precondition for the moral agreement. Without the proviso, rational utility-maximizers would find moral cooperation irrational.

The proviso that Gauthier suggests is a *Lockean proviso*. The Lockean proviso is an interpretation done by Robert Nozick of an argument advanced by John Locke. In a discussion regarding the appropriation of land, Locke argues that the appropriation of an external good must be done following specific rules. One of such rules is that the appropriation must leave enough and as good left of the good appropriated for other persons. The act of appropriation must be such that it does not constitute an injury to another, who becomes unable to enjoy the good appropriated.

Nozick, in his turn, resums Locke's discussion and advances an interpretation of Locke's argument. He formulates the proviso as a restriction on the acquisition of previously unowned goods. The proviso states that an acquisition will not be translated into private property if it worsens the position of those who become unable to use the good acquired.

The constraint that Gauthier's proviso introduces is, in short, that individuals ought not to exploit one another in the state of nature. To define exploitation, he adopts Nozick's interpretation of the proviso with a slight alteration. The proviso forbids individuals of improving their situation by worsening that of another. The alteration that Gauthier introduces is an exception to the

---

104 See Gauthier (1986), p. 16.
prohibition. In Gauthier's view, one is allowed to worsen the situation of another if doing so is necessary to avoid worsening one's own situation. What the proviso requires, then, is that individuals ought not to worsen the situation of another person, except to avoid worsening their own situation. The standard that Gauthier suggests to determine whether a given action or omission worsens the situation of another person is a comparison between what the individual did and would have happened in the absence of the individual.

Gauthier incorporates Nozick's interpretation of the proviso in his contractarian theory with the alteration we considered above. In his own framework, the proviso applies to what he calls the initial bargaining position. Gauthier conceives the moral agreement as the result of a bargain to divide the benefits that emerge from moral cooperation, or, as he calls it, the cooperative surplus. The initial bargaining position defines each individuals' endowments (or utilities) at the start of the bargain. It serves to identify the utilities or shares of each individual that are not part of the cooperative surplus. The proviso sets some constraints on what individuals may bring to the bargaining table. By imposing such restriction, the proviso is restricting what individuals may do in the state of nature. What individuals have in the initial bargaining position, what they bring to the bargaining table, are the fruits of their interactions with each other in the state of nature. The constraints of the proviso, then, are constraints on how individuals may interact in the state of nature.

In Gauthier's view, the proviso moralizes the state of nature. The proviso shifts the position of individuals from a Hobbesian state of nature to a Locke's state of nature. The shift is performed by the introduction of a rudimentary structure of rights and obligations in the state of nature through the proviso. The restriction imposed by the proviso can be translated as a system of rights and obligations whose normativity is sustained by the proviso's command of 'Thou shall not worsen the situation of the other except if you must do so to avoid worsening your own'. The prohibition of worsening the situation of an individual can be interpreted as a right that this individual has. If Eve has an apple on her hand, and Adam takes the apple from her, Adam's action

109 Consider two cases. In the first case, John and James are having dinner, and someone poisons John's food. When John starts to suffocate because of the poison, the shock of seeing his friend choking paralyzes James, who is unable to help John. The consequence is that John dies. In the second case, James poisons John's food, and John dies since James does not help him when he starts to suffocate. To consider whether James has worsened the situation of John we must see how things would have gone should James not be present in the dinner. In the first case, John would still have died. However, in the second, he would have survived. Given this difference, Gauthier thinks that only the second case constitutes a violation of the proviso.
111 The idea is that Hobbes' state of nature of widespread conflict and strife becomes Locke's state of nature of people living a communal life under certain rights. We considered both views in chapter one.
is worsening Eve's situation. If Adam observes the proviso, he does not take the apple from Eve, or, in other words, Adam accepts that he ought not to take away Eve's apple. The obligation to refrain from taking Eve's apple can be understood as a rustic form of right that Eve has over the apple and against Adam. On the other side, if an individual wants to perform an action that does not worsen the situation of any person, then there are no restrictions to this action. We can interpret this liberation, then, as a liberty, or a right to perform such action, i.e., that one ought not to hinder the individual from performing the action he desires. If Adam is walking around a forest and comes across a previously unknown apple tree full of apples, the act of taking an apple for himself worsens no one, since there are lots of other apples in the tree. In this case, there is no prohibition of acquiring one of the tree's apples. We may say, then, that Adam has a right to acquire an apple from the tree.  

The objective of the proviso is to assure that moral cooperation is rational for everyone. Without the constraints of the proviso, Gauthier thinks that moral cooperation is impossible. He argues that moral cooperation is only possible if the interactions that lead to the initial bargaining position were done in observance of the structure of rights and obligations set by the proviso. That is, in essence, the reason why rational agents are ready to accept the proviso. Without the proviso, they are not able to leave the state of nature. We turn now to consider why is it rational for individuals to accept the proviso.

Although the proviso represents a pre-agreement constraint, it is justified by reference to the agreement. What Gauthier argues is, in short, that rational agents will observe the proviso in the state of nature for the following rationale. Rational agents want to enjoy the benefits that moral cooperation, and moral cooperation alone, produces. The proviso sets restrictions that individuals must follow in their interactions with one another in order for moral cooperation to be rational for everyone. If the proviso is not respected, moral cooperation becomes impossible, and the benefits that it affords will lie beyond the reach of everyone. To follow the proviso is, then, a requirement for the emergence of moral cooperation and the benefits it produces. If a rational agent wants to enjoy the benefits of moral cooperation, he must respect the proviso in his interactions with other rational agents in the state of nature.

Gauthier believes that the proviso presents a requirement for the rationality of the agreement

113 Gauthier discusses in some length the actual rights that follow from the proviso. The rights are rights similar to those discussed by Locke and Nozick. The proviso affords each individual a right to the use of his body, powers, physical and mental capacities. Moreover, Gauthier thinks that the proviso affords individuals a right in the fruits of their labor so that they are rightfully entitled to full compensation for the fruits of their labor. Lastly, the proviso also grant individuals exclusive rights to land and external goods. In any case, the actual rights that are produced by the proviso are not relevant for our discussion here. See Gauthier, pp. 210-214, 227.

for one central reason. He thinks that the individuals who were exploited, who were taken advantage of, during the interactions in the state of nature refuse participation in the moral agreement because they were exploited. The exploited would demand the introduction of the proviso in order to participate in the moral agreement. The refusal of the exploited, Gauthier believes, blocks the possibility of a moral agreement for everyone, since the refusal means that a state of nature condition remains between the exploited and the others. By blocking the moral agreement, the exploited hinder the other individuals from enjoying the benefits brought about by moral cooperation. If the non-exploited want to abandon the state of nature and enjoy the benefits that moral cooperation affords, they must cede to the demands of the exploited in order to engender their acceptance to participate in the agreement. Gauthier believes that it is rational for the non-exploited to cede to such demands.

The refusal of the exploited is, we believe, easy to understand. The exploited are the individuals who were taken advantage of during the competition in the state of nature. They are the ones who were forced to perform what Gauthier calls an unproductive transfer. An unproductive transfer is a transfer between two individuals in which only one of them has gains in utility, whereas the other only loses utility. The paradigmatic case of a clear transference is a transference forced by coercion. The individual who performs the transfer under coercion is an individual that leaves the interaction worse than he was before it happened. In a Hobbesian state of nature, unproductive transfers are but one of the manners with which individuals can improve their position regarding utility. In the same vein, the natural distribution that Buchanan writes about is a distribution that emerges from a series of factors that include unproductive transfers among them. Whenever individuals with different levels of power are left to interact in a state of nature, coercion done by those with more power over those with less power with the purpose of improving their situation is something that will happen. The final result is a distribution of goods in which the share of goods of some contains fruits of such unproductive transfers. Gauthier explicitly rejects Buchanan's natural distribution as an appropriate baseline for the initial bargaining position. Should the exploited accept the natural distribution as the baseline for the initial bargaining position, they are, in Gauthier's view, sanctioning predatory activity done by some to improve their position at their own expenses. That is, for Gauthier, irrational since the result of such an agreement would be a moral order that the exploited would regard as inherently unfair since it started from a distributional baseline in which the some individuals were allowed to retain the fruits of their predatory

At this point, it is useful to make a distinction that Gauthier himself does not make. The distinction appears in Stemmer's discussion of Gauthier's argument and also in Rafael Iturrizaga's discussion of it. It is a distinction between strong individuals and weak individuals. The strong are, in essence, the individuals at the profitable end of the unproductive transfers, whereas the weak are those at the losing end of such transfers. What Gauthier's proviso attempts to achieve is to neutralize the effects that differences in power have over the determination of the initial bargaining position. If all individuals accept the proviso, their differences in power are nullified because they are not able to exact coercion upon one another to improve their position. The proviso forbids improving one's own position by worsening that of another. The distinction between strong and weak is essential because, at least prima facie, the proviso favors one group, the weak, over the other, the strong. An initial bargaining position determined by interactions without the proviso is Buchanan's natural distribution; it is the distribution of goods based on power. The natural distribution emerges from a set of interactions that involve, among other forms of interactions, unproductive transfers. The introduction of the proviso causes a shift from the natural distribution to another distribution, one in which the effects of differences in power are nullified. This shift is a shift that makes the strong worse off and the weak better off.

In the context of the refusal of the weak to take part in the moral agreement, Gauthier evokes a distinction between broad and narrow compliance. Broad compliance is the disposition to comply with cooperation if the benefit expected from cooperation, whatever the terms of the cooperation, are higher than the benefits one would get from not cooperating. The broad complier always cooperates if cooperation is more advantageous than non-cooperation. Narrow compliance, on the other side, is the disposition to cooperate whenever the benefits of cooperation approach the benefits that fair and optimal cooperation would afford. Gauthier believes that it is irrational for individuals to adopt the broad compliance disposition. He justifies this affirmation with the argument that a broad complier is someone that invites others to exploit oneself. A broad complier is not satisfied with what Gauthier calls a fair share of the cooperative surplus, but he will demand a larger share instead, the lion's share, in Gauthier's own words. To prevent the strong from

118 See Iturrizaga (2007), pp. 303-304. Rafael Iturrizaga makes an important note here. It cannot be the case that the transfer itself is irrational or unacceptable for the weak. If an individual approaches another with a gun and forces his victim to give him his wallet, the individual that hands over the wallet is acting rationally. He prefers to give the robber the wallet rather than to be shot. The person who is forced to perform an unproductive transfer by the someone else does so rationally. What is unacceptable in terms of rationality for the exploited is not the performance of the transfer, but the initial conditions of the negotiation that includes the unproductive transfer.


120 In the discussion of Gauthier's proviso, power refers to the capacity individuals' have to exploit other individuals.


behaving as broad compliers, Gauthier thinks that the weak must adopt the disposition of narrow compliance. Otherwise, they invite the strong to exploit them.

Gauthier defines a fair outcome as an impartial outcome. Impartiality is defined, in his view, by the Lockean proviso and the principle of minimax relative concession that guides the bargain for the cooperative surplus. In the context of interactions in the pre-moral state, fairness is defined by the proviso. Fair cooperation is cooperation done in observance of the structure of rights and obligations set by the proviso. The disposition to narrow compliance involves, hence, the disposition of observing the Lockean proviso.

In Gauthier's argument, it is irrational for the weak to start cooperation from an initial bargaining position determined by a natural distribution. It is irrational for them, Gauthier argues, to cooperate with individuals that hold in their shares fruits of unproductive transfers that the weak had to make. The key to assuring the participation of the weak, then, is to annul the effects that exploitation had in determining the distributional pattern of resources. Since the unproductive transfers the weak were forced to make constitute the reasons for their refusal, they will become cooperative if such transfers are undone. If we assume that individuals are in a natural distribution, what the weak will demand in order to take part in the moral agreement is that the fruits of predation that are part of the shares of the strong be returned to them, the 'rightful' owners of such fruits. That is what the Lockean proviso requires. Every transfer that worsened an individual must be undone, and the individual that was made worse off by the transfer must be made as better off as he was before the transfer. The proviso demands the strong to return to the weak the fruits of the coercive exercise of power over them. The exploited are ready to take part in the moral agreement only if such alteration in the distributional pattern is made.

That the weak prefer a distribution that annuls the effects of the difference in power between them and the strong is no surprise. For the weak, a distribution endorsed by the proviso is necessarily better than the natural distribution. To accept a shift, then, from the natural distribution to a distribution that observes the proviso is always rational for the weak and exploited. What is surprising is that Gauthier believes that performing the transfers required by the proviso is also rational for the strong. As we argued before, here lies the problem that the contractarian method faces with regards to beneficence, in the form of a problem of distributive justice, namely to convince the strong or wealthy to transfer resources to the weak or poor. In the context of Gauthier's discussion, the difficulty in demonstrating the rationality of such transfers for the strong lies in the fact that they are, for the strong, unproductive transfers, as Iturrizaga has noted. When the strong

---

return the fruits of their superior power to the weak, they are transferring goods without receiving nothing in return. These are transfers that make the strong worse off. If the weak are lead to rejecting the agreement by reason of such transfers, we should expect the strong to refuse to perform such transfers and to reject the agreement if it is conditioned to the performance of the transfers. Gauthier argues, however, that the unproductive transfers the strong must make are compensated by a particular factor that makes it rational for the strong to perform the transfers and take part in the agreement.

The strong, Gauthier believes, are led to accept the demands of the weak for two reasons. The first reason is that they want to enjoy the benefits that only moral cooperation can afford. As long as the weak refuse participation in the moral agreement, the strong are unable to enjoy such benefits. The second reason is that they want not only a condition of moral cooperation but also a stable condition of moral cooperation. For Gauthier, moral cooperation is only stable if it emerges from a distributional baseline that every individual regards as fair. If the strong were to force the weak into cooperation, then, they would establish an unstable system of moral cooperation.

The instability that affords the strong a reason to cede to the demands of the weak proceeds from Gauthier's distinction between compliance and acquiescence. Compliance, he argues, is the disposition to accept fair and optimal cooperative agreements. Acquiescence, on the other hand, is the disposition to accept cooperative agreements that are unfair in order to ensure mutual benefit. If the exploited were to accept the natural distribution as the baseline for the moral agreement, they would acquiesce, and not comply, with the moral agreement. Gauthier argues that cooperation based on acquiescence is less stable than cooperation based on compliance. Compliant individuals are ready to withdraw their support for the agreement as soon as doing so is advantageous for them. He thinks that since fair agreements elicit full-blown compliance from every party, individuals will not press for renegotiation or withdraw from the agreement in virtue of shifts in the distribution of power that underlies the moral agreement. The idea is that cooperation based on acquiescence is completely based on the power structure that underlies the cooperative system. In such a scenario, cooperation is rational insofar as the power structure makes it rational to cooperate. When the power structure changes, the incentive to cooperation will change as well. Individuals that were weak and became powerful, for example, will prefer to withdraw from the cooperative system and

128 As we discuss in chapter four, shifts in the distribution of power are a key factor that affect individuals' perception of how rational it is to comply with the moral agreement. Such shifts can drive individuals to become non-compliant, what creates a problem of stability for the moral agreement. We explore this problem later. Nonetheless, it is useful to mention that Gauthier attempts to solve this problem of stability with an appeal to the ideas of the proviso and fairness.
rebuild it under new terms, terms more favorable to themselves. That is why cooperation based on acquiescence is unstable, in Gauthier's view. Cooperation based on compliance, on the other hand, is more stable.

In sum, Gauthier's justification of the proviso appeals to the surplus of moral cooperation. The weak, on the one side, accept the proviso because the proviso makes them better off and ensures that the initial bargaining position is fair. The strong, on the other side, accept the proviso because doing so is the only manner they have of enjoying the benefits of cooperation and assuring that cooperation will be stable. Without the proviso, the weak refuse to take part in the moral agreement and trap the strong in the state of nature, thus preventing them from benefiting from moral cooperation.

3.2 Gauthier's Proviso Considered

Gauthier's theory is a contractarian theory. For this reason, he has only one manner to justify the proviso. He must show that utility-maximizers will find it rational to constrain their behavior in accordance with the proviso. In order to achieve this end, he must demonstrate that the proviso maximizes the utility of every individual in the state of nature, and thus that everyone prefers to be under its constraints rather than without them. To justify the proviso, then, Gauthier must show that the proviso is indeed such requirement for the agreement. To do this involves two different tasks. The first task is showing that it is irrational for the weak to engage in moral cooperation with the strong when the distributional baseline for cooperation is a natural distribution. For if the weak are willing to cooperate with those that profited from unproductive transfers, then the proviso is not a requirement for moral cooperation in any sense. The second task concerns the demonstration of the rationale of the strong for accepting the proviso. Gauthier must show that it is indeed rational for the strong to accept the constraints set by the proviso. We believe that Gauthier fails in both tasks.

The central reason for the weak's rejection of a moral agreement with the natural distribution as the baseline for the initial bargaining position lies in the unproductive transfers that they had to make in the state of nature. These transfers made the weak worse off. From the perspective of the weak, a distributional pattern established without coercion and unproductive transfers is a pattern where they are necessarily better off than they are in the natural distribution. The rationale of the weak for the proviso is quite simple. They are better off in interactions under the proviso than without it. It surely is rational for them to favor the proviso. However, although it is clear why weak individuals dislike the natural distribution, it is unclear whether their dislike of unproductive
transfers is sufficient to move them to reject moral cooperation. Unproductive transfers may be insufficient to lead the weak to refuse cooperation for two reasons. The first concerns whether the dislike of unproductive transfers is a reason strong enough to reject cooperation. The second problem is whether cooperation is not rational even if the starting baseline is a natural distribution.

The first problem is whether the occurrence of unproductive transfers offers a reason that can be used to refuse cooperation. We must stress that Gauthier's argument refers to individuals in the state of nature. For this reason, the only normative criterion they have to evaluate states of affairs is utility-maximization. The only sort of evaluation that individuals can do is to identify a state of affairs as advantageous or disadvantageous for them. In this respect, the weak's negative view of unproductive transfers must be based on utility-maximization and nothing more. However, it is odd that an individual rejects a specific baseline because he is better off in another baseline. In this case, one could question why would the weak, or any other individual, accept a specific baseline instead of another baseline in which one is better off. If the weak can refuse the natural distribution because they are worse off in it rather than in a distribution constrained by the proviso, then the weak should also be able to rationally refuse a distribution in which they are made worse off by luck or accident, for example.

Consider the case of a man who has his arm ripped off by a wild beast in the state of nature. Losing the arm made him worse off. If a distributional baseline for moral cooperation is suggested after he lost his arm, it is the case that this baseline is worse for him than a baseline in which he still has his arm. Nonetheless, the man can conceive a baseline in which other individuals compensate him for the loss of the arm, and thus make him as better off as he would be in a baseline in which he still had his arm. If such baseline where he is compensated is better for him than that in which he gains no compensation, then why should he accept to cooperate under any baseline except that in which he is compensated for the arm lost? The rationale for the refusal of cooperation here is the same that the weak use to reject the natural distribution. It is a distribution in which they are worse off than they would be in another, the distribution constrained by the proviso, in the case of the weak, and the distribution that involves compensation for the lost arm, in the case we considered.

A second example is that of a person without any talents or skills. Since the reason for refusing the agreement is that the individual would be better off in another distribution rather than in the baseline suggested by the strong, it seems somewhat arbitrary to restrict the demands of the individual to cases that he is made worse off by unproductive transfers alone and leave aside cases of, for example, natural accidents. Why should the weak restrict themselves to ask for compensation for unproductive transfers alone, and not also for all the events that made them worse off in general? It should not matter to them what is the source that made them worse off, but only the fact
that they were made worse off. There is no normative criterion to evaluate transfers except that afforded by utility-maximization. So that in the same manner that a weak individual has its condition worsened by an unproductive transfer, an unlucky individual has his situation worsened by bad luck. From the perspective of the individual that became worse off, it does not matter what the source of the evil that afflicts him is, but only that his situation was worsened.

One could argue here that what underlies the refusal of the weak is not only the fact that the natural distribution and unproductive transfers make them worse off, but that both of them are *unfair*, whereas the case of the unlucky individual has nothing to do with fairness. An appeal to fairness would, then, serve us to differentiate the complaint of the weak against unproductive transfers from that of the unlucky individual. Such an appeal, however, cannot be done inside Gauthier's framework. Fairness and justice are, as we argued in this chapter, moral categories. They are supposed to have normative force over individuals. However, Gauthier is discussing interactions in the state of nature, i.e., the pre-moral state. He cannot claim that an unproductive transfer is unfair, from the perspective of individuals in the state of nature, because fairness and unfairness are themselves defined by the proviso. The idea that a transfer is unfair because it is an unproductive transfer can undoubtedly be endorsed by the weak in the pre-moral state. However, the strong would certainly not endorse the same idea. Without the endorsement of the strong, it is impossible to speak about fairness or unfairness. The sole reason the weak have to reject a distribution brought about by unproductive transfers is that such transfers make them worse off. And if we allow such reasons to determine the initial bargaining position, it is unclear why individuals should accept any baseline except the most advantageous baseline possible for them.

That is one of the problems with Gauthier's arguments for the proviso. The proviso is supposed to be the criterion to define what fairness is in the state of nature. However, Gauthier seems to appeal to fairness as one of the reasons for which it is rational for individuals to act in accordance with the proviso as if fairness and its normative force were anterior to the agreement and the proviso. If the proviso defines what fairness is, one cannot justify the proviso itself with an appeal to fairness. The proviso must be justified with arguments that show that it maximizes the utility of every individual to act following it.

The second problem with the argument for the proviso is whether it is indeed irrational for the weak to take part in an agreement based on a natural distribution. Gauthier's argument emphasizes the cooperative surplus that the weak deny the strong by refusing to cooperate. He never considers, however, that the refusal of the weak also prevents themselves from enjoying the benefits brought about by cooperation. One should not forget that engaging in moral cooperation from a natural distribution is profitable for every individual, i.e., both the strong and the weak.
benefit from moral cooperation. In this sense, refusing to take part in the agreement could be depicted as an irrational choice for the weak, since doing so is to refuse to maximize one's utility. By refusing the agreement, the weak refuse their share of the cooperative surplus, which would make them better off in comparison with how they are in the state of nature. Such refusal would be, we believe, irrational.

Gauthier attempts to justify the refusal of the weak employing the distinction between broad compliance and narrow compliance. He argues that it is irrational for every individual to be disposed to broadly comply. Everyone should, instead, be a narrow complier. It is unclear, however, why is it rational for the strong to adopt the disposition of narrow instead of broad compliance. To be broadly compliant, Gauthier argues, is irrational because it invites others to exploit oneself and it also invites others to demand what he calls the 'lion's share' of the cooperative surplus instead of a fair share. If the individual is strong, then broad compliance assures him the chance of exploiting others to increase his own share and also of acquiring the so-called 'lion's share' of the surplus from cooperation. These advantages make it rational for the strong to be broad, and not narrow, compliers. Narrow compliance presents itself as an attractive alternative for the weak because it offers them the chance of avoiding coercive demands from the strong. The narrow complier, Gauthier argues, always compares the prospects of some cooperative outcome with the prospects of a fair and optimal cooperative outcome. Gauthier introduces, again, an unjustified idea of fairness in such a distinction. The references to a 'fair share' and a 'fair outcome' are displaced here. From the perspective of the weak, fairness depicts here a property of some outcomes that are better for them than other outcomes, and nothing else. In this sense, fairness has no appeal or attractiveness for the strong, since striving for fairness requires them to make themselves worse off. And if the strong refuse to adopt narrow compliance, it follows that the weak will have no gains by choosing to be narrow compliers. In sum, what narrow compliance assures the weak is unclear. The rationale for adopting narrow compliance depends on the strong accepting the demands of the weak. Should the strong refuse to comply with what the weak want, it becomes immediately irrational to be narrowly compliant for the weak, since doing so deprives them of the gains of cooperation. We move now to the case of the strong.

The second key element of the argument is the rationale of the strong to accept the proviso. In this point, there are two reasons that Gauthier suggests that lead the strong to accept the proviso. One is that the weak will prevent them from enjoying the cooperative surplus should they not accept the proviso. The other is that cooperation started from the natural distribution will be unstable.

As mentioned before, Gauthier argues that the weak prevent the strong from enjoying the benefits of cooperation by refusing to take part in the moral agreement. The weak condition their
participation in the moral agreement with the introduction of the Lockean proviso to determine the initial bargaining position. We saw, however, that the sole reason that the weak have to refuse moral cooperation is that they would be better off with a baseline under the Lockean proviso rather than without it. Apart from their gain, they have no other reason to persuade the strong to accept the proviso. A shift from the natural distribution to a distribution constrained by the proviso, however, has the opposite effect for the strong than it has for the weak. It is a shift that makes them worse off since their shares are reduced to augment those of the weak. The strong can launch against the weak the same reason they gave in favor of the proviso, i.e., that the shift to a baseline under the proviso makes them worse off. Why, then, should the strong cede instead of the weak? Remember that the weak, and the strong as well, cannot appeal to fairness to reinforce their reason. What Gauthier considers to be unfair transfers are mere interactions that make the weak worse off. It is unclear what the strong gain from accepting the proviso.

There is an essential element of the relationship between strong and weak that Gauthier does not consider. When the weak refuse to cooperate, they hinder the strong from acquiring the benefits of cooperation. However, the refusal of the weak does nothing to prevent the emergence of cooperation between the strong themselves. The strong can convene and establish a moral framework to regulate their own interactions with one another, excluding the weak from the framework since they refuse to cooperate. Such cooperation, of course, does not provide the full-blown benefits that general cooperation in the system would provide.\textsuperscript{129} Since the weak are not included, the strong remain in a state of nature situation with them. Nonetheless, the strong manage to acquire some level of security and stability by cooperating among themselves. Moreover, given that the individuals excluded from the framework are necessarily the weak individuals in the system, it must be the case that the disadvantages the weak can impose on the strong are not so considerable. The possibility of cooperation among the strong indicates that the weak cannot hinder the strong entirely from enjoying the benefits of cooperation. Consequently, it is possible to notice that the power of bargain of the weak over the strong is not as considerable as it appears to be. The weak can deny the strong the benefits they would get by cooperating with them, but not the benefits that the strong can get by cooperating between themselves. It is questionable, then, whether the weak’s refusal is a sufficient reason for the strong to accept the proviso, since they manage to reap some cooperative benefits even if the weak refuse to cooperate with them.

The second aspect of the rationale for the proviso of the strong concerns the stability of moral cooperation. Gauthier argues that cooperation on terms that are less than fair is an unstable form of cooperation. He thinks that the stability of a cooperation based on terms that are not fair is

\textsuperscript{129} We explore the theme of limited forms of cooperation in the state of nature in chapter three.
entirely dependent on the structure of power underneath it. As soon as shifts in the power distribution occur, individuals will tend to press for renegotiation of the terms of cooperation or even withdraw from cooperation. A fair distribution, in Gauthier's view, is more stable. It is not clear why it is that fair distributions are more stable and not subject to destabilization caused by shifts of the distribution of power. Gauthier's view of fairness, we have already seen, is somewhat unclear and hard to grasp. The strong have good reason to avoid any agreement to justify fairness in the sense that Gauthier understands it since it is necessarily an agreement that makes them worse off. A rational agent's commitment to moral cooperation depends entirely on the advantages that cooperation affords him. If a change in the distribution of power is such that an agent could press for a better position in the cooperative scheme, there seems to be no reason for him to not press for the changes that would improve his condition, regardless of how fair or unfair is the cooperative scheme.

Another aspect is that an unstable cooperative framework is still profitable in some sense, especially in comparison with the state of nature. An unstable framework refers to a condition in which some individuals are strongly drawn to defect from cooperation. Such tendency turns the enjoyment of the benefits of cooperation uncertain in the future. An unstable framework is indeed less advantageous than a stable one. That is true. Nonetheless, an unstable framework is still more advantageous than the state of nature. After all, to enjoy some of the benefits of cooperation is better than having none of such benefits available. The fact that a moral order based on the natural distribution would be unstable, then, does not seem to indicate a solid reason for the strong to accept the proviso.

In sum, it seems that Gauthier's attempt to justify a proviso is not successful. Gauthier's proviso is treated with skepticism by a series of authors. It is unclear why exactly the weak should refuse participation in the agreement, or why the strong should cede to the demands of the weak, or what is precisely the role of fairness in the whole argument. The justification attempt of the proviso seems to fail at several steps. The failure of Gauthier's attempt has at its core the difficulty that contractarianism has to handle beneficence. Gauthier's discussion, it is true, is about distributive justice and not beneficence per se. However, what hinders the justification of Gauthier's proviso is the same factor that hinders the justification of moral rules of beneficence. The strong are required to make transfers to the weak that make them, the strong, worse off. As a consequence, the model of practical rationality as utility-maximization is incapable of endorsing the performance of such a transfer as a rational action.

4 CONCLUSION OF THE CHAPTER

In this chapter, we examined what the problem of justice and its connection with contractarianism are. We saw that the topic distributive justice, as Aristotle conceives it, can be incorporated into contractarian discussions without problems. A conception of distributive justice determines how a specific good ought to be distributed. The determination is done under a standard of entitlement. The just distribution is the one in which the proportion of each individual's share of the good is equal to the proportion each individual's entitlement to the good. From the contractarian perspective, a conception of distributive justice can be justified if the normative demands created by the standard of entitlement it includes can be justified through the contractarian method. With reference to James Buchanan's discussion of private property, we suggested a conception of justice that we called justice as natural liberty. JNL states that the just distribution is that in which each individual's share was obtained under the rules of private property. Although JNL is justifiable from the contractarian perspective, we noticed that it is considerably economic in terms of redistributive demands. Such characteristic indicates what we observed to be a problem for contractarianism in general. The contractarian methodology faces considerable difficulty to justify principles of distributive justice that produce substantive redistributive requirements, such as a principle that demands income redistribution in favor of the poor. The failure of Gauthier's attempt to justify a Lockean proviso is evidence of such hardship. Our objective in this work is precisely to overcome this difficulty and suggest a conception of distributive justice that has more substantive normative requirements on the matter of redistribution than JNL. In the next chapters, we start the presentation of our argument.
CHAPTER THREE: THE CIRCUMSTANCES OF MORAL JUSTIFICATION

OVERVIEW OF THE CHAPTER

The goal of this chapter is to propose the circumstances of moral justification. This idea is a contractarian interpretation of a view that is present throughout the social contract tradition. It is the view of the circumstances of justice. The focus of our interpretation is to depict such circumstances as the circumstances that must be in place in order for moral cooperation to be justifiable for every person in society.

We start the chapter with the presentation of part of the philosophical tradition regarding the circumstances of justice. For this purpose, we briefly consider the views of David Hume, H.L.A. Hart, John Rawls, and Thomas Hobbes. The historical presentation is followed by a section in which we advance the three main circumstances that compose the circumstances of moral justification. The circumstances suggested are partially a rearrangement of ideas present in the works of Hobbes, Hume, Hart, and Rawls and partly an alteration of these ideas. The contractarian interpretation results in three critical circumstances of moral cooperation: unsocial sociability, the social balance of power, and the absence of widespread poverty. The third section presents and addresses critiques from D. Clayton Hubin and Peter Vanderschraff against the idea of the circumstances of justice. We consider their critiques insofar as they affect our interpretation of the circumstances. The consideration of Hubin's critiques is particularly important for our argument. Hubin suggests five different interpretations of Hume's circumstances of justice. He argues that in neither of the five interpretations the scope of justice is limited by Hume's circumstances. We consider Hubin's arguments in each of the five interpretations and then suggest a sixth interpretation of the circumstances. That is the interpretation that considers the circumstances as the
circumstances that enable the *justification* of moral rules.

The importance of the circumstances of moral justification for our argument is clarified in the next chapters. For now, we can say that the circumstances are critical for the justification of more substantive principles of distributive justice. As it is argued in chapter four, the circumstances have an intrinsic connection with the stability of the moral order. And it is with regards to such stability that the justification of substantive principles of distributive justice is possible.

**1 The Circumstances of Justice**

In this section, we consider the philosophical tradition on the subject of the circumstances of justice. For this purpose, we cover the views of four authors. We start with David Hume's account of the circumstances. Following Hume's account, we move to H.L.A. Hart's and John Rawls' discussion of the circumstances. Lastly, we consider Thomas Hobbes' views on the matter. It is true that Hobbes is the first author to discuss the topic of the circumstances of justice, albeit implicitly. Nonetheless, we have two reasons to consider Hobbes' view after having presented those of the other authors. The idea that there are circumstances that enable justice – or morality in general – for persons appears in Hobbes' work only implicitly, whereas it is in Hume's work that the view is explicitly articulated. Additionally, our purpose in this topic is to present a contractarian interpretation of the circumstances of justice. As Hume and Rawls present them, the circumstances refer to justice in particular, and not to morality as a whole. Our interpretation aims at considering the circumstances as factors that enable moral cooperation as a whole, and not only justice, as it is in the case of Hume and Rawls. Hobbes, as a contractarian philosopher, offers us a treatment of the circumstances that helps us to understand how the circumstances can be considered from a contractarian perspective. Hobbes provides us with the link to move into the second section, in which we present the contractarian interpretation of the circumstances.

**1.1 David Hume's View**

In *A Treatise of Human Nature* (1738), Hume writes that an inconvenience for human life arises from a conjunction of the *natural temper* of persons with the *outward circumstances* in which they find themselves. The inconvenience he refers to is the impediment to life in society that the
conjunction of the two factors produces.\textsuperscript{131} For Hume, the natural temper of persons is marked by both selfishness and confined generosity, whereas the outward circumstances are characterized by the easiness with which external goods can change hands and the scarcity of these goods in comparison with the desires that persons have.\textsuperscript{132} Such factors hinder persons from enjoying life in society. Life in society, Hume believes, is the only means by which persons can rise to be superior to other creatures, and also achieve a considerable level of satisfaction and happiness.\textsuperscript{133} To live in society provides every person with additional force, ability, and security, all of which provide individuals with a higher potential for fulfilling their interests and desires. However, despite these advantages that life in society offers, persons cannot join themselves to live in society and enjoy these advantages easily.

The inconvenience appears from the conjunction between the natural temper that persons have and the outward circumstances in which they live. The natural temper of persons is marked by selfishness and limited generosity. People care firstly for their interests and those of a limited group of persons, such as parents and loved ones. Regarding strangers, however, people either care little or are indifferent toward their fate. Because of the natural temper that people have, each tends to be partial toward himself and those for whom he has some affection. In a dispute with a stranger, people will often be biased toward themselves and their loved ones. However, this 'natural' bias is not \textit{per se} a hindrance to life in society. It contributes to the appearance of the inconvenience because people live in a world where the goods they need, desire, and covet are scarce and can easily change hands.\textsuperscript{134} The world is such that it is impossible for every individual to fulfill every desire he has at the same time. By reason of the scarcity of external goods, we can be sure that if every individual covets a particular good, whereas some will obtain the good others will end up without it. Moreover, the fortunate individual who manages to acquire the good should not be too happy. Since goods can generally change hands easily, he should expect that one of the individuals without the good will attempt to snatch it from his hands. Since every individual is partial toward oneself, he will try to acquire every good he wants for himself, regardless of the effects of his acquisition to strangers. Since the world in which these selfish individuals live is such that the available goods are scarce and can easily change hands, not every individual will be able to fulfill

\textsuperscript{131} See Hume, T.3.2.2.7, SBN, p. 486.
\textsuperscript{132} See Hume, T.3.2.2.7, SBN, p. 487. In the context of justice, Hume speaks of a group of goods which he calls 'external goods'. He differentiates between three classes of goods, 'the internal satisfaction of our mind', 'the external advantages of our body', and 'goods we have acquired by our industry and good fortune'. The 'external goods' refers to the third class of goods, for Hume believes the goods of the third class are the ones that can be exposed to the violence of others and may change hands without being damaged or suffer alteration, and, at the same time, are scarce.
\textsuperscript{133} See Hume, T 3.2.2.3, SBN, p. 485.
\textsuperscript{134} See Hume, T.3.2.2.7, SBN, pp. 487-488.
his needs, and those that do manage to acquire a good will be fearful that others will attempt to take it away from them. The natural temper combined with the outward circumstances will create a situation of competition and conflict for goods between persons. Hume characterizes this situation as one of unstable possession of goods.\footnote{See Hume, T 3.2.2.7, SBN, p. 488.} It is this instability, produced by the natural temper of people with the outward circumstances of their lives, that is the impediment to life in society.

The inconvenience is vanquished, in Hume's view, through an artifice provided by judgment and understanding. The artifice is a \textit{convention of justice}.\footnote{See Hume, T 3.2.2.9, SBN, p. 489.} The convention eliminates the inconvenience by terminating the instability of possession. That is done through the introduction of the normative requirement that individuals ought not to interfere with the possession of the others without their permission. It is from this prohibition that the ideas of justice, injustice, property, right, and obligation appear. In Hume's view, the core rule of the convention of justice is the rule of private property.\footnote{See Hume, T 3.2.2.9, SBN, pp. 490-491.} It is by turning mere possession of goods into private property that the instability in possession is eliminated and people can thus live in society. Hume considers that the convention of justice derives its origin from the circumstances that create the hindrance to life in society, the \textit{selfishness} tempered by \textit{confined generosity} and the \textit{scarcity of goods} and their \textit{easy exchangeability}.\footnote{See Hume, T 3.2.2.18, SBN, p. 495.}

In \textit{An Enquiry Concerning the Principles of Morals} (1751), Hume reaffirms the view that justice has its origin in the conjunction of the natural temper of people and the outward circumstances of their life. However, he emphasizes this time that the origin of justice is in its \textit{public utility}.\footnote{See Hume, M 3.12, SBN, pp. 183-188.} To demonstrate that, Hume refers again to the factors of selfishness, confined generosity, and scarcity of goods. He argues that these factors are responsible for making justice \textit{useful} for persons, whereas justice would be rendered useless if the factors were altered, and thus justice would never appear. The usefulness to which Hume refers arises from the fact that justice allows individuals to live in society, and therefore to have access to the superior force, stability, and security that social life provides. However, in the \textit{Enquiry}, Hume introduces a new factor in the group of conditions that make justice useful for individuals. The factor introduced is the \textit{equality of powers}.\footnote{See Hume, M 3.18, SBN, pp. 190-191.} Hume considers the fictional case of interactions between humans and rational creatures that are considerably inferior regarding the strength of body and mind than humans. Because of the weakness of the creatures, there is no instability of possession between them and the humans. The beings are neither strong nor smart enough to be a threat for humans in a competition and conflict.
over scarce goods. Hume argues that there can exist no society between humans and these beings, for society requires a degree of equality in powers between its participants. In this case, humans would not be bound by the laws of justice in their interactions with these beings because such laws would be, from the perspective of humans, completely useless.

In sum, Hume believes that justice – its rules, rights, duties, and institutions – has its origin in its usefulness for persons. Justice is what enables individuals to live in society and thus to enjoy the fruits that only life in society can provide. The usefulness of the convention is a product of the conjunction between the natural temper persons have and the outward circumstances in which they find themselves. In this sense, justice is useful in virtue of three factors: *selfishness tempered by confined generosity, scarcity of external resources and goods*, and *equality of powers between persons*. Without the presence of any of these factors, justice is rendered useless. These factors comprise what is referred to by the philosophical tradition as the *circumstances of justice*.

Hume is often understood to be a critic of the doctrine of the social contract.\(^\text{141}\) Recently, however, some authors have argued that Hume's position is nearer to contractarianism than philosophers have previously thought.\(^\text{142}\) Philosophers have suggested that the convention of justice which Hume writes about can be interpreted as a moral agreement. A strong reason to endorse this interpretation is that Hume seems to connect the emergence of a convention of justice with the utility of such a convention for individuals. He explicitly argues that the convention arises because it is *useful* for persons. Such an argument for the appearance of justice resembles the contractarian method of justification that we presented in the first chapter. Contractarians discuss morality whereas Hume discusses justice. Nonetheless, both call attention to the fact that morality, in the case of contractarians, and justice, in Hume's case, is something useful for persons. Hume, however, does not explicitly discuss the justification of the convention of justice, whereas contractarians always refer to such a concept in their theories. Some authors interpret Hume's theory as an account of the origin and development of the social and legal institutions.\(^\text{143}\) Such interpretation distances Hume's theory from the social contract tradition. Under this interpretation, the circumstances of justice are empirical circumstances whose presence is a necessary condition for the emergence and development of institutions connected with justice. However, whether Hume's theory of justice can be interpreted as a contractarian theory or not, Hume's account of the circumstances of justice was imported into the social contract tradition. In the social contract view, the circumstances would be

\(^{141}\) For Hume's criticism of the idea of the social contract, see Hume (1996), pp. 274-292. See also Brownsey (1978) for an overview of Hume's arguments against the doctrine of the social contract.

\(^{142}\) There is a discussion about whether David Hume can be considered a contractarian in the line of Hobbes. In this respect, see Gauthier (1979) and Whelan (1994).

the conditions that make the social contract useful for everybody. John Rawls has explicitly endorsed the idea of the circumstances in his social contract theory. We consider now the views of H.L.A. Hart and Rawls on the circumstances. Both authors have enlarged the list of the circumstances in comparison with Hume's account, and their discussions serve us to understand how Hume's idea can be transposed into the social contract tradition.

1.2 H.L.A. Hart's and John Rawls' Views

Hume's view has been resumed and further developed by a series of authors in the philosophical tradition. Such authors have provided a richer description of the factors that compose the circumstances than Hume. We consider here two authors in particular, H.L.A Hart and John Rawls. Hart refers to what he calls 'truisms' from which the minimum content of natural law can be inferred, whereas Rawls writes about the circumstances of justice, which are circumstances that make social cooperation necessary and possible.

Hart writes about five truisms that make particular rules of justice necessary. The truisms are: Human vulnerability; approximate equality of powers, limited altruism, limited goods, and limited understanding and strength of will. Hart's emphasis is that these natural facts make some rules of justice necessary for persons – which Hart refers to as the minimal content of natural law. Examples of such rules are the prohibition of murder and the institution of private property. With Hart's view, it is possible to observe that the circumstances about which Hume writes not only give us a formal idea of justice, i.e., a convention to solve conflicts, but that they also provide us some material content for the convention. A law to prevent the occurrence of murders, for example, reveals itself necessary when we consider that individuals are mutually capable of killing one another since they have approximate powers and similar vulnerabilities.

Hart's idea of truisms that indicate a minimal content of the natural law augments the number of circumstances from three to five. Rawls' discussion of the circumstances presents an even greater augmentation than Hart's. He writes about conditions that make human cooperation both possible and necessary. He refers to these conditions as the circumstances of justice. Rawls

144 See also J.R. Lucas (1966), pp. 1-10 and Simon Hope (2010) for additional discussions about the circumstances of justice.
145 Hart believes that the fundamental reason for which individuals form societies is survival. He believes that individuals have better prospects of survival in society rather than in isolation. In this respect, his argument is close to Hume's, who argues that the cornerstone for a convention of justice is the individuals' desire to live in society. 146 Hart (1994), pp. 193-200.
147 Rawls seems to be the first to refer to Hume's idea of the inward and outward circumstances that make justice useful for persons as the 'circumstances of justice'. The philosophical tradition seems to have adopted Rawls
divides these circumstances into two categories, objective and subjective. The objective conditions are: coexistence at the same time and geographical location, similarity in mental and physical powers, vulnerability to attack, and moderate scarcity of goods. As for the subjective conditions, they are: similarity of needs and interests; different plans of life; limits in knowledge, thought, and judgment; and a diversity of religious, philosophical, political, and social doctrines. Rawls' account of the circumstances of justice greatly expands on Hume's. We believe, however, that some of the circumstances listed by Rawls can be deemed to be circumstances of justice only in the context of his discussion, but not generally.

In Hume's view, the circumstances of justice are the circumstances that render justice useful for persons. The usefulness of justice is drawn from the fact that it is justice that allows individuals to overcome the competition and conflict for external goods caused by the circumstances. Rawls, in his turn, claims that the circumstances are conditions under which social cooperation is possible and necessary. Such a definition is not different from Hume's. If we understand that the 'life in society' to which Hume refers amounts to what Rawls calls 'social cooperation', then both philosophers are writing about the same subject. Nonetheless, some of the circumstances that Rawls lists can hardly be depicted as circumstances of justice. In particular, it is difficult to see the relevance of the conditions of similarity of needs and interests, different plans of life, and diversity of religious, philosophical, political, and social doctrines as circumstances of justice. To see why it is problematic to consider such conditions circumstances of justice, consider the following example. Three persons form a society, a Catholic monk, a socialist politician, and a compulsive gambler. It should be clear that these three individuals have wildly different needs and interests as well as wildly different plans of life. However, there is no reason to think that social cooperation is impossible or unnecessary between them, even though they are wildly different persons. Social cooperation, and thus justice, remains useful for all three of them. Considering the diversity of religious, philosophical, and social doctrines, it is also hard to see how this condition can be one of the circumstances of justice as well. Persons in feudal Europe were generally

149 The exact relation between the circumstances and justice is subject to interpretation. In the last topic of this chapter, we consider D. Clayton Hubin's discussion on the matter. It is clear, nonetheless, that regardless of the exact connection between the convention of justice and the circumstances, utility is an element involved in the connection. Hume puts strong emphasis in the view that the convention of justice is useful for individuals in virtue of the circumstances in which they find themselves. Due to Hume's emphasis, we consider here, and throughout the discussion, until we come to discuss Hubin's arguments, that the circumstances are responsible for making justice useful for individuals in Hume's view.
150 One could argue that Rawls refers here to interests such as the fundamental general interests that Stemmer suggests. However, if Rawls is indeed referring to such interests and desires, the condition is simply redundant. These are desires and interests that are obviously shared by all humans. To imagine someone that does not have an interest in not being killed or does not have a desire for food is to imagine someone that is not human.
Christian almost without exception, and it is clear that the people then still enjoyed social cooperation and were in need of justice. Even if we imagine an ideal society where every person has the same religion, philosophy, and social doctrine, it seems evident that social cooperation is possible and necessary for them, provided the other circumstances of justice are in place.

The circumstances that were present in the arguments of Hume, Hart, and Rawls, are essential for the contractarian view of morality. As argued before, contractarianism justifies the normativity of moral rules referring to the interests of individuals to whom the rules are supposed to be applied. Contractarians argue that a rule is justified for a person if it is rational for that person to accept the normative constraints of the rule. What justifies the rule, then, is the fact that the rule is advantageous for the person whom it constrains. The idea that underlies the contractarian method of justification is similar to that of the origin of Hume's convention of justice, namely usefulness. Moral rules are justified in virtue of their usefulness for individuals in the same manner that the convention of justice emerges because of its usefulness for persons. It is in an appeal to usefulness that the contractarian method and Humean view of justice have a common ground. Rawls' account of the circumstances allows us to notice that. Rawls defines the circumstances as the conditions that make social cooperation both possible and necessary. In other words, individuals have a sufficient reason to cooperate only in the case that they find themselves in the circumstances of justice. Otherwise, cooperation is either impossible or unnecessary for them, i.e., it affords them no advantage or benefit. Rawls provides us the link to import Hume's view into the social contract tradition. They are the circumstances that make the social contract useful for individuals, i.e., they enable the social contract.

In the next topic, we consider the view of Thomas Hobbes on the subject of the circumstances of justice. Although Rawls indicates which role the circumstances play in the social contract tradition, it is in Hobbes' view that we observe how the circumstances of justice are relevant for the contractarian branch of the social contract tradition.

1.3 THOMAS HOBBES' VIEW

Gregory Kavka offers an excellent summary of Hobbes' argument for the depiction of the state of

151 A difference between Hume's arguments about the convention of justice the contractarian views of the moral agreement concerns the importance of the individual. As we argued, the fundamental contractarian norm of justification is a unanimity principle, represented in the thought-experiment by the moral agreement. Hume does not, at least explicitly, depicts the convention as a pact between individuals. His argument is not necessarily committed with a rule of unanimity in justificational endeavors, in the same sense that the contractarian argument is.
nature as a state of war. Kavka's summary is useful for us because it stresses the aspects of Hobbes' argument that we can connect with the idea of the circumstances of justice. Kavka argues that the structure of Hobbes' argument is such that five specific factors create the situation of conflict. Three of these factors appear explicitly in Hobbes' argumentation, and two of them are implicit in it. The three explicit factors are the equality in physical and mental faculties, the conflicting desires that individuals have, and the fact that individuals are forward-looking. The two implicit factors are the advantage that anticipation offers individuals in the situation of conflict and the fact that individuals have limited altruism.

The five factors that Kavka lists are the factors responsible for creating the state of war. Notice that the Hobbesian state of war is related to social cooperation in the same manner that Hume's view of the instability of possessions is. They are both hindrances to social cooperation. They hinder individuals from enjoying the benefits of social cooperation. In Hobbes' case, such a hindrance would not exist if the five factors summarized by Kavka did not define the state of nature in which individuals find themselves. If their physical or mental powers were substantially unequal, or their desires were in complete harmony with each other, or even if individuals were not self-interested, but highly altruistic, in any of such cases, there would not be any hindrance to social cooperation. Nonetheless, since such conditions are present, they thwart cooperation. The circumstances create the suboptimality of the state of nature that we stressed in the first chapter. It is because individuals are constrained by the circumstances that they find themselves in a suboptimal state regarding the promotion of their interests, namely the war of all against all. Consequently, we can connect the circumstances with the central reason individuals have to take part in a social contract, namely to overcome the suboptimality of the state of nature.

Notice that like the convention of justice in Hume's argument, Hobbes introduces the social contract because it is useful for individuals. It is introduced because it helps individuals overcome the hindrance to social cooperation of the war of all against all. Here we find the connection between contractarianism and the circumstances of justice. The usefulness of the contract can be traced to the five factors that Kavka stresses. It is because those conditions are in place that there is a hindrance to social cooperation, and it is because there is such a hindrance that the social contract is useful for individuals. Since the circumstances are responsible for making the social contract useful for individuals, we can argue that they are the conditions that make participation in the social contract rational for every person. The contract is advantageous, and thus participation in it

---

152 To be forward-looking in Kavka's summary is but to be self-interested and concerned with the satisfaction of one's present and future desires.
154 We presented Hobbes' argument in chapter one already. All the references to Hobbes' view can be found there. For this reason, we do not go over the argument here again.
rational, because individuals find themselves in the circumstances of justice. Otherwise, the contract would have no usefulness for individuals, and hence participation in the social contract would be irrational either for everybody or for a particular group of individuals. In both cases, nonetheless, a social contract would be impossible.

In the context of contractarianism, we can interpret the circumstances that define the usefulness of the moral agreement as the *circumstances of moral cooperation*. They are circumstances in which participation in a moral agreement, i.e., moral cooperation, is rational for every individual in the social system. Now that we verified in which sense Hume's circumstances can be applied to contractarianism, our next step is to analyze each circumstance. Our analysis involves a reformulation of the circumstances. However, we do not think that we improve much on Hobbes' or Hume's description of the circumstances. The general goal of the reformulations done here is to adapt Hume's circumstances to the contractarian outlook. Our three circumstances are: unsocial sociability, the social balance of power, and the absence of widespread poverty. Naturally, other minor conditions are part of the circumstances of moral cooperation, such as physical proximity, existence at the same time, a capability of language, a shared language, and so on, but since these are self-evident conditions, we will leave them aside.

## 2 The Circumstances of Moral Cooperation

We come now to our interpretation of the circumstances of justice as the circumstances of moral cooperation. In the section above, we examine the circumstances of justice as they are formulated by Hume, Rawls, Hart, and Hobbes. Our task in this section is to offer our own formulation of the circumstances. In some sense, our reformulation amounts to a rearrangement of the factors proposed by the four authors considered above. That is true especially in the case of the first and third circumstances we propose, unsocial sociability and absence of widespread poverty respectively. It is mainly in the context of the second circumstance, the social balance of power, that we approach critically some factors suggested by the philosophical tradition.

### 2.1 Unsocial Sociability

The first condition is the hardest to formulate. It encompasses, but it is not restricted to, the majority
of the factors suggested by Hume, Hobbes, Hart, and Rawls. From Hume's view, our first condition contains what he called selfishness and confined generosity. From the five factors of Hobbes' argument, our first circumstance draws from the fact that persons are forward-looking and that they have limited altruism. From Hart's account, it refers to the truisms of human vulnerability, limited altruism, and limited understanding and strength. From Rawls' exposition, it contains the circumstances of similarity in mental and physical powers, vulnerability to attack, a similarity of needs and interests, limits in knowledge, though, and judgment. The first condition refers primarily to internal factors that play a significant role in defining individuals preferences regarding the subject of social and moral cooperation.

The first circumstance was covered in some sense in the first chapter. When we discussed the instrumental view of practical rationality, such a discussion was already about the first circumstance of moral cooperation. We stressed then that contractarianism depicts individuals through a combination of the utility-maximization view of practical rationality and a constellation of fundamental general interests. Such a depiction is a more complex and elaborated form of Hume's and Hobbes' accounts of individuals' selfishness. In this section, we want to elaborate further the account of individuals' that we suggested in the first chapter. For this purpose, we refer to an idea of Immanuel Kant. Although a contractualist, and not a contractarian, Kant depicted individuals through an idea that suits the contractarian outlook perfectly. That is the idea of *ungsellige Geselligkeit*, which captures the central aspects of human nature that we wish to stress with the first condition of moral cooperation.

In Kant's view, persons incline toward social life and, at the same time, have an inclination to separate themselves from society. He argues that the inclination to enter society springs from the fact that it is in the society that humans develop their natural dispositions. However, alongside the inclination toward social life, persons have are drawn to isolation in virtue of the tendency they have of wishing to arrange everything following their will. This tendency causes other people to resist the individual since others also want to arrange things in accordance with their own will. The unsocial sociability depicts a contradictory inclination that pushes the individual toward social cooperation with one another and simultaneously makes social cooperation challenging to execute.

With Hume and Hobbes, we saw that social cooperation offers advantages and benefits for persons that they cannot achieve in isolation. Life in society provides individuals with a variety of goods – material and abstract – that can be neither acquired nor produced outside of society. Hobbes, for example, explicitly argues that, outside of society, advantages such as farming, trade,

---

technological development, arts, and the development of knowledge are not possible. Hume, as we saw, thinks that life in society offers individuals additional force, ability, and security, all of which provide individuals with a higher potential for fulfilling their interests and desires. It is the fact that individuals cannot access such benefits that makes the state of nature suboptimal. What both philosophers indicate is that there are particular advantages and benefits that individuals can obtain only through social cooperation with one another. Nonetheless, despite the attractiveness of such benefits, individuals are prone to face difficulties in cooperating. One of the factors that produce such difficulty is the inclination to unsociability that people have.

It is the existence in a limbo between sociability and unsociability that makes moral cooperation both profitable and hard to execute for individuals. Sociability indicates that there are benefits that persons can only acquire by cooperating, whereas their unsociability means that cooperation does not happen spontaneously and out of their own nature. Were humans perfect sociable creatures, they would enjoy the benefits of social cooperation without ever feeling a single antisocial inclination. As perfect sociable creatures, they would interact with one another without any need for external constraints, such as morality or justice. Perfect social persons would be persons that treat every stranger in the same manner that we treat our most intimate and beloved friends. The individuals that inhabit Kant's ideal of the kingdom of ends (Reich der Zweck) are an interesting representation of what perfect sociable humans would be. An individual in the kingdom of ends treats every other individual always as an end and never as a means. In such a realm, individuals require no moral law to regulate their actions. The reason for that is that the actions that the moral law requires of us, actions whose performance we consider to be duties, are actions that they perform naturally, without feeling the slightest temptation to do otherwise. Moral cooperation for perfect social beings is unnecessary. They can reap the advantages of social cooperation without requiring to engage at the same time in moral cooperation.

In case humans were perfect unsocial beings, moral cooperation would also reveal itself to be unnecessary for them. Perfect unsocial beings are beings that can reap no considerable benefit or advantage from social cooperation or moral cooperation. Their nature does not contain dispositions that can be developed through social cooperation. Hume speaks in the Enquiry of a society of ruffians as an example of a society composed of individuals with antisocial tendencies. However, even a society of criminals is still a society, and its members still have something to gain through social cooperation. Perfect unsociability involves more than a tendency to break the law. It involves

157 In the sixth chapter of this work, we consider in more detail the relation between moral cooperation and social cooperation.
158 See Kant (1785), p. 439.
159 See Hume M 3.9, SBN, p. 187.
a biological and psychological nature such that the individual either has nothing to gain from social cooperation or that the advantages that social cooperation afford him are outweighed by disadvantages brought about by it.

A fictitious case that serves to illustrate the case of perfect unsociability is that of the vampire. The vampire is a rational creature that requires the blood and flesh of humans to survive. We can imagine that the vampire and the human could profit from some cooperative endeavors. The human could protect the vampire during the day, while the vampire protects the human during the night, for example. However, despite the potential benefits that both could acquire from social cooperation, it is impossible for them to cooperate rationally. To enable social cooperation, the human and the vampire must first engage in moral cooperation. It is through moral cooperation that the difficulties to establish social cooperation are overcome, as Hume and Hobbes argue. However, moral cooperation between the two implies that one of the following two outcomes must follow. Either the vampire will be prohibited from attacking the human, and it will starve, or it will be allowed to attack and kill the human. Whatever the benefits of social cooperation are, the drawbacks that it brings either for the human or the vampire are so substantive that neither of them can rationally agree to engage in moral cooperation.

It is interesting to notice that perfect sociability and perfect unsociability have different effects on the possibility of moral cooperation for individuals. Perfect sociability makes moral cooperation unnecessary, but it does not make it impossible. Kant's kingdom of ends could contain a rulebook of moral norms for cases where the individuals have trouble to coordinate their actions to achieve a goal desired by everyone, even though such rules are likely to be useless. No aspect of perfect sociability explicitly hinders the existence of moral norms. Perfect unsociability, on the other hand, makes moral cooperation both unnecessary and impossible. Whereas perfect social beings might be indifferent regarding the existence of moral rules, perfect unsocial beings always prefer that moral rules do not exist rather than the contrary. As argued before, perfect unsociability is marked by the fact that persons cannot reap any benefit from social cooperation, and also that cooperation has negative consequences for them. For a perfect unsocial being, engaging in moral or social cooperation is something necessarily irrational, since doing so makes him worse off necessarily.

There is another important factor involved in the unsocial sociability that we want to stress before proceeding to the next circumstance of moral cooperation. That is the limitation in the knowledge that individuals have. Such a factor is stressed by two of the philosophers we considered. Hart writes about the truism of limited understanding and strength of will and Rawls about limits in knowledge, thought, and judgment. What such factors emphasize is that persons do
not have full-blown knowledge about the empirical circumstances that surround them. Although they are rational and they can understand the world around them, their cognitive and perceptual powers are such that their comprehension of the world is limited and flawed. For this reason, their capacity of predicting the future is also limited and flawed. The suggestion of such limits to individuals' knowledge and cognitive capacities compose one of the central tenants of Friedrich A. von Hayek's thinking.

Hayek was a fierce critic of what he called constructivist rationalism. In Hayek's view, constructivist rationalism is the approach characteristic of economic theory. Economists tend to refer to models where agents are assumed to be, among other things, perfectly rational and to have perfect knowledge of the situation in which they find themselves. In the analysis of a market, for example, economists start from a model of perfect competition, which involves the assumption that every consumer and producer in the market has complete and immediate knowledge of market prices, their own utility functions and those of other individuals, among other factors. Game theoreticians often refer to perfect knowledge as one of the assumptions of their model. Consider the prisoner's dilemma between the two farmers that we suggested in the first chapter again. In it, both farmers are assumed to have perfect knowledge about the outcomes that may result from their actions, about their own utility expectations and those of the other player involved in the dilemma. It is such form of assumption that Hayek criticizes.\textsuperscript{160}

Instead of the assumption of perfect knowledge, Hayek favored a type of epistemological pessimism regarding individuals' capacities to know and understand social phenomena.\textsuperscript{161} He stresses the 'fact of the necessary and irremediable ignorance of everyone's part of most of the particular facts which determine the actions of all the several members of human society'. And also the 'fact of our irremediable ignorance of the particular facts which determine the processes of society'.\textsuperscript{162} We want to incorporate such facts into the circumstance of unsocial sociability. Individuals' knowledge about the mechanisms operating in society is limited.

In sum, the first condition that is part of the circumstances of moral cooperation is the unsocial sociability of individuals. Individuals are neither perfect social creatures nor perfect unsocial creatures. Instead, they lie in a middle ground between these two hypothetical cases. Unsocial sociability indicates that individuals are beings for whom social cooperation is advantageous but also difficult to execute. Individuals are drawn to engage in cooperation with one another in virtue of the profits they acquire from doing so, but their nature offers hindrances to establishing cooperation. It is because of such contradictory tendencies that the introduction of

\textsuperscript{161} See Gamble (2006) for an in-depth discussion of Hayek's views on knowledge.
moral rules and moral cooperation, in general, appears as viable endeavors for individuals. What the condition of unsocial sociability stresses are the internal factors that govern the individuals' preferences which draw them to see social cooperation as profitable and, at the same time, to have difficulties to cooperate.

2.2 THE SOCIAL BALANCE OF POWER

The second condition that composes the circumstances of moral cooperation refers to power and its relevance for moral cooperation. The intuition that power is one of the key figures of the circumstances of moral cooperation appears in Hobbes, Hume, Hart, and Rawls. Hobbes writes that there are no significant inequalities in physical and mental faculties of individuals and that this fact is key to the production of diffidence among persons. Hume, in his turn, writes that a society can only exist between beings with similar levels of strength and cleverness. Hart refers to human vulnerability, approximate equality of powers, and a limited understanding and strength of will as some of the truisms that give minimal content to natural law. Rawls includes the vulnerability to attack, and limits in knowledge, thought, and judgment as among the circumstances of justice. The fact that these elements are included among the circumstances of justice reflect – directly or indirectly – that the distribution of power in the social scheme plays a vital role as a circumstance of justice. However, although there is a group of circumstances connected with power in the views of the four philosophers, they are often summarized in the tradition as one single condition, namely the equality of powers.

2.2.1 EQUALITY OF POWERS

The idea behind equality of powers is that the distribution of power in the social scheme must be egalitarian in order for moral cooperation to obtain. When persons have a similar amount of power, it follows that they are mutually capable of imposing their interests and desires over each other, and also that they are mutually vulnerable to be victims of each other regarding the imposition of interests and desires. In this sense, the more powerful an individual is, then the abler he is to fulfill and promote his interests regardless of the opposition from other individuals, and the

163 As we observed in chapter one, the same intuition appears throughout the contemporary contractarian tradition. Gauthier's theory and Stemmer's contain explicit references to an equality of powers as a requirement for moral justification. See footnote 60.
less vulnerable he is to have his plans frustrated by the interference of others. If the distribution of power is egalitarian, then individuals are equally vulnerable to having their plans frustrated by the others and are equally capable of frustrating the plans of one another.

It is interesting to notice that part of the benefits that cooperation produces for persons arises directly from this condition. Hobbes noticed that social instability follows from the egalitarian distribution of power. Hobbes stresses that it follows from the equality of powers – conjoined with the other circumstances – that persons fearing that others will interfere with their interests have a high incentive to anticipate the interference of others and act first. In this sense, persons in a setting of equality of powers are highly motivated to interfere with others even when the interference provides no direct benefits, but serves as a deterrence measure to stop others from interfering with one's plans. To introduce moral cooperation in an environment where people have this perverse incentive is to introduce a measure that removes the rationale for anticipation. When moral cooperation is introduced, the expectation that others will interfere with one's plans disappears, and with it the incentive to anticipate others as a deterrence measure. With the elimination of the incentive, people can enjoy thus a degree of stability in the promotion and fulfillment of their plans that was not possible without moral cooperation. This stability is one of the advantages brought about by moral cooperation.

Although we are in accord with the four philosophers about power being a key element in moral cooperation, we think that the image of an equal distribution of power between individuals is not the best representation of such a circumstance of justice. The equality of powers view has two problems. The first problem is that it fails to capture a central aspect of power in social relations, namely the role that coalitions play in such relations. The image of equality of powers is too focused on the individual level. As it is proposed by the authors considered, it refers to a roughly equal distribution of power between individuals. As we argue ahead, if we allow for the possibility of coalitions in the social system, an equal division cannot sustain moral cooperation necessarily, and thus must be rejected as one of the circumstances of moral cooperation.

The second problem that we see in the circumstance of equality of power is that it leaves contractarianism without any means to account for moral cooperation between individuals with a considerable difference in their amounts of power. The inability to account for such form of moral cooperation has been pointed as a reason to reject contractarianism as a whole by some philosophers. Derek Parfit, for example, argues that the Hobbesian social contract tradition has a sensitivity to differences in power that makes it impossible for theories in the tradition to justify rules that are included in the moral beliefs of most people, such as rules that favor the poor,
children, or those with physical disabilities.\textsuperscript{164} Parfit thinks that the inability of the contractarian doctrine to account for such rules offers us a reason to reject it.

Martha Nussbaum advances a similar critique to Parfit's against the Hobbesian tradition of the social contract.\textsuperscript{165} Her critique is directed at contractualism, but it is also applicable to contractarianism. She argues that no social contract doctrine includes in the group of people that are part of the contract those with severe and atypical physical and mental hindrances. Such individuals are excluded from the group that selects the rules, and also excluded from the group of people for whom the rules are chosen. They are neither participants of the social contract nor members of the moral community created by the contract. In her view, such exclusion is a direct consequence of the social contract theorists insistence that equality of powers is a circumstance of justice. Given contemporary moral attitudes toward people with impairments, the exclusion that follows from the social contract method reveals itself a problem for the theory, argues Nussbaum. Although Nussbaum's argument is directed against political contractualism, it is, in essence, the same argument that Parfit launches against contractarianism. They both argue that the fact that contractarianism is unable to account for a widespread moral belief that people have, namely that persons' with impairments are part of the moral community, is a problem for the doctrine.

Parfit's and Nussbaum's critiques against contractarianism are elaborations of the more general problem we presented in the second chapter, namely the difficulty that contractarianism has to justify moral rules of beneficence. As we considered, rules of beneficence and rules of distributive justice with substantive redistributive demands can hardly be justified by the contractarian method because they lack the sort of reciprocity that the method requires. The critiques of Parfit and Nussbaum indicate that the lack of reciprocity can be traced to the circumstance of the equality of powers. In the case of a rule against murder, for example, we are only able to justify the rule, and also the rights and obligations that entail from it, for a group of individuals whose members are mutually capable of murdering one another. If among these individuals there is one that is not powerful enough to murder another individual if he so wishes to do it, then the others have nothing to gain from including him in the moral community. If they were to do so, they would grant the weak individual a right not to be murdered by them and in exchange acquire a right to not be murdered by the weak individual, which is something that cannot come to pass whatsoever. The strong gain nothing by including the weak in the moral community. It is for such reason that, as we considered, Gauthier's idea of a Lockean proviso fails. The contractarian framework seems to have no means of justifying normative demands for the strong in favor of the

weak, since a substantive inequality of power subsists between the two groups.

We propose ahead a reformulation of the circumstance of equality of powers. The central purpose of the reformulation is to, first, allow for the possibility of coalitions between individuals and, second, to address the problem stressed by Parfit and Nussbaum. Our reformulation, however, does not solve completely the problem indicated by Parfit and Nussbaum. Instead, it suggests that contractarianism is not committed to the exclusion of the weak of the moral framework necessarily. But before proceeding to the reformulation itself, we want to discuss the concept of power that underlies our discussion.

2.2.2 The Concept of Power

Throughout our discussion, we have been referring to 'power' without a careful treatment of the concept. We move now to consider the concept with some care. The reason to perform this task is that power is, as we have been arguing, a central aspect of contractarianism. In chapter two, we saw that JNL is, in essence, a conception of justice anchored in a specific idea of power. In later chapters, our argument considers in detail the effects that power has over social interactions and persons' attitudes toward moral cooperation. Due to the importance that this concept has for our discussion, we must consider it in some detail.

Keith Downing differentiates two concepts of power, namely outcome power and social power. Outcome power is the capacity or ability to realize or help in the realization of outcomes. Downing also refers to it as 'power to'. That is the concept of power implicit in our previous discussion of justice that culminated in JNL. Social power, on the other side, is the capacity or ability to deliberately alter the structure of incentives of another person or group in order to produce an outcome. Downing also refers to social power as 'power over'.\footnote{166 See Downing (1996), pp. 5-8.} In essence, both forms of power are about bringing about outcomes. Downing argues that whenever an exercise of outcome power affects other persons, both forms of power would be the same were not for one single factor. In an exercise of outcome power, the effects over other individuals are irrelevant for the actor's action. The actor desires merely to bring about an outcome. In social power, on the other hand, other individuals are deliberately affected in order to bring about the desired outcome. The effects over other individuals are a constitutive part of the intention to bring about an outcome in the case of social power, but not in outcome power. Coercion is a major example of social power. Through coercion an individual, the stronger, uses his strength to force another, the weak, to do something
the strong desires. Without the presence of coercion, the weak find it irrational to acquiesce with the strong's desire. The introduction of coercion serves to change the incentive structure of the weak so that acting in a manner that the strong require becomes rational for the weak. Social power is the form of power that figures in Buchanan's discussion of private property and Gauthier's discussion of the proviso. Buchanan's natural distribution is the situation where individuals are unable to predate each other to increase their shares. They do not have the social power to alter the incentive structure of other individuals in order for them to cede part of their shares. In a similar sense, what stands at the core of Gauthier's proviso is the idea of an unproductive transfer. Such a transfer is but an exercise of social power. In it, the strong use their social power over the weak, changing the incentive structure of the weak to perform the transfer.

Downing's differentiation helps us understand to which type of power the circumstance of equality of powers refers. In this respect, the type of power that philosophers have in mind when they identify equality of powers as a circumstance of justice is social power. In Hume's thought-experiment of the interaction between humans and feeble rational creatures, what hinders social cooperation between them is a discrepancy in social power. The humans can impose their desires upon the feeble creatures. They can change the incentive structure of the feeble creatures at their will and thus force them to bring about outcomes they desire. The weak creatures, in their turn, are not able to face the humans in the same manner. Their social power with regards the humans is minimal.

The idea of social power can be expanded if we consider Robert A. Dahl's analysis of power. Dahl defines power as a relational concept. In his view, power is a relation between individuals or group of individuals. He refers to power with this intuition: A has power over B to the extent that he can get B to do something that B would not otherwise do. That is the same notion of 'power over' or social power that we saw above in Downing's definition. The usefulness of Dahl's concept for our discussion is that it introduces a manner to compare powers. As we see ahead, this idea is fundamental for the analysis of the circumstance of equality of powers.

In Dahl's concept, power between persons involves four different factors. One can speak of the basis of A's power over B, of the means used by A to exact power over B, of the amount of power over B, and of the scope of power over B. The basis of one's power refers to the resources one can employ to affect the actions of another individual. The means are the manners in which the basis can be used to determine the behavior of another individual. The means can include, for example, a promise or a threat involving elements from the basis. As for the amount of one's power, Dahl refers to it as a probability. The amount of power indicates the odds an individual has of

successfully coercing another individual to act in a desired manner. Lastly, the scope of one's power refers to the range of actions or behavior that one is capable of compelling another to perform. Lastly, Dahl argues that comparisons in power are always comparisons done for two or more individuals regarding one of these four factors. It is by referring either to the basis, means, amount, or scope of two or more individuals' power that we can affirm that one individual is more powerful than another.168

It is of central importance for our discussion Dahl's assertion that power, with regards to its amount, has a probabilistic nature. In this sense, when we claim that a person has power over another, the claim is not that the person with power can always and under every circumstance change the incentive structure of the other. The claim indicates merely that one person has some capacity of affecting the other's incentives. Whether this person is always, sometimes, or rarely able to influence the other person depends on the amount of power that she has. The higher the amount of power, then the higher are the chances that she has of changing the incentive structure of the other person to produce a specific outcome. The lower the amount, then the lower is the probability of success in altering the incentive structure of the other. It is not the case that one necessarily has or not 'power over' another. Social power appears in different amounts.

The scope of social power with which we are concerned in our discussion can be extracted from Hobbes' view of the state of nature. Hobbes argues that diffidence arises from the fact that two persons covet at the same time a scarce good and that they are equally powerful. The scarcity of the good means that the desires of both individuals cannot be fulfilled at the same time. Diffidence arises because the one that fails to acquire the good still desires the good and is powerful enough to attempt to get the good from the individual who acquired it. The scope of power that concerns such relation is the power of forcing the other individual to cede the good one desires for oneself. To put it more generally, it is the power of forcing the other individual to act and behave in accordance with one's interests and desires in general. The power relation that stands at the center of interactions between individuals in the state of nature is the power to force other individuals to act in conformity with one's interests, should they refuse to do it. It is to such scope of power that the equality of powers view is connected. We define, then, the scope of the power relations we have in mind not by reference to particular actions or behavior of the coerced individual, but as the set of actions and behaviors whose performance by the coerced individual are required for the fulfillment of the agent's interests and desires.

It is with regards to the scope of power above considered that one's amount of power should be understood in the case of the circumstance of equality of powers. Dahl connects the amount of

168 See Dahl (1957), pp. 203-204.
power of an individual to the probability of successfully forcing another individual to perform an action that he would not otherwise perform. In this definition, power is not something that a person either has or not. Persons have power on different levels. In our discussion, the levels of power refer to the power of coercing another to act in accordance with one's interests and desires. Then, the amount of power an individual has is determined by his odds of successfully forcing the other individual to act in conformity with his interests and desires, should the other individual refuse to do so voluntarily. The higher the odds, the more powerful the individual is.

Downing's and Dahl's discussions of the concept of power are useful for the discussion that follows. The type of power that is under consideration when we discuss the circumstances of moral cooperation is social power. It is the power to change the other's incentives deliberately so that a specific outcome is produced. Such power involves four factors, a basis, a means, an amount, and a scope. In the context of the equality of powers view, the scope of power that is relevant is that of forcing other individuals to produce the outcomes that one desires. As for the amount of power, we saw that power has a probabilistic element. One's social power indicates the odds that one has of successfully influencing the incentive structure of other individuals to act as they require. Dahl's idea of amount and scope of power can be used to define the view of equality of powers. What Hume and Hobbes argue is, in essence, that a requirement for the emergence of moral cooperation is that individuals have roughly equal amounts of power over one another to alter the incentive structure of each other in order to produce outcomes they desire. With such considerations in mind, we can now move to the analysis of the circumstance of equality of powers.

### 2.2.3 The Problem with the Equality of Powers View

The idea of equality of powers endorsed by Hume and the others, we believe, is flawed. It cannot correctly represent a circumstance of moral cooperation nor a circumstance of justice. The equality of powers view is overly individualistic. It states that the amount of power of each individual must be roughly equal in order for moral cooperation to be both possible and necessary for them. It is important to notice, however, that the suggestion that individuals' social powers must be equal is a reference to a more fundamental factor, namely the absence of *ascendancy* in the system.

Dahl argued that the probability one has of successfully forcing the other person to do something that she would not have done otherwise represents one's amount of power. Consider two groups, Athens and Mytilene. The social power of Athens over Mytilene is such that Athens is always able to alter the incentive structure of Mytilene. Whenever Athens wants to move Mytilene
to produce a specific outcome, it can do it successfully. Mytilene, on the other hand, has no social power over Athens. Mytilene is entirely unable to change the incentive structure of Athens to produce outcomes that it desires. In such a scenario, we may say that Athens has ascendancy over Mytilene. If we were to consider moral cooperation between Athens and Mytilene, such cooperation would be impossible. The discrepancy in social power between the two is so vast that Athens would have no gain in participating in a moral framework with Mytilene. Recall that in both Hobbes' and Hume's arguments, the rationale for moral cooperation arises from its usefulness for individuals. Morality is useful because, without it, individuals are in a suboptimal state regarding the promotion of their interests. If Athens has ascendancy over Mytilene, then there is no suboptimality for Athens. Athens can exact its will upon Mytilene whenever it desires, and Mytilene cannot do the same in return to Athens. Mytilene is unable to threaten Athens' in any meaningful way. For this reason, the introduction of a moral framework that would constrain Athens and Mytilene equally in their interactions would be prejudicial to Athens, although beneficial to Mytilene. In this sense, Athens would regard a moral agreement in the same manner that the strong regard Gauthier's proviso as we argued before, i.e., as an unjustified normative constraint.169

Ascendancy is the central aspect of the circumstance of moral cooperation that refers to power. What the view of equality of power attempts to formulate is that moral cooperation is impossible between two individuals when one of the individuals has ascendancy over the other. That is the case of Athens and Mytilene and of humans and the feeble creatures in Hume's thought-experiment. In both cases, one of the parties in the system has overwhelming social power over the other. It can successfully change the structure of incentives of the other whenever it likes, whereas the other party is unable to the same thing in return. Along with Hume's case, the case of Athens and Mytilene depicts an absolute form of ascendancy. Notice, however, that such an absolute level of social power is not a requirement for establishing ascendancy. What must be the case for ascendancy to appear in the social system is the existence of a considerable gap in power between individuals. A considerable gap indicates that an agent has high odds of successfully coercing another agent to acquiesce to his interests and desires in cases where the second agent would not acquiesce voluntarily, and also that the second agent has low odds of doing the same to the first agent.170 That is a consequence of Dahl's probabilistic treatment of power, which we incorporate

169 It is important to keep in mind that ascendancy only hinders moral cooperation when we conceive moral cooperation as we usually do, i.e., involving rules such as the prohibition of murder, private property, and so on. A cooperative framework can appear between an ascendant individual and a weak individual. But such framework will not contain the rules we mentioned. The terms of cooperation are dictated exclusively by the ascendant individual whereas the weak person has to accept the terms merely because of the ascendant individuals' power.

170 One could investigate how large must the gap be in order for ascendancy to appear. Such investigation is not, however, important for our discussion here and we do not pursue it.
into our discussion. An individual's amount of power over another is represented by the odds this individual has of coercing the other individual into doing his bidding. In this sense, ascendancy appears not only when such coercion is always successful, as in Hume's case of humans and feeble creatures, but also when a considerable gap in power appears between two individuals.

Consider the following example. Athos and Aramis are fencing because Athos wants to enter a tavern and Aramis does not allow him to. Regarding social power, they both want to alter the incentive structure of the other to bring about a goal they desire, entering the tavern for Athos and fobbing Athos' entrance for Aramis. If Aramis wins the fencing, that does not mean that he has ascendancy over Athos. Instead, it means that he had a higher amount of power over Athos than Athos had over him. Ascendancy would appear if we considered that Aramis was dueling using plate armor and a steel sword, whereas Athos had ordinary clothes and a wooden stick. In such case, the situation is such that the gap in power between Aramis and Athos is so big that Aramis is highly likely to change Athos' incentive structure, whereas Athos' chances of doing the same to Aramis are small. Such a case, however, is an illustration of a more realistic case of ascendancy, i.e., one that does not involve the absolute ascendancy that we saw in the case of Athens and Mytilene. Athos could win the duel with the well-armored Aramis wearing clothes and wielding a wooden stick. Doing so is not impossible, as it is for Mytilene to affect Athens' incentives. The difficulty is that Athos' chances of successfully changing Aramis' incentive structure are very dim.

What stands in the center of Hobbes' and Hume's' view of equality of powers is that when power is distributed equally, there is not an ascendant individual in the system. Without an ascendant individual in the system, individuals cannot engage in social cooperation in virtue of the natural situation of conflict. Social cooperation can only be achieved, then, with the introduction of a moral agreement. That is the argument suggested by both Hobbes and Hume. Equality of powers, however, is not an indication that there is no ascendancy in the system. And consequently, that moral cooperation is possible and necessary. That is the central problem with this view. We suggest three examples below to demonstrate how equality of powers does not necessarily imply the absence of an ascendant individual.

Consider a society composed of three persons: Anthony, Brutus, and Caesar. In this society, the division of power is such that all three individuals have five power units. Assume that a difference of three power units between two individuals indicate that the individual with a surplus of three power units has ascendancy over the weak individual. A difference of three power units indicates that it is not clear whether moral cooperation between the strong and the weak is rational for the strong. Assume further that all the other circumstances of moral cooperation obtain in this society. Now, this three-person society has a fully egalitarian distribution of power. And since all the
other circumstances of moral cooperation obtain, we should expect the rise of moral cooperation between the three individuals. However, moral cooperation does not necessarily follow from the given circumstances. To observe that, consider the possibility of Anthony and Caesar forming a coalition. A coalition is an alliance between the two for some common goal, such as defense or attack. With the coalition formed, it contains a total amount of ten power units, i.e., five more than Brutus. The surplus of five power units that Anthony and Caesar have over Brutus signifies that they can consider themselves somewhat safe against Brutus, and it enables them also to exact their interests over Brutus with a high margin of success. They acquire ascendancy over Brutus. With the coalition formed, Anthony and Caesar have no reason to engage in moral cooperation with Brutus, for doing so will bring no relevant benefits for both of them. The presence of Brutus does not make the social system unstable since Brutus' power of affecting either Anthony or Caesar is practically nullified with the creation of the coalition. We have, thus, a case in which a social system where all circumstances of moral cooperation obtain between its individuals, including the equality of powers, but in which moral cooperation does not necessarily entail. Moral cooperation fails to appear because ascendancy appears in the system in the form of the coalition between Anthony and Caesar. The ascendancy that the coalition provides both of them turns any prospects of moral cooperation with Brutus unnecessary. As a consequence, moral cooperation cannot appear in this system. Unless we establish an arbitrary restriction on the formation of coalitions, there is no relation of necessity between the equality of powers, the absence of ascendancy, and the occurrence of moral cooperation.

One could immediately claim here that we rely on an instrument, coalitions, that is not available in a Hobbesian state of nature. In a Hobbesian state of nature, rational agents would be unwilling to form coalitions with one another because of the risk of exploitation. Hobbes himself seems to be skeptical about the possibility of coalitions in the state of nature. To enter an offensive or defensive coalition requires a degree of trust between the members of the coalition. The decision to enter or not a coalition is, in essence, a prisoner's dilemma like the one we presented between the two farmers in the first chapter. In this sense, if Caesar decides to start a coalition with Anthony, Caesar must disarm himself in his mutual relations with Anthony. To do this, however, represents a risk for Caesar. If Caesar disarms himself, he necessarily opens up himself to be exploited by Anthony. In virtue of this risk, one could argue, Caesar would not accept to form up a coalition with Anthony. Since coalitions would be shown to be irrational in a state of nature, one should dismiss the introduction of such devices as we did in the example above, and thus the equality of powers view would be saved.

Gregory S. Kavka discusses the theme of coalitions in a Hobbesian state of nature in some length. He argues that the possibility of coalitions in the state of nature are closed if we maintain certain assumptions. The first assumption is that rational agents facing a decision to enter a coalition are facing a one-shot prisoner's dilemma. In such case, Kavka argues, participation in a coalition is indeed irrational. As with every one-shot prisoner's dilemma, to refuse cooperation is a dominant strategy, and thus involvement in the coalition cannot be justified. The second assumption is that rational agents in a state of nature, unlike us, possess extreme levels of knowledge and rationality. Kavka shows that even in an iterated version of the prisoner's dilemma in the context of coalitions, rational agents would defect from cooperation if we maintain strong assumptions regarding the rationality and information that the agents have. In our discussion, however, none of the two assumptions are made. Firstly, we do not have any commitment to the formulation of the state of nature as a one-shot prisoner's dilemma. Instead, we prefer to formulate it in an iterated version. In truth, the examination of the rationale for a coalition is more reasonable to be carried in the context of an iterated prisoner's dilemma rather than a one-shot version. The decision to take part in a coalition or not is made in a dynamic context of competition for goods, i.e., the decision assumes that future interactions with other individuals will take place. Secondly, recall that the circumstance of unsocial sociability indicates that individuals have limited cognitive and rational powers, Hayek's so-called epistemological pessimism. Individuals are indeed assumed to be rational, but their cognitive capacities to act rationally are limited. Although they always act in the most advantageous manner for themselves that is possible, they are ignorant about some aspects that shape the circumstances in which they must decide. Their decisions are often not optimal, given that they do not possess sufficient cognitive powers to examine the circumstances that surround their choices perfectly. Moreover, the definition of coalition hat Hobbes has in mind in his

172 See Kavka (1986), pp. 126-146.
173 See Kavka (1986), p. 130 “Suppose now, that we have a game of iterated prisoner's dilemma of some definite number of moves, n, to be played by two players who are rational and interested solely in maximizing their personal payoffs and who will remain so throughout the game. It is assumed that this is all common knowledge among the players, that is, each knows it, knows the other knows it, knows the other knows he knows it, and so forth. Given the assumptions of self-interest and rationality, and the dominance of noncooperation on a single play, it follows that a player will cooperate on a given play only if he believes that doing so may induce his partner to cooperate on some later play or plays. Since each party, being rational, knows this, each party knows his opponent will not cooperate on the nth (i.e., last) play, for there are no later plays on which cooperation by one's opponent could be induced by a cooperative move on the nth play. But then each party knows that a cooperative move on the n – the 1st play could induce future cooperation by his opponent and, being rational, he will not cooperate on the n – 1st play. But his opponent, knowing this, will have no reason to cooperate on the n – 2nd play and will not so cooperate. By similar reasoning, we work our way back step-by-step (or by mathematical induction) to the very first play and conclude that each party will make the noncooperative move on every play.”
174 The decision to take part in a coalition resembles the decision that the two farmers face, which we explored in chapter one. The farmers cooperate because they foresee the disadvantages they will have in the future if they defect from cooperation. In the same manner, members of a coalition can cooperate because they can anticipate the disadvantages they will have if they betray their coalitions partners.
discussion, and which Kavka considers, is a particular one. In his discussion, a coalition is a defensive agreement based on mutual trust.\textsuperscript{175} It does not involve the introduction of any set of rules to ensure security between the members of the coalition. Under these terms, it is indeed hard to conceive how coalitions can endure in a state of nature. However, we can conceive coalitions differently. Coalitions can be considered to be a limited and primitive form of moral cooperation. We entertained this possibility in chapter two when we discussed Gauthier's Lockean proviso. That is a possibility entertained by James Buchanan.\textsuperscript{176}

In Buchanan's discussion of the emergence of private property, he argues that the natural distribution may be conceived on several levels of aggregation.\textsuperscript{177} A pure natural distribution, Buchanan argues, is the natural distribution that involves every individual in the system, it contains no coalitions. That is the case of the initial moment of our example with Anthony, Brutus, and Caesar. However, as the competition for goods starts, individuals can sign limited constitutional contracts to form coalitions against other individuals or groups in the system, as Anthony and Caesar did in our example. As the conflict between the groups moves through time, it will lead to a 'final' constitutional contract in which the shares of each individual are assigned. In this sense, the coalition between Anthony and Caesar can be considered as a form of cooperation based on a moral agreement. The coalition is supposed to work as a form of cooperation that provides both individuals with security. For this purpose, the coalition must include provisions that are typical of moral agreements, such as a pact of non-aggression, for example. Consequently, the coalition can be characterized as a form of moral agreement because it dissolves in some measure the instability that is typical of pre-moral situations. In our example, the instability is dissolved in two steps. The first step concerns the establishment of security between Anthony and Caesar by the introduction of the coalition and its rules, such as the pact of non-aggression. With it, the rationale for exploitation and competition between the two disappears. The second aspect is the security acquired between the members of the coalition and Brutus. This security is achieved by the power that Anthony and Caesar have together. It minimizes Brutus' capacity of interfering with both of them. Brutus becomes, then, like one of Hume's feeble creatures, who pose no threat to humans, and social stability is thus achieved.

The example of Anthony, Brutus, and Caesar stresses that an equal distribution of power between every individual in the system does not necessarily imply the absence of ascendancy in the system. Utilizing coalitions, ascendant individuals may appear, and thus the possibility of moral

\textsuperscript{175} See Kavka (1986), p. 127.
\textsuperscript{176} The possibility is also present in Hobbes' view. Hobbes discusses covenants of mutual trust in the state of nature. However, he explicitly argues that if a common power is introduced over the members of the coalition, the coalition pact becomes valid and sustainable in the state of nature.
\textsuperscript{177} See Buchanan (1975), pp. 41-43.
cooperation in the system disappears. It is true, nonetheless, that the coalition that emerges represents a form of moral cooperation. It is limited to Anthony and Caesar, but it is, nonetheless, a cooperative framework. However, since Anthony and Caesar have the same amount of power, one could still argue that equality of powers is still a condition for moral cooperation. Although no moral order emerged in the whole system, a limited one emerged between Anthony and Caesar, and this moral order emerged because Anthony and Caesar have equal amounts of power. That is true. Nonetheless, we reaffirm that equality of powers is not a requirement for moral cooperation. We suggest another example to show it.

Consider a three-person society again, but now with a different distribution of power. Assume now that Anthony has four power units, Brutus has seven power units, and Caesar has one power unit. Consider one more time that all circumstances of moral cooperation obtain, with the exception, of course, of the equality of powers. A highly unequal distribution of power marks this society. With this distribution, Brutus has ascendancy over Anthony and Caesar, and Anthony has ascendancy over Caesar. In a Humean or Hobbesian reading of the circumstances of moral cooperation, to engage in moral cooperation with Anthony and Caesar is useless for Brutus, and participating in moral cooperation with Caesar is useless for Anthony. The condition of equality of powers, evidently, is not present in this society, and thus we should conclude that moral cooperation is neither possible nor necessary in this social system. This conclusion, however, is wrong. To see that, we need to entertain the possibility of coalitions again. Given the aforementioned distribution of power, it seems reasonable to suppose that Anthony and Caesar would form a coalition again. To do so enables them to achieve a total of five power units, what leaves them still two power units short of Brutus, but sufficiently powerful to balance Brutus' power and end his ascendancy over both of them. In the same manner of the first case, the coalition between Anthony and Caesar can be considered a limited form of moral cooperation. It creates a state of security and stability between them. However, in this second case, the coalition is unable to ensure Anthony's and Caesar's security against Brutus. Whereas there is security between Anthony and Caesar, insecurity remains present between their coalition and Brutus, since Brutus and the coalition have similar amounts of power. The introduction of a moral agreement that encompasses all three individuals is an

178 We can discuss whether a coalition based on mutual trust between Anthony and Caesar would be more stable than in the first example. In the first example, breaching the coalition leads Anthony and Caesar to an outcome in which they do not have ascendancy over Brutus anymore. Despite that, each remains as powerful as Brutus is. In the second case, a breach of the coalition signifies that Brutus regains his ascendancy over Anthony and Caesar, and thus that both of them become completely exposed to Brutus' power. The consequence for breaching the coalition seems to be much more severe in the second case than in the first. Given such an outcome, it seems reasonable to suppose that the coalition will be much more stable and enduring in the second case than in the first, even if it is based on mutual trust alone. The threat of returning to a condition in which Brutus' power is overwhelming should work as a strong enough reason to make Anthony and Caesar remain loyal to each other in the coalition.
alternative to overcome such insecurity. Such an agreement is what Buchanan has called the 'final constitutional agreement'. It will define the overall cooperative framework that will shape social interaction between the three. Again, unless we stipulate that coalitions cannot be formed, what is an arbitrary and unjustified restriction, moral cooperation arises independently of the presence of an equal distribution of power. This second case stresses with clarity that moral cooperation can appear even in social systems that have a highly unequal distribution of power. It shows that the dynamics of power that underlie the rationale for moral cooperation are more complex and intricate than the idea of equal distribution of power depicts.\(^{179}\)

The second case provides us with material to deal with critiques against contractarianism such as those suggested by Parfit and Nussbaum. The case shows a social system in which moral cooperation involves both a strong individual, Brutus, and a weak individual, Caesar. Although Caesar is very weak in comparison with Anthony and Brutus individually, and \textit{prima facie} that would be a reason to exclude him from the moral community or to incorporate him with a lower normative status, the overall distribution of power makes it beneficial to Anthony and Brutus to engage in moral cooperation with him. Caesar is in the group of individuals that design the cooperative framework and also in the group for whom the framework is developed. His status in the cooperative framework is the same of Anthony and Brutus, even though both of them have ascendancy over him. Such conclusion demonstrates that the exclusion of the weak from the cooperative framework is not a requirement of contractarianism. Considerably weak individuals can be incorporated into the moral community insofar as they help to balance the powers of individuals stronger than them. It is true, nonetheless, that the conclusion does not warrant us to affirm that the inclusion of the weak in the framework is \textit{always} contemplated by contractarianism. The first case that we considered, for example, includes the exclusion of an individual from the framework due to his weakness. The difficulty posed by Parfit and Nussbaum, however, can be drastically reduced if we allow for the introduction of altruism in the model.

Consider a third case. In this case, the distribution of power is such that Anthony has one power unit, Brutus has seven, and Caesar has eight. With such a distribution, it is rational for both Brutus and Caesar to engage in moral cooperation, whilst it is irrational for them both do the same with Anthony. In this case, there is no coalition with Anthony whose formation can be rational.

\(^{179}\) One could launch here, again, the criticism that the equality of powers appears as a circumstance of justice in the sense that cooperation is based on two coalitions with roughly equal powers. Such critique is right. Nonetheless, it does not help Hume, Hobbes or any of the other defenders of the view of equality. They argue explicitly that equality of powers refers to powers between individuals. Hobbes, for example, refers to an equal distribution of physical and mental faculties. What the example shows is that a rough equality of power is important at the level of the coalitions formed, but unimportant at the level of individuals' power. The example demonstrates that we do not need an equal distribution of powers between individuals in order for moral cooperation to be useful for everyone.
either for Caesar or Brutus. What one should expect, then, is for moral cooperation to arise between Brutus and Caesar and Anthony be excluded from it or incorporated with a lower normative status. In this case, contractarianism indeed faces a difficulty to include the weak into the framework. If Anthony is to be included in the moral framework, we must postulate an additional element in the scenario. We can hypothesize, for example, that Caesar and Anthony are close friends, and thus that Anthony's well-being and safety are essential for Caesar. With this postulate, it is reasonable to assume that Caesar would move to form a coalition with Anthony in order to protect him from Brutus and to assure that he would enjoy the benefits of moral cooperation.

To assure that weak individuals that find themselves in social systems like that of the third case will be included in the moral agreement, it is necessary to introduce altruistic factors. As we argued in chapter one, contractarians often dismiss such factors since they are not widespread in individuals. However, it is not because contractarians reject such factors that they ought not to be incorporated in the treatment of particular cases, such as those of moral obligations toward children. But here we can see how altruistic interests can play a role in contractarian arguments. We must not suppose that both Caesar and Brutus have an altruistic interest in Anthony's well-being. To justify the inclusion of Anthony in the moral community, it is enough that Caesar alone is interested in Anthony's well-being. Altruistic interests can appear in the contractarian justification procedure insofar as we do not need to suppose that every individual in the system has such interests in order to use them in the method of justification. It is enough that some individuals in the social system have altruistic interests and that these individuals condition their participation in the moral agreement to the inclusion of those contemplated by their altruistic interests in the contract.

In any case, we do not pursue any attempt of incorporating altruistic interests here. But the possibility we discussed now indicates that contractarianism is not helpless regarding the justification of moral obligations toward weak individuals.

---

180 With the postulate, we assume that Anthony's well-being ranks high among Caesar's preferences, so that Caesar is assumed to lose a high amount of utility whenever Anthony's well-being is negatively affected.

181 This third case constitutes also the basis for a contractarian argument for the moral consideration of non-human animals. As long as certain groups with sufficient power impose as a condition for their participation in the moral agreement the inclusion of obligations toward non-human animals in the contract, we can justify moral duties toward animals. In the same manner that Caesar forces Brutus to accept Anthony into the moral community, a powerful group could demand the inclusion of non-human animals in the moral community from the others members of the community. Andrew Cohen develops a contractarian argument for the moral standing of non-human animals precisely in this line. See Cohen (2007).

182 In some sense, the use of altruistic interests in the procedure of justification, in the manner that we suggested, is similar to Gauthier's argument for the Lockean proviso. Gauthier's central argument for the proviso is that the exploited will withhold their participation in the moral agreement as long as the exploiters have possession of the fruits of exploitation. The 'altruists' employ a similar strategy to the exploited in order to achieve the inclusion of the weak into the moral agreement. They withhold their participation in the agreement to constrain the 'non-altruists' to accept the inclusion of the weak. A difference is that in the case of the proviso, the individuals who withhold the participation in the agreement are the weak. In the case of altruism, those that withhold participation are not the weak themselves, but individuals interested in the well-being of the weak.
One could criticize us by arguing that the use of 'power points' in our examples makes the illustrations presented too abstract. The reference to power units leaves open what exactly are the gains involved in the formation of coalitions. To assert that individuals sum up their 'power points', and thus become more powerful contains no indication of what has effectively changed in the interactions between them. In this sense, when we argue that Anthony and Caesar band up in a coalition against Brutus, and as a consequence Anthony and Caesar emerge with ascendancy over Brutus because they have more power points than him, it remains unclear what do Anthony and Caesar actually gain which creates a considerable gap in power between them and Brutus. As individuals, Anthony, Brutus, and Caesar remain the same as they were before the coalition is established. They suffer no bodily or mental alterations. A question that remains to be answered, then, is what is it about coalitions that enable the appearance of gaps in power between individuals, which did not exist before, when they are formed.

Coalitions can be interpreted as devices to increase or create a power gap between individuals insofar as they are a form of cooperation between individuals that enables certain possibilities for the members of the coalition, which were not available previously. The first clear advantage that coalitions enable is the coordinated use of each individuals' power to a single shared purpose. We can refer here to Dahl's view of the basis of power to better see how this advantage appears. The basis of power is composed of the elements that individuals have at their disposal that can be employed by the means of power to affect other individuals. In a coalition, we can say that individuals augment their basis of power by having at their disposal the assistance and help of other members of the coalition. It is reasonable to think that if Anthony has a chance of forcing Brutus to do something for him, the odds he has of coercing Brutus increase sharply if he has available for him the support of Caesar. In this sense, the power gap between Anthony and Brutus increases when a coalition is formed because Anthony acquires access to Caesar's basis of power, the coalition creates a common basis of power composed by a union between the basis of every member. What changes is that Anthony has access to a basis of power that was not accessible before.

A second significant advantage that the coalition confers to its members is the establishment of a condition of security between them. As we observed with Buchanan, coalitions can be considered to be limited forms of moral cooperation, i.e., 'smaller' constitutional contracts. When they become members of a coalition, individuals do not have to protect themselves against the other members. As a consequence, any effort or goods that they were supposed to invest in defending themselves from one another is freed by the establishment of the coalition to be employed elsewhere. The liberation of goods and effort should be translated, we believe, as an augmentation of the individuals' power concerning those outside of the coalition, since now the individual has...
goods available to use against the outsiders that were not available before. The security established by the coalition can be interpreted as an enlargement of the basis of power of individuals in relation to those outside of the coalition.

In sum, the augmentation of power that coalitions provide can be interpreted in two different and complementary manners. Both of them refer to the augmentation of the basis of power of a member of the coalition in relation to someone outside of the coalition. The first interpretation refers to the fact that the coalition creates a common pool that mixes the basis of power of every individual in the coalition. When an individual joins a coalition, he has at his disposal the basis of power of the other members of the coalition to employ for his purposes. The second interpretation refers to the 'freeing' of goods that were being used against the other members of the coalition before the coalition was established. Since individuals do not have to employ efforts and goods to protect themselves from those that are in the coalition with them, these goods are now a part of their basis of power when dealing with individuals from outside of the coalition.

We can conclude, then, that the view that equality of powers is a circumstance of moral cooperation is false. The three cases we examined above seem to be sufficient to endorse a rejection of the view. The equality view is too narrow to accurately capture the role that power has as a circumstance of moral cooperation. What the cases show, and in particular the second case, is that the power element is better represented by an image of equilibrium or balance between every individual in the system. The balance can be one based on an equal distribution of power among every individual in the system, but it must not be necessarily so, as the second and third cases show. What must be in place in the social order if individuals are to find it rational to engage in moral cooperation is a balance of power, an equilibrium in which no individual can to interfere with the others and be sure that the others are unable to do the same to him. What we suggest, then, is that the presence of a social balance of power, a situation in which the power of each individual is somewhat in equilibrium with that of the others, is one of the circumstances of moral cooperation.\textsuperscript{183}

\textbf{2.2.4 A Balance of Power}

The concept of a balance of power is native to the field of international relations.\textsuperscript{184} It has been long used to describe some aspects of the dynamics of power between states in the international realm.\textsuperscript{185}

\textsuperscript{183} An interpretation of Hume's condition of 'rough equality of powers' in this line appears is suggested by Peter Vanderschraff. See Vanderschraff (2006), p. 331.

\textsuperscript{184} See Little (2007) for an in-depth analysis of the concept in the field of international relations.

\textsuperscript{185} For an interesting formalization of the balance of power as a system, see Chatterjee (1972).
Authors in the field, however, employed the concept with a variety of different meanings.186 For our discussion here, we understand the balance of power to be the state of affairs in which individuals are arranged in coalitions in such a manner that no individual has ascendancy. The arrangement does not require an equal division of power at the individual level. Inequalities in power may be present insofar as individuals are arranged in coalitions in a manner that nullifies substantive inequalities between them.

The idea of balance and equilibrium naturally evokes a sense of equality. As our examples above demonstrate, equality must be present in the system in a particular form in order for moral cooperation to be rational for everyone. It must be present in the arrangement of coalitions. Although considerable inequality in powers can be present between individuals, the same cannot happen for coalitions. Moral cooperation is only rational for everyone when individuals are arranged in coalitions that have roughly equal amounts of power. It is when such arrangement is present that a balance of power is present and, consequently, it is rational for every individual to cooperate with every other individual, regardless of the specific amount of power that each individual has.187

If we consider individuals in a state of nature, it is always the case that either there is or there is not a balance of power in the system. If there is a balance of power, and the other conditions of moral cooperation obtain, moral cooperation is rational for every individual in the system. If there is not a balance of power, it is necessarily the case that the individuals in the system are rearranging themselves to establish an equilibrium. As long as the equilibrium does not come about, certain individuals will regard moral cooperation as irrational, since they are in a position to use their power to obtain advantages at the expenses of other individuals. It must be noted that the absence of a balance of power is always a temporary condition. The social system is a balance of power system. When equilibrium is absent, it always rearranges itself to produce it. It is during such 'rearrangement time' that exploitation and attack are profitable and thus rational for some individuals, for there are then gaps in power considerable enough to make some individuals less vulnerable to retaliation from the others. There is, however, a single possibility where equilibrium is established but moral cooperation is not possible. That is the case of ascendancy, when there is considerable power gap between a single individual and the other individuals in the system, either individually or combined. In this case, the situation of equilibrium is one in which the powerful

---

186 See Claude (1962), pp. 13-28, for a discussion of the meanings attributed to the concept of balance of power.
187 Our reformulation of equality of powers into a balance of power should be seen as a mere adjustment of traditional view that is present in the social contract tradition. We do not abandon completely the idea of equality of powers, as we argued, but added more complexity to the idea by introducing the possibility of coalitions. Although we dismiss equality at the individual level as a relevant factor for the rationale of moral cooperation, we still maintain that equality is a fundamental factor for such rationale at the level of the coalitions.
individual has control over all the others in the system. Notwithstanding this exception, the state of nature is a balance of power system that always tends to establish an equilibrium. To see that, consider the following example.

Consider a state of nature with four individuals, Frodo, Sam, Merry, and Pippin. They all have the same level of power and are engaged in a competition for goods. They compete among themselves until they reach an equilibrium in the division of goods. Because they have similar amounts of power, the equilibrium situation is such that they have roughly equal shares of goods. When Frodo notices that equilibrium has been reached and he cannot improve his situation further, he approaches Sam and proposes a coalition. Forming the coalition improves both Frodo and Sam's power, what allows both of them to acquire more goods and thus break the equilibrium established and move the distribution to another equilibrium that is more favorable to them. From Sam's perspective, accepting the coalition is rational, since it allows him to augment his share of goods. However, when Merry notices that a coalition between Frodo and Sam was formed, he finds himself in a crossroad. If Merry remains idle, then Frodo and Sam will move to augment their share of goods at his and Pippin's expenses. To prevent a loss of goods, Merry must either approach Frodo and Sam to bargain an entrance in their coalition or he must approach Pippin to establish a coalition with him to balance Frodo's and Sam's newly acquired social power. To bargain with Frodo and Sam is unlikely to be a profitable endeavor, at least in comparison with the prospect of making a coalition with Pippin. Because of the difference in power between Merry and the coalition of Frodo and Sam, it is likely that Merry will have to concede some goods to them in order to join the coalition, what means that Merry will be worse than he was in the equilibrium situation. A coalition with Pippin, on the other hand, is likely to assure a better outcome for Merry. The formation of such a coalition will likely not demand a worsening of Merry's situation. Pippin finds himself in the same situation of Merry. He must either bargain his entrance in the coalition or he must band up with Merry and thus form a coalition to balance against Frodo and Sam. As it is the case with Merry, to enter Frodo and Sam's coalition is likely to be unprofitable for Pippin. The formation of a coalition between Merry and Pippin, thus, seems to be the most rational option for both Merry and Pippin. When this coalition is formed, a new equilibrium is reached. The combined power of Merry and Pippin balances that of Frodo and Sam.

One could discuss here the terms on which this coalition is established. One could introduce a problem regarding how the surplus of goods that will be acquired by the new level of power reached should be divided between the members of the coalition. This problem does not pertain the discussion, so we will ignore it. As long as each individual in the coalition augments his share, even if he does so for a small margin, participation in the coalition is rational for him.

The distribution of goods under the new pattern may be the same as that of the original equilibrium, but it may also be different by reason of the bargains that the four individuals have done to form their coalitions.
The Hobbesian state of nature is a balance of power system in the sense that it is a system that moves to a state of balance naturally. Such characteristic of the system is grounded on individuals' attitude toward power. Because of their interests and limitations, each individual in the system seeks to increase in his favor the gap of power between himself and the others. However, as the case above demonstrates, this tendency has two aspects. The first aspect is the augmentation of one's power in comparison with that of the others. Frodo and Sam represent this aspect. The second aspect appears in how Merry and Pippin act. The second aspect concerns actions to reduce the gap in power between oneself and the others. As soon as a considerable gap in power appears, those at the bottom of the power distribution will move to close the gap. To not do so is, given their interests and limitations, irrational. That is how a state of nature situation is a balance of power system. The units in the system, the individuals, will always move to establish an equilibrium in the system, not because they desire the equilibrium itself, but because they seek to break the equilibrium in their favor. Whenever the equilibrium is broken, those disfavored by the new status quo will move to restore it. To establish or restore equilibrium to the system means, in essence, to introduce a balance of power between those in the system. Power is in balance when the equilibrium occurs, i.e., when no individual or group has ascendancy over the others and is thus not capable of effectively interfering with the other's affairs without being vulnerable to the same thing when perpetrated by other individuals. Notice, however, that the establishment of a balance of power does not mean that individuals are unable to interfere with each other. What it represents is a state of affairs in which individuals are mutually vulnerable to interference, and no one can consider oneself safe from the others. It is precisely this balance of power, meant as a state of affairs, that serves as a proper image for the second condition of moral cooperation.

We may now conclude the analysis of the second circumstance of moral cooperation. We propose to substitute the traditional image of the rough equality of powers to one of a social balance of power. As argued, an egalitarian distribution of power between individuals seems insufficient to assure that moral cooperation entails between them. One must assume that individuals are forbidden to establish coalitions with one another in order to claim that equality of powers is a circumstance of moral cooperation, which is, of course, an unwarranted assumption. Equality of powers is also associated with the necessary exclusion of weak individuals from the framework of morality. Some philosophers argue that contractarianism cannot account for moral obligations toward weak individuals if it is assumed that equality of powers is a circumstance of justice. The substitution of the equality view for the image of the balance of power enables us to overcome the restriction on coalitions with which the equality view must commit itself. Moreover, it also improves the position of contractarianism to address the complaint that contractarianism excludes the weak from the
moral community. As a circumstance of moral cooperation, the balance of power states that moral cooperation is possible and necessary when no individual or group has ascendancy over the others since the power of every individual or group is in check by the power of another individual or group. The amount of power that the single individual has is as significant as the coalitions he can form to protect himself. Through coalitions, individuals can stand up against stronger individuals, and in this manner, the weak have an opportunity of being accepted into the moral community, i.e., as balancers that help individuals stronger than them to put the power of even stronger individuals in check.

2.3 ABSENCE OF WIDESPREAD POVERTY

We come now to consider the third and last circumstance of moral cooperation. It refers to the availability of goods in the environment where the social system is. Hobbes wrote about diffidence arising from common desires for a good that cannot be shared. Hume referred to this circumstance as the scarcity of provisions and the easy exchangeability of goods, Hart to the truism of limited resources, and Rawls referred to the condition of moderate scarcity of goods. All such philosophers shared the intuition that the availability of goods is an essential factor for the emergence of cooperation. The intuition they all had is that although persons have an infinite number of interests and desires, the goods they require to fulfill those interests and desires are scarce. And because they are scarce, people will fight over them. We share the same intuition. In this section, we formulate such intuition in the form of a circumstance of moral cooperation in a slightly different form than that formulated by the philosophers we mentioned. The difference of our formulation is that we introduce a conception of poverty into the formulation.

2.3.1 EXTREME SCARCITY AND EXTREME ABUNDANCE

Hume refers to moderate scarcity as a circumstance of justice. Moderate scarcity refers to a middle area in a spectrum with two extremes, namely extreme scarcity and extreme abundance. Hume, Hart, and Rawls believed that moral cooperation would be unnecessary and impossible in a system where goods were either extremely scarce or extremely abundant. Extreme scarcity is the condition in which individuals – a majority of them – suffer deprivation. Extreme abundance, on the other hand, is a circumstance in which so much of everything is available that individuals do not come in
conflict with one another to obtain the goods they desire. Whenever the environment is such that there is neither extreme scarcity nor extreme abundance, moral cooperation is both possible and necessary. The scarcity that works as a circumstance of moral cooperation is, then, to be understood as this middle ground between the extremes. It is easy to see how the extremes of the spectrum hinder moral cooperation by depicting moral cooperation as a matter of agreement on distributional patterns of goods.

Consider a system with two individuals, Gaius and Lucius, and two units of a particular good. Moreover, consider that Gaius and Lucius meet to start a moral agreement. In order to cooperate, they must first select a distributional baseline of the goods. It is on top of such a baseline that cooperation will take place. The selection of the baseline obeys two conditions. The baseline selected must, firstly, be such that moral cooperation over it does not make anyone worse off in comparison to how the person is in the state of nature. If this condition fails, at least one of the individuals will prefer to remain in the pre-moral state and hence to refuse participation in the moral agreement. The second condition is that the moral agreement must improve the condition the individuals, i.e., it must improve individuals' standings in comparison to how they are in the state of nature. If the first condition obtains and the second one does not, we have a situation in which both individuals are indifferent to the establishment of a moral agreement, and as a consequence, we have a situation in which there is no clear rationale for moral cooperation. Now suppose that the other two circumstances of moral cooperation obtain, and also suppose that the good in question is a vital good that both Gaius and Lucius require, so that deprivation of this good leads the individual to death. Given these assumptions, there is only one baseline that can sustain moral cooperation. It is the pattern in which Gaius and Lucius have one good each. In all other baselines, at least one individual will refuse to engage in moral cooperation, and will thus prefer to remain in the natural condition. The individual who would refuse the agreement refuses it for the following reason. Moral cooperation is to be understood as the establishment of restrictions on the means of power that individuals have, recalling Dahl's concept of power. When compared to the state of nature, individuals in the moral state have their means to acquire goods belonging to other individuals severely restricted. A moral agreement is to be interpreted as raising of 'tall walls' around each individual's share of goods. Consequently, to enter a moral condition means for a deprived individual entering a condition in which the acquisition of goods becomes more difficult than it was in the state of nature. If we assume that the balance of power obtains in the state of nature, and power is a matter of how likely one can coerce another into doing his bidding, we may infer that a

There are six different possible distributional baselines for this system. Either Gaius has the two units and Lucius none, or Gaius has one and Lucius has none, or Lucius has two and Gaius has zero, or Lucius has one and Gaius has none, or both Gaius and Lucius have one, or both Gaius and Lucius have zero.
deprived individual will always retain some likelihood of escaping deprivation through attacks against other persons. Conversely, if the deprived person is not allowed to use attacks as a means of acquiring the goods she desires, it is reasonable to suppose that her chances of escaping poverty diminish. A moral agreement represents for the deprived a reduction in their chances of escaping poverty. If the deprived are rational, then, they must reject the agreement.

Extreme scarcity appears when no baseline can sustain moral cooperation. When such a situation occurs, individuals find themselves in an environment with so few essential goods that in every possible pattern there will always be a significant majority of individuals that will refuse moral cooperation and prefer to remain in the conflict over goods and thus in the state of nature. Consider the following system. In a system, there are four individuals and four units available of the good R. Consider additionally, that every individual needs to possess at least three units of R to fulfill his most basic needs. There are several possible baselines for this system. Nonetheless, in none of them it is possible to provide three units of R to more than one individual, so that every configuration of the pattern in which one individual has three units is unacceptable to the other individuals who have either nothing or one unit of R. This is why extreme scarcity hinders moral cooperation. In such circumstances, moral cooperation is useless for a large share of the individuals in the social system.

Hume also believes that extreme abundance has the same effect over the usefulness of moral cooperation.\textsuperscript{191} We do not, however, follow Hume in this respect. Extreme abundance is a condition in which the overwhelming majority of possible baselines is acceptable for every individual in the system. It is a very different case from that of severe scarcity. What happens in the case of extreme abundance is that the matter of selecting a baseline is redundant and virtually useless since practically every baseline will please every individual. So many goods are available that individuals have no reason to struggle with one another for them. Since conflict is absent, they require no moral agreement to pacify and stabilize their interactions. That is how Hume seems to see the situation. Hume's argument is formally correct. However, the situation of abundance that it refers to is hugely counter-factual. Hume's discussion of abundance leads us to believe that he is speaking of goods such as food, houses, land, and so on. Nonetheless, if the abundance about which Hume writes refers exclusively to such goods, it will not hinder moral cooperation from arising. Human desires are not restricted to external goods connected with survival only; they also refer to more abstract and complex goods – such as social recognition or aesthetic recognition, for example. In this sense, if abundance is to make moral cooperation useless, it must involve an abundant provision of \textit{every}

\textsuperscript{191} See Hume, M 3.8, SBN, pp. 186-187.
type of good, including these more abstract and complex goods.\textsuperscript{192}

\subsection*{2.3.2 Poverty}

The critical factor that hinders moral cooperation is extreme scarcity and not extreme abundance. However, it is not scarcity of every good that makes moral cooperation useless for rational agents. If two persons want to buy the last blue shirt in a store, they will not conflict over it seriously. The one who manages to get the shirt first will be able to keep it without having to appeal to the system of sanctions of the social scheme to protect his acquisition. On the other hand, if two starving individuals desire the last apple available on a tree, the individual left without the apple will not care that the other person has acquired the apple first. The person deprived of the apple will fight for it and will take it from its holder if she can. It is in connection with such observations that we introduce a new element in the formulation of the intuition that the availability of goods is a crucial factor for cooperation. In our view, the goods whose severe scarcity make moral cooperation useless are goods that individuals require to be free from poverty. It is the scarcity of goods connected with poverty that we believe to be connected with the impossibility of moral cooperation.

We connect the idea of poverty with that of deprivation. A deprived person is a person whose life is at risk. The source of the risk that individuals face lies in the inability to command goods to fulfill one's basic needs. Inability to fulfill one's basic needs is what poverty denotes. A basic need is, in its turn, a biological need whose unfulfillment causes a malfunction in the organism and, ultimately, death. It is a need whose fulfillment is connected with the proper functioning of the body. Moreover, the unfulfillment of the need not only leads the organism to death but also produces suffering and pain during the process. Examples of basic needs are the needs for food, water, shelter, and protection against the weather.

Poverty, under the definition given above, is something that we may suppose every rational agent wants to avoid. Due to the connection between basic needs and poverty, it seems reasonable to affirm that every individual is as interested in being free from poverty as he is in not being murdered or robbed. To be free from poverty, we believe, is a fundamental general interest in the sense proposed by Stemmer, which we considered in the first chapter.\textsuperscript{193} A fundamental general interest in being free from poverty can be explained further by reference to Rawls' idea of basic

\textsuperscript{192} Unless we entertain some science fiction hypothesis, such as Robert Nozick's thought-experiment of the experience machine, it seems hard to see how such abundance is possible. See Nozick (1974), pp. 42-45.

\textsuperscript{193} Stemmer actually suggests specific fundamental general interests that could be interpreted as a fundamental general interest in being free from poverty, when considered together. They are the interests in food, clothing, and protection from the weather and environment. See Stemmer (2000), p. 197.
primary goods. The concept of basic primary goods is complementary to the concept of fundamental general interests. Rawls argues that although individuals have wildly different tastes, preferences, and projects of life, some goods are necessary for the realization of every project of life – or conception of the good, as Rawls calls it. A primary good is a good without which no project of life can be realized. Because they are so important, it is reasonable to assume that every rational agent desires these goods, irrespectively of the agent's particular conception of the good. That is the case, we believe, with the goods individuals require to be free from poverty. Without such goods, no particular conception of the good can be realized. We may refer, then, to an abstract primary good called 'freedom from poverty'. A person has freedom from poverty whenever she has the means to command the goods she requires to fulfill her basic needs. To lack freedom from poverty is to lack such means.

An environment where freedom from poverty is hugely scarce is an environment where the goods required for the fulfillment of basic needs are also exceptionally scarce. In such an environment, the majority of individuals in the social system is poor and destitute. They are unable to fulfill their basic needs, and hence they are engaged in a fierce struggle for the few goods available, so that they may acquire the freedom from poverty they want. As we argued before, to introduce moral cooperation in such an environment is to put the goods the destitute need entirely out of their reach. Provided the other circumstances of moral cooperation obtain, an individual in the natural condition always has some chances of acquiring the goods to provide himself with freedom from poverty. Such chances depend, among other factors, on the use of coercion, violence, threats, and fraud. For if extreme scarcity of freedom from poverty is the case, the few individuals in possession of freedom from poverty are certainly not willing to give their goods away voluntarily to the others. The goods can only be acquired against the will and interests of those that have them. To establish a moral agreement in such circumstances is, from the perspective of the destitute, to rule out of the table the only means of acquiring the goods that the destitute have. Hence taking part in moral cooperation must be irrational for the poor.

One could claim that extreme scarcity of other types of good can make moral cooperation irrational in the same manner that scarcity of freedom from poverty does. However, it is hard to see what goods those could be. There are other primary goods beyond freedom from poverty, that is correct. We can conceive of primary goods associated with every fundamental general interest. In this respect, we can conceive primary goods such as 'freedom from violence' or 'safety from theft', for example. However, such other primary goods have an essential difference in relation to freedom from poverty. They are goods that are naturally scarce in any conceivable state of nature. In a state

194 See Rawls (1999), p. 79.
of nature, primary goods such as 'safety from theft' or a 'right to help in emergencies' are virtually nonexistent because such goods are products of moral rules, which do not exist in the state of nature.\textsuperscript{195} The availability of freedom from poverty, on the other hand, is not dependent on the introduction of moral norms as the others primary goods are.\textsuperscript{196} What determines whether freedom from poverty is scarce or not is the amount of external goods associated with the fulfillment of basic needs that is available. For this reason, it seems to be the case that it is only scarcity of freedom from poverty that may hinder moral cooperation.

We may, then, claim that extreme scarcity thwarts moral cooperation when it is about freedom from poverty. But despite the adverse effects that poverty has over the rationale for cooperation, it does not work as a hindrance to moral cooperation in every case.\textsuperscript{197} The threat that poverty poses to cooperation is entirely dependent on how widespread in the social system it is. In a system with millions of individuals in which a single individual is poor, the general rationale toward cooperation is not affected. Although the single poor individual feels discontent with cooperating on terms that leave him in poverty, there are no other options for him. If he refuses cooperation, he will find himself in a state of nature facing a coalition – a moral order – composed by millions of individuals. Regarding his poverty, refusing or accepting the agreement makes no difference whatsoever. With or without the agreement, he remains unable to escape poverty. However, as the number of poor persons in the system increase, the effects of poverty over the rationale for cooperation start to change.

The more poor individuals there are in the system, the more significant is their refusal to cooperate. In a system with ten individuals, if five of them are poor, then it becomes unclear whether a moral order that encompasses every person in the system can emerge. The five poor individuals represent a considerable amount of social power, an amount that can thwart the emergence of a moral order and thus trap the other five individuals in the state of nature, then forcing a redistribution of goods, if such a thing is possible. The amount of power that the poor hold collectively is determinant for the viability of cooperation. If they have similar or more power than the non-poor, they can thwart moral cooperation. For this reason, it is not the occurrence of poverty

\textsuperscript{195}The creation of such goods could be listed as one of the reasons that lead individuals to engage in moral cooperation. It is only through moral cooperation that individuals can promote some of their fundamental general interests adequately. The promotion is done precisely by the creation of such primary goods. The promotion of the fundamental general interest in not being killed, for example, is effectively carried out when individuals introduce a moral rule that forbids murder, for example. With the rule, every individual acquires a right not to be murdered that he cannot have outside of the moral condition.

\textsuperscript{196}The supply of goods and resources that grant individuals freedom from poverty can be augmented through the moral agreement, however. The moral agreement enables individuals to cooperate more efficiently and, consequently, it allows them to be more productive. More productive individuals lead to a more productive society, which, in its turn, leads to a situation of more abundance of resources and goods.

\textsuperscript{197}In the next chapter, with analyze with more precision and depth the effects that poverty have over the rationale of cooperation.
in general that hinders moral cooperation, but instead the occurrence of widespread poverty, i.e., the occurrence of poverty among a considerable number of individuals in the social system.

Some authors have argued that what Hume called moderate scarcity should be referred to as variable moderate scarcity instead. Such a suggestion can be found, for example, in David Gauthier's *Morals by Agreement*. He argues that if scarcity could not be alleviated by the cooperation that entails from a moral framework, then morality would not pertain to the provision of mutual benefits, but only the avoidance of mutually destructive conflicts. Gauthier's argument is correct. Scarcity can be alleviated by cooperation because cooperation allows individuals to become more productive. Through cooperation, individuals can increase the overall amount of goods available in their social system. However, the variability of scarcity cannot be considered an essential aspect of the circumstances of moral cooperation. These are the circumstances that make moral cooperation rational for every individual, and if the only thing morality would afford individuals were the avoidance of mutually destructive conflicts, this provision would already make moral cooperation rational for everyone, even if scarcity was not variable. The variability of scarcity merely strengthens the rationale for cooperation, but it is not an indispensable condition of it.

We may now conclude the exposition of the third circumstance of moral cooperation. Hume, Hart, and Rawls wrote about the scarcity of provisions as one of the circumstances of justice. In our analysis, we concluded that this condition is better expressed as the absence of widespread poverty. The first thing we noticed is that it is only extreme scarcity that provides a definitive hindrance to moral cooperation. Extreme abundance makes moral cooperation superfluous, but it does not hinder it necessarily. We then suggested that it is not the extreme scarcity of every type of good that is a hindrance to moral cooperation, but that it is the extreme scarcity of freedom from poverty that fulfills this role. Hume's condition of moderate scarcity, we concluded, is better referred to as the absence of widespread poverty.

### 2.4 The Three Circumstances of Moral Cooperation

We have now what a complete contractarian interpretation of the circumstances of justice, which we called the circumstances of moral cooperation. The circumstances of moral cooperation are the circumstances that make moral cooperation both possible and necessary for every individual in the social system. In other words, they are the circumstances that ensure that moral cooperation is

---

rational for every individual in the system. There are three such circumstances. The circumstance of unsocial sociability is the first one. This circumstance states that social cooperation is beneficial for individuals, i.e., they gain when they cooperate. However, despite the advantages of social cooperation, individuals face hardships to cooperate due to unsocial tendencies they have. The second circumstance is the presence of a social balance of power. The second circumstance states that there must be a balance of power in the social system so that no individual or group has ascendancy over another individual or group. We argued that the image of the equality of power fails to define how power is relevant to ensure moral cooperation properly. The image of a balance of power is more suitable to fulfill this role. Lastly, the third circumstance is the absence of widespread poverty. We argued that it is the severe scarcity of specific primary goods, freedom from poverty, that hinders moral cooperation. If other goods were scarce, individuals would still regard moral cooperation as rational.

3 CRITIQUES

We move now to consider critiques that were launched against the idea of the circumstances of justice. D. Clayton Hubin and Peter Vanderschraff suggested the critiques we examine ahead. Hubin criticizes Hume's account of the circumstances in the belief that they do limit the scope of justice, as Hume seems to believe. Vanderschraff, on the other hand, believes that Hume's account of the circumstances cannot be considered to make moral cooperation both possible and necessary.199

3.1 D. CLAYTON HUBIN'S CRITIQUE AND THE JUSTIFICATIONAL INTERPRETATION

Hubin's discussion concerns the scope of justice.200 He argues that Hume's circumstances constitute an attempt to restrict the field of justice. The idea is that it is only in the circumstances of justice that justice can be present. It is not clear, however, in which sense the circumstances limit justice.

199 From now on we will refer to ‘justice’ and ‘moral cooperation’ interchangeably.
200 It is important to stress one aspect of Hubin's argument. Hubin's discussion is partially a hermeneutic discussion about Hume's work – a discussion about what Hume meant with the circumstances of justice – and partially an analytic discussion about the idea of the circumstances. What concerns us here is the analytic part of Hubin's discussion. In this sense, Hubin's aims to show that the circumstances of justice limit the scope of justice in neither of the five possible interpretations of the circumstances, but instead that we can speak of justice in contexts where the circumstances are not present, regardless of how they are interpreted. What Hubin is arguing, then, is that the circumstances of justice, as Hume describes them, do not limit the scope of justice.
Hubin believes that the relation between the circumstances and scope of justice can be interpreted in five different senses. The possible interpretations are: logical, epistemic, ontological, deontic, and the utility interpretation. We consider each interpretation and, in the end, suggest a new one.

The logical interpretation states that the concept of justice logically presupposes the existence of the circumstances. In this sense, to apply a concept connected with that of justice – such as unjust or just – in a situation in which the circumstances do not obtain would be a mistake. Hubin believes this interpretation does not limit the scope of justice. To argue for that, he refers to two hypothetical cases. The first case is about people trapped in a lifeboat with scarce supplies, and the second case refers to the society of ruffians that Hume himself suggests. In the first case, Hubin argues that waiting for the others to sleep and then killing them to use their bodies as food would be an 'unjust way' of solving the shortage problem, whereas drawing lots to determine which persons should survive is a 'just resolution'. As for the case of a society of ruffians, Hubin considers the prospect of some people working hard to create valuable objects from the goods available. Given such prospects, Hubin claims that he 'feels no temptation' of stating that justice regarding private property is inapplicable in this society. Both cases are suggested to demonstrate that even when the circumstances of justice do not obtain, it is possible to make sense of the concept of justice.

In general, we agree with Hubin's conclusion. The circumstances of moral cooperation do not limit the scope of morality under the logical interpretation. Despite this, we think that Hubin's two hypothetical cases do not warrant him to draw this conclusion. Let us see first where Hubin is wrong. In the first case, Hubin argues that it is still possible to make sense of the idea of justice there is extreme scarcity occurs. However, as we saw in chapter two, justice is defined by the agreement. We can only speak of justice and injustice by reference to an agreement that defines the standard of justice. In the lifeboat case, there are, prima facie, no moral rules in place. So that when an individual claims that 'killing the others on the boat is unjust', he has no moral rules to refer in order to substantiate the justice about which he is speaking. Without moral rules available, the individual's claim amounts to something similar to 'killing the others on the boat is something that displeases me'. With no moral rules to appeal to, it is difficult to see how one can make sense of the concept of justice. The use of a random procedure on the lifeboat, nonetheless, can be said to be just, if every individual agrees to it. In the contractarian view, as we saw in chapter two, what defines whether a rule is just is the fact that individuals agree with the rule. However, in the case of

203 See Hume, M 3.9, SBN, p. 187.
204 We assume here that the lifeboat case represents a state of nature scenario in Hubin's case.
the ruffians, it is hard to see how one could speak about justice in the case. Since there are no moral rules in place, there is no manner of meaningfully argue about private property in that society.

Hubin's conclusion, however, seems right to us. To see that, we must stress the difference between two hypothetical scenarios. The first scenario is that of the traditional state of nature, the original condition of individuals. On the other hand, the second scenario is that in which the circumstances of moral cooperation were in place in the past but ceased to be at some moment. The first scenario indeed limits the scope of justice in the logical interpretation. However, the second scenario does not. In the second scenario, individuals have the ideas and concepts of morality and justice because they had norms and rules in the past. The presence of the circumstances does not necessarily limit the scope of justice. People can still make sense of the concept of justice in situations where the circumstances are not present, provided they existed in the past.

The second interpretation is epistemological. With the epistemic interpretation, the circumstances of justice refer to conditions that must be in place for a society to develop the concepts connected with justice and morality. Regarding this interpretation, we also agree with Hubin's argument. Hubin argues that the interpretation cannot restrict the scope of justice. Hubin argues that the factual development of concepts does not depend on the existence of actual circumstances in which the concept has some practical utility or applicability, as in the case of justice. The reason for that is that intelligent beings with the power of imagination can conceive concepts based on views of the world that are not part of their immediate experience. In the same manner that Hume can imagine the inapplicability of justice under the conditions of extreme scarcity, beings that lived under such conditions could imagine their life under the conditions of moderate scarcity and thus some concept of justice.\(^{205}\)

The ontological interpretation is the third one. It states that institutions of justice can only exist in environments where the circumstances obtain. Hubin divides this interpretation into two others. One is what he calls the *genetic presuppositional model*, which states that institutions of justice cannot develop where the circumstances of justice are not present. The other is the *sustaining presuppositional model*, which states that institutions of justice can exist only under the circumstances of justice. Hubin argues that in both interpretations, the circumstances of justice do not limit the scope of justice.\(^{206}\) He believes that institutions of justice can originate and be sustained in conditions where the circumstances of justice do not obtain. He dismisses two reasons for why one could endorse the epistemological interpretation of the circumstances. The first reason is that without the circumstances of justice, no institution of justice would arise because such institutions

---


would not be socially useful. Against this reason, Hubin argues that many institutions originate and are sustained in conditions which are detrimental to human well-being. A second reason is that institutions of justice can only be effective when the circumstances of justice are in place. Against this second reason, Hubin recalls the circumstance of the equality of powers. He argues that if we imagine a society of persons similar to average humans, who are neither wholly selfish nor wholly selfless, institutions of justice could continue to exist even if some individuals were invulnerable to the rest under one condition. The condition Hubin has in mind is that those who are invulnerable must support the institutions.

We partially agree with Hubin's conclusions regarding the ontological interpretation. Hubin does not specify what are the 'institutions of justice' that he writes about, and that causes some vagueness in his discussion. On the one side, the institutions of justice in Hubin's mind can be the institutions that we usually associate with justice, such as courts, laws, the judiciary, and so on. If Hubin is referring to such a particular set of empirical institutions, his argument is right. These institutions might be created and sustained by, for example, a conjunction of power and caprice in an environment where the circumstances of moral cooperation do not obtain. However, if the 'institutions of justice' that Hubin is referring to are institutions whose people under them recognize and accept as moral institutions, then the matter is more complicated. We will return to this case later on, and suggest a sixth interpretation based on this possibility. In any case, Hubin's conclusion seems right in the first case, that of institutions of justice as a particular set of empirical institutions. These institutions can be created from conditions where the circumstances of justice are not present if we introduce, for example, benevolent tendencies into very powerful individuals. Such individuals would use their power to enforce the creation of institutions of justice and also to sustain them through time. Also, due to their power, these institutions would undoubtedly be effective to maintain allegiance to themselves. In both the genetic and the sustaining models, the ontological interpretation of the circumstances of justice does not manage to limit the scope of justice.

The fourth and the fifth interpretations, the deontic and utility interpretations respectively, are considered together by Hubin. The deontic interpretation states that rules of justice are morally obligatory only if the circumstances of justice are in place. The utility interpretation states that justice is only useful in the circumstances of justice. However, because Hume connects both elements, Hubin considers them together. For Hume, duties of justice are founded on their public utility.207 In this sense, if justice is not socially useful, its demands are not morally obligatory. Hubin argues that justice can be useful even if the circumstances do not obtain, and consequently that it can create obligations outside of the circumstances. To do so, he refers to two cases. The first is a

207 See Hume, T 3.2.6.8, SBN, pp. 530-1, M 2.17, SBN, p. 180, and M 3.1, SBN, p. 183.
case of extreme scarcity. He argues that in extreme scarcity, there are rules of distributive justice that can enable all – or a majority – of individuals to have the minimum amount of goods they need to have a decent life. However, some rules increase the number of individuals who can have the minimum, and these rules would count as rules of justice, even though they are not useful for every individual in the system. The other case Hubin examines is again the case of the lifeboat. In this case, the plan to draw lots and determine who should stay on the boat and who should jump in the water to be killed by sharks is a just plan if it is accepted voluntarily for everyone in the boat. Additionally, Hubin argues that if respect for rules of justice is productive of human-well being under the condition of equality of powers, then the importance of such respect should be even greater when this condition does not obtain, i.e., and some individuals have ascendancy over others. For when there is no equality or balance of powers in the social system, some persons, the weak, are much more vulnerable to other persons than otherwise. Consequently, their well-being is in danger. For this reason, Hubin thinks that in such situation the respect for rules of justice has considerable importance. With these cases, Hubin believes to have shown that justice does not necessarily require that the circumstances about which Hume writes.\(^{208}\)

We agree with Hubin's conclusion. In the first case, the one about scarcity, it is shown that rules of justice might be useful for some individuals outside of the circumstances of justice. The few individuals who would be favored by the establishment of a moral agreement would undoubtedly regard the agreement as useful. However, since such an agreement would be disadvantageous for the majority of individuals, the agreement cannot be established. Similarly, when there is no social balance of power in the social system, and thus an individual or group has ascendancy over the others, a moral agreement is undoubtedly advantageous and profitable for the weak individuals in the system. For them, the moral agreement means protection against the power of the strong. For the strong, however, the moral agreement is an unjustified restriction on freedom, since it has no benefits or advantages for them.\(^{209}\)

With the lifeboat case, Hubin attempts to show that justice may emerge where there is extreme scarcity or widespread poverty. Hubin attempts to show that under extreme deprivation of goods individuals could still agree on a moral agreement. The object of the plan is to distribute a good that we can call safety from starvation. By drawing lots, it is decided which individuals remain in the boat, the ones that receive the good, and which individual will not receive the good. Notice that the distribution is being carried out in front of an extreme situation, if individuals do not do it, everyone will perish. The plan can be justified insofar as everyone in the boat does not want to

\(^{208}\) Hubin (1979), pp. 16-19.
\(^{209}\) Unless we assume the strong to possess altruistic preferences.
perish, and the only manner of avoiding this is to risking one's own life either in a lottery or in a
fight for the goods available. The lottery is a better alternative rather than the fight only under
certain strong assumptions. The first assumption is that every individual in the boat is equally
powerful so that no individual would have a higher chance of surviving through physical conflict
rather than through the lottery. The second assumption is that individuals, regardless of how
powerful they are, cannot form coalitions with one another to gain advantages in power and thus
increase their chances of survival by means of conflict. Given these unrealistic assumptions,
participating in the lottery is rational for everyone because it spares individuals from the effort and
suffering of the conflict. Additionally, there is one factor that Hubin does no take into account.
Individuals who take part in the lottery know that the loser will not comply with the result. He will
start a fight to acquire the good he needs. In this sense, the outcome of the lottery is, in any case, a
fight, since it is irrational for the loser to submit oneself to the lottery's result when he can start a
fight. If individuals expect this to happen, it is unclear whether the lottery has any normative value
at all for them.

We return now to an aspect of the ontological interpretation whose consideration we
postponed before. That is the case of how important are the circumstances of justice for the
development and maintenance of institutions of justice, which are considered here to be moral
institutions by those that live under them. According to this interpretation, what makes a given
institution an institution of justice is the fact that individuals accept to have their freedom restricted
by such an institution, i.e., the restriction of freedom performed by the institution is not merely a
sheer manifestation of power. In this view, then, the question at hand is whether individuals can
agree to have their freedom restricted by an institution outside of the circumstances of justice. Such
a question indicates that what defines an institution of justice is that its normative demands are
justified for those who are constrained by them. And although connected with the ontological view,
this question opens the way to a sixth interpretation of the circumstances of justice, the
justificational interpretation.

The justificational interpretation states that the circumstances of moral cooperation or the
circumstances of justice are the circumstances in which moral institutions can be justified for the
whole social spectrum. It is only when a rule or institution is capable of justification that we may
refer to it as a moral rule or moral institution. Otherwise, we are speaking of a rule or institution
whose normativity is based solely on power. Notice that the justificational interpretation of the
circumstances must necessarily restrict the scope of justice. For a moral institution can only be
considered moral, or just, insofar as the normative requirements it produces can be justified. It
seems to us that to speak about an 'unjustified' moral institution makes little sense. Unless we refer
to Hubin’s ontological interpretation as we analyzed it, when we considered an institution of justice to be an institution that belongs to a particular set of institutions, it is not possible to say that a given institution is at the same time an institution of justice and it is not justifiable. If we take a tribunal to be an institution of justice regardless of the circumstances in which it appears, then we can speak of an institution of justice that is not justified. By pulling justification to the center of the idea of justice or morals, we necessarily limit the scope of justice or morality to the extent that a moral rule or an institution of justice can be justified.

The question that must be answered, then, is whether moral justification can be done outside of the circumstances of moral cooperation. We believe it cannot be done. To justify a moral rule means, from the contractarian perspective, to demonstrate that this rule and the normative requirements it produces are profitable and advantageous for the individual to whom the rule is supposed to apply. We can see, then, that the justificational interpretation is connected with the utility interpretation. It is only insofar as moral rules and moral cooperation are useful for every individual in the system that we can say that they are justified. Justice and morality can still be seen as useful even when the circumstances fail, but then it can be useful only to a particular group in the social system, and never for everyone in it. That is the case when, for example, there is no balance of power in the system, and a group of individuals has ascendancy over the others. In this case, the weak individuals find the establishment of moral rules profitable and advantageous. However, it is only in the circumstances of moral cooperation that every individual regards moral cooperation as rational. Hence, we can safely argue that it is only when the circumstances of moral cooperation obtain that moral rules or institutions of justice can be justified. Outside of these circumstances, justification cannot happen because there will always be a group of individuals that will regard moral cooperation as irrational. The justificational interpretation of the circumstances of moral cooperation is the main view of the circumstances that we will adopt throughout our discussion. For this reason, we refer to the circumstances of moral justification, instead of cooperation, from now on.

3.2 Peter Vanderschraff’s Critique and the Difference between Justification and Compliance

Another author who levied critiques against the idea of the circumstances of justice is Peter Vanderschraff. Vanderschraff argues that the conditions that Hume suggests – which he calls the
In order to demonstrate the insufficiency of the conditions suggested by Hume, Vanderschraff refers to Hobbes' description of the state of nature. In the Hobbesian state of nature, Vanderschraff argues, there is a scenario in which the conditions of the standard account are in place, and despite that, justice remains an impossible endeavor. The reason for this impossibility is the fact that moral cooperation would only be possible with the presence of a Leviathan, a power that assures that every individual remains cooperative. In this sense, the Leviathan itself should be one of the conditions that compose the circumstances of moral cooperation, in Vanderschraff's view. Since the standard account does not include this condition, it is incomplete, and hence the conditions of the standard account are insufficient to assure that moral cooperation is possible.

Vanderschraff does not discuss in which sense the circumstances are 'necessary' and 'sufficient' for justice. With Hubin's critique, we saw that the relation between the circumstances and justice could be interpreted in five different manners, plus the sixth additional manner that we suggested. We also saw that in five of those interpretations, the circumstances of justice do not limit the scope of justice, i.e., justice does not require their presence to appear in the social system. It is only in the justificational interpretation that the circumstances of justice show themselves to be necessary for justice, or morality. It is this interpretation that we will have in mind in the discussion here.

When discussing whether the circumstances are 'sufficient' for justice, Vanderschraff writes at one point that the Hobbesian state of nature blocks the possibility of the existence of justice between individuals. On another moment – when he refers to Hobbes' state of nature as a prisoner's dilemma-like situation – he argues that 'justice cannot exist between rational parties matched in a prisoner's dilemma'. However, Vanderschraff's affirmation that the Leviathan itself should be one of the circumstances if the circumstances are to be sufficient for justice is puzzling for us. He fails to notice that the Leviathan is, in Hobbes' view, part of 'justice', i.e., part of what the circumstances of justice enable. If one claims that the circumstances of justice are not sufficient for justice to exist, one is also claiming that a Leviathan cannot come into existence from the state of nature, and Hobbes' argument is the contrary of that.

Hume emphasizes that the circumstances of justice represent the conditions that make justice useful for individuals. In other words, it is when individuals find themselves in the circumstances that a convention of justice can emerge. In this sense, when it is claimed that the circumstances contain sufficient conditions for moral cooperation, that means that the presence of

---

the circumstances suffices to make justice rational, or useful, for every individual. However, moral cooperation on justice can be understood here in two different aspects, and Vanderschraff's argument seems to be directed at one of these aspects only. The first aspect concerns the justification of moral rules, i.e., the justification of the normative requirements such rules produce. On the other side, the second aspect refers to the incentives individuals have to act in conformity with the justified moral rules. In other words, the second aspect is about whether it is rational or not to conform to justified moral rules. Although similar, the questions are different. The first question concerns the matter of justification, whereas the second is about compliance.

Byeong-Uk Yi differentiates these two aspects in a transparent way. He does so in the context of Gauthier's moral theory. Yi separates the question of the rationality of forming a moral disposition – what corresponds to the rationality of participating in the moral agreement – from that of the rationality of acting in conformity with the disposition – or acting in conformity with the moral agreement. Yi argues that Gauthier's theory lacks a 'connecting principle', i.e., a principle that states that if it is rational to create a cooperative disposition – to be a constrained-maximizer – then it is also rational to act in conformity with the disposition.213 Yi is that compliance does not follow automatically from justification, and, consequently, that the question of whether it is rational to participate in a moral agreement – whether a rule can be justified – is a different question from whether it is rational to act in accordance with the rule. Vanderschraff fails to notice that these are two different questions, and the failure leads him to think that the Leviathan must integrate the circumstances of justice.

The difference can be noticed by stressing that it is possible to say that a moral rule can be justified and that compliance with such rule can be irrational in some instances. Justification refers to demonstrating whether accepting the normative requirements of the rule is useful, i.e., rational, for every individual in the system. In the contractarian view, this task amounts to demonstrating that the rule can be the object of a moral agreement. Compliance, in its turn, does not automatically follows from justification. In the first chapter, we considered the case of dilemma involving the two farmers. Although both agree to adopt a rule to constrain them to cooperate, it is a different matter whether they will act in conformity with the rule. For violating the rules can be profitable, if they manage to do it whilst the other party acts in conformity with the rule. It is because there is a risk of individuals refusing to comply that sanctioning mechanisms are introduced in the system.214

213 See Yi (1992), pp. 491-495.
214 In chapter two, we considered Hobbes' view on distributive justice. We saw that he associates distributive justice with the figure of the arbitrator. Our argument for the irrelevance of the arbitrator for justifying a rule of distributive justice is based on the difference between justification and compliance that we consider now. The arbitrator is irrelevant for the justification of rule, since individuals can agree by themselves on the rule. Nonetheless, the arbitrator is not irrelevant to assure compliance with the rule. For although individuals agree with the rules, the

125
Leviathan is a sanctioning mechanism. It is introduced to assure that individuals will conform to the rules they agreed to. However, notice that the Leviathan is itself a product of the agreement. It is because individuals regard the rules as justified that they want to create the Leviathan. In other words, the rationale for creating the Leviathan is itself a product of the circumstances of moral justification. The Leviathan finds its justification in its capacity of ensuring that individuals will comply with justified rules. We do not need the Leviathan to justify the moral rules themselves. They and the Leviathan are justified by reason of the benefits they produce for individuals. Vanderschraff fails to notices this. He confuses justification with compliance, and the confusion leads him to claim that the circumstances of justice are not sufficient for moral cooperation, what is false since they are sufficient for moral cooperation and the introduction of the Leviathan. In sum, the standard account of the circumstances of moral justification contains sufficient conditions to make moral cooperation useful, and thus a rational endeavor, for every individual. The Leviathan is not, and cannot be, one of the circumstances of moral justification.

Vanderschraff also formulates his argument against the standard account with the resources of game theory. He refers to the traditional representation of Hobbes' depiction of the state of nature as a prisoner's dilemma and claims that justice cannot exist between rational agents trapped in a prisoner's dilemma. The game theory view also does not help Vanderschraff's argument. For Vanderschraff does not notice the limitation of Hobbes' approach when it is translated in the language of game theory.215

The first problem with this approach is that it wrongly emphasizes a connection between the circumstances and the fact that uncooperativeness is a dominant strategy for every player in a prisoner's dilemma. What the circumstances represent is the fact that the players are in a condition in which they would be better off cooperating rather than mutually defecting, i.e., that cooperation is Pareto superior to mutual defection. The circumstances are the conditions that make moral cooperation useful for every individual. They are sufficient for moral cooperation in the sense that cooperation is Pareto superior to mutual defection. One may question, of course, how are the players supposed to assure the cooperative outcome without being double-crossed by their opponents. That both would be better off cooperating – by establishing moral cooperation – rather than struggling, it is clear for both players. However, how they can do that, that is a different question, the question that Hobbes' Leviathan attempts to solve. But notice, again, that the establishment of the Leviathan is only rational for both agents in a prisoner's dilemma because the cooperative outcome is Pareto optimal when compared with the uncooperative outcome. The violation of the rule may be profitable for them in particular conditions. The arbitrator is an important figure to ensure compliance when such conditions that may lead to violation of the rule appear.

relation of Pareto superiority between these two outcomes is a representation of the circumstances that players find themselves in, the circumstances of moral cooperation.

The second problem refers to the limitation of Hobbes' view when imported to game theory. Russel Hardin discusses these limitations.216 Hardin compares Hobbes' and Hume's approaches as game theoretic approaches to morality, and he argues that Hobbes' approach was limited by the fact that Hobbes has a static outlook of moral cooperation, whereas Hume introduces a dynamic outlook. The difference can be depicted as the fact that Hobbes has in mind a one-shot prisoner's dilemma, while Hume has in mind an iterated version of the dilemma. As we have already considered in chapter one, in a one-shot prisoner's dilemma, to always defect from cooperation is always the rational option for both players. However, in an iterated version of the same dilemma, it is not so clear that defection is the best strategy and that, as a consequence, both individuals remain in the state of nature.217 In chapter one, we considered the reasons for this difference. The prospects for cooperation that an iterated form of the prisoner's dilemma offer indicate that the circumstances of moral justification would be sufficient for moral cooperation in both the senses of justification and compliance. The reason for that is that individuals may be lead to cooperate even without the presence of a Leviathan. The perception that cooperation in the present is intrinsically connected with the prospects of cooperation in the future can lead individuals to cooperate to avoid mutual defection. Such prospects indicated that Vanderschraff's argument against the sufficiency of the circumstances for moral cooperation is flawed.

In sum, Vanderschraff's arguments against the standard account of the circumstances of justice cooperation fail. The claim that these circumstances do not suffice for the appearance of moral cooperation between rational agents is false. Moral cooperation can be seen from two perspectives, the perspective of justification and that of compliance. In the first perspective, we argued that the circumstances are sufficient for moral cooperation and that Hobbes' description of the state of nature does nothing to disprove that. As for the second perspective, we saw that the circumstances might not be sufficient to assure cooperative behavior, what indicates that another factor, such as Hobbes' Leviathan, may be necessary to ensure the cooperative behavior of rational agents. Nonetheless, the circumstances refer to the conditions for the justification of moral rules, and not of compliance. For this reason, even though they are insufficient to assure compliance from rational agents, it would be wrong to consider the Leviathan a circumstance of moral cooperation.

Vanderschraff advances a second batch of arguments against the standard account. In this

217 See Hume T 3.2.5.8, SBN, pp. 519-520. Hume himself considers the problem of moral cooperation in a circumstance of an iterated prisoner's dilemma in his case of the so called Farmer's Dilemma.
second batch, he targets the view that the circumstances of justice are necessary for justice. In the context of moderate scarcity, Vanderschraff argues that Hume's view that justice serves no purpose in conditions of severe scarcity applies only in the context of rules of private property, but not for justice more generally conceived. Regarding moderate selfishness, Vanderschraff argues that the parties need justice regardless of whether they are selfish or altruistic, in his view what matters is whether there is a conflict of interests between the parties, and such a conflict may arise in cases where the parties are extremely selfish or extremely altruistic. Lastly, regarding the condition of rough equality of powers, Vanderschraff argues that it is not necessary for justice because what justice requires is not that every party has equal powers, but that every subgroup of the individuals can, when united, have some hindering effect on any individual in another subgroup. In this sense, individuals' must not have equal powers, since they can unite with other individuals.

Vanderschraff's argument against moderate scarcity being necessary for justice may seem right at first. However, we believe it is wrong. Formally, scarcity is only relevant for justice when justice is restricted to private property. If we imagine other duties such as a duty of not committing murder, of not torturing, not injuring, and so on, these are duties that have nothing to do with external goods. One could conclude, then, that it is only for duties related to private property that scarcity is relevant. That is false, however. The external goods that the condition of scarcity refers to are goods that individuals need in order to live and survive. Assured possession of these goods is vital because without them individuals have higher prospects of suffering and even death. In the absence of a rule that assures the stable possession of goods, individuals find themselves in a scenario of fierce competition for these goods. In these circumstances, force and fraud are, as Hobbes says, cardinal virtues. In this sense, murder and torture, for example, are means that individuals can employ in the competition for goods. To renounce the use of such means is an irrational action. By renouncing these means one reduces the means available to oneself to succeed in the competition for goods, and thus one augments one's chances of lacking the goods one needs in the future. In this sense, as long as individuals are engaged in the fierce competition for goods, moral cooperation on matters other than private property is not possible. A consequence, then, is that moderate scarcity is a necessary condition for justice or moral cooperation as a whole.

Regarding the condition of moderate selfishness, Vanderschraff has a point in emphasizing that what is an essential condition for justice is the presence of conflicts of interests, and not how selfish or unselfish are individuals. We agree with Vanderschraff in this matter. However, we prefer to formulate this condition as we did before, i.e., in terms of unsocial sociability. We believe that the two central factors of persons' dispositions regarding the circumstances are the advantages they get

from cooperation and the difficulty they have in cooperating. Conflict of interests is an essential element that makes cooperation advantageous. However, it is not every type of conflict of interests that make cooperation rational for those involved in the conflict. Recall the case of the human and the vampire, for example. They experience a conflict that cannot be resolved by cooperation, since cooperation between them is necessarily prejudicial either for the human or the vampire, depending on the terms on which they cooperate.

As for the condition of equality of powers, we are entirely in accordance with Vanderschraff's critique. As it was argued before, this condition captures an important intuition regarding the circumstances of moral cooperation, namely that power is a vital aspect for this matter. However, the formulation that an equal distribution of power is a necessary condition for moral cooperation is false. Vanderschraff calls attention to the fact that alliances and coalitions between individuals with different levels of power may have. In this sense, the power condition is better expressed as the necessary presence of a balance of power between every individual in the system.

4 CONCLUSION OF THE CHAPTER

We conclude now the exposition of our interpretation of the Hume's idea of the circumstances of justice. We interpret this idea in the form of the circumstances of moral justification. They are the circumstances in which we can justify moral rules to rational individuals. There are three central circumstances: the unsocial sociability of individuals, the social balance of power, and the absence of widespread poverty. Individuals are characterized by tendencies that, at the same time, draw them to social cooperation and isolation. They can reap considerable benefits from social cooperation, but they face difficulties to cooperate by reason of antisocial tendencies they have. Cooperation can only ensue between them if there is a balance of power in the social scheme. It is only when there is not any individual with ascendancy over the others that moral cooperation is rational for everyone. Lastly, it is necessary that there are enough goods available for the vast majority of individuals to fulfill their basic needs. When such a situation does not obtain, rational cooperation is irrational for all the poor in the system. The reason for that is that in such condition of scarcity, cooperation means, for the poor, renouncing the sole means they have of escaping poverty. Moral rules can only be justified when these three circumstances obtain. In the contractarian perspective, a moral agreement is only possible when individuals find themselves in the circumstances of moral justification. If a single of these circumstances fails, the consequence is
that some individuals will refuse to take part in the moral agreement and, as a consequence, moral cooperation is impossible.

We move now to consider a subject that has been mentioned in this chapter, compliance. In the next chapter, we examine the matter of stability of cooperation, which is, in essence, the matter of compliance. In the discussion, we recall the idea we developed in this chapter and connect it with the matter of stability. The connection indicates a problem for JNL that we examine in the fifth chapter.
CHAPTER FOUR: THE RATIONAL CONCEPTION OF STABILITY

OVERVIEW OF THE CHAPTER

In this chapter, we introduce and discuss the topic of stability. The objective is to present the idea of stability in cooperation and to contextualize it in the contractarian perspective. The chapter is divided into three parts. The first part of the chapter is dedicated to present the idea of stability. We perform the presentation in concert with John Rawls' and David Gauthier's discussions of stability. We refer first to Rawls' two different conceptions of stability, as well as to his ideas of overlapping consensus and *modus vivendi*. From Rawls' views of stability, we move to Gauthier's thoughts on the subject to understand what stability refers to in a contractarian context. Although we agree with Gauthier's views on the concept, we argue that the view is incomplete. We argue that Gauthier's view of stability is focused on a particular rationale to defection from cooperation, but that there is another type of such rationale that does not appear in the discussion. We distinguish between a rationale toward cheating and one toward rebellion. Gauthier's argument contemplates cheaters, whereas rebels do not appear in it. The second section of the chapter is dedicated to examining the effects that the rebel has over the stability of a moral principle. In this section, we argue that the rebel poses a challenge to stability that sanctions alone cannot solve. The source of the rebel's non-compliance lies in an aspect that sanctions cannot affect. The rebel regards the moral order as unjustified, and that is why he acts against it. In the third section, we consider how the rationale for rebellion arises. Here, we resume the topic we explored in the third chapter, the circumstances of moral justification. We argue that rebels appear in the system as a result of alterations of the circumstances.
1 STABILITY

Referring to Hobbes, Vanderschraff argued that the circumstances of justice are insufficient to elicit compliance with moral rules from rational agents. Compliance, he argued in Hobbes' line, requires an enforcing agent that threatens the individuals with sanctions in case they do not cooperate. Without enforcement, compliance with moral rules is likely to be irrational for rational agents. In Vanderschraff's view, such a conclusion undermines the idea that the circumstances of justice are sufficient for moral cooperation. In our response to Vanderschraff, we observed that although his observation is correct in a sense, it involves a misunderstanding of the idea of moral cooperation. He fails to distinguish the rationality of moral cooperation in the sense of justification of moral rules from that of compliance with moral rules. The argument seems to imply that if compliance with a rule cannot be assured, then the rule cannot be justified. We disagree with such a conclusion. There is no such relation between justification and compliance. If the moral order is justified, that does not mean that compliance with its rules is always rational for every individual. If we refer to the metaphor of the moral agreement, we may affirm that although a rational agent may find it rational to take part in the moral agreement, he might also find it rational to breach the moral agreement at some later moment. Although compliance and justification are connected, as we see ahead, they are slightly different topics. In the context of the metaphor of the agreement, we may refer to Hobbes and Hume to further explain where Vanderschraff's mistake is. Both Hobbes and Hume write about a hypothetical individual whose view toward moral cooperation suits the present discussion. These hypothetical individuals are the Fool and the Sensible Knave.

1.1 THE FOOL, THE SENSIBLE KNAVE, AND COMPLIANCE

The Fool appears in Leviathan.219 He is an individual who accepts the distinction between justice and injustice, i.e., he accepts that individuals should observe the covenants they make with one another. The Fool reluctantly accepts that respecting covenants can be called justice, whereas breaking them may be called injustice. He also accepts that acting as justice requires is rational. However, the Fool argues that the practice of injustice is also rational in certain circumstances. In the Fool's view, the circumstances that warrant the practice of injustice are those in which acting unjustly affords the agent more advantages than doing otherwise. In short, when injustice

maximizes the utility of the agent, he should act unjustly, that is the Fool's opinion.

The Sensible Knave that Hume writes about possesses an attitude to justice similar to that of the Fool.\textsuperscript{220} The Sensible Knave admits that the existence of justice is advantageous for himself, and also that persons should be just to one another in general. However, similarly to the Fool, the Sensible Knave argues that acting unjustly is rational under certain circumstances. The Sensible Knave believes that injustice should be performed when doing so is, on the one side, profitable to the agent and, on the other side, does not threaten the continuity of the whole system of justice.

The Fool and the Sensible Knave are abstractions that help us understand the difference between justification and compliance. Both of them regard participation in a moral agreement as rational. In other words, they both think that moral rules are justified. Their willingness to take part in a moral agreement indicates that morality is, in their view, advantageous for themselves. Moreover, they also maintain that compliance with the moral agreement is rational in general. After all, if compliance could not be deemed rational, they would not argue that participation in the agreement is advantageous for them. Despite that, they maintain that violating the agreement can be rational in certain circumstances. In the case of the Sensible Knave, the commitment with the moral order is more apparent than in the case of the Fool. The Sensible Knave affirms that the performance of injustice can be deemed rational only when it is profitable \textit{and} it does not undermine the whole scheme of moral cooperation. The concern with the maintenance of the moral order indicates a full-blown commitment with it, which we may deduce to be a product of the advantages the order produces for the Sensible Knave. Nonetheless, the commitment with the moral order does not preclude the possibility that immoral actions can be rational, as Vanderschraff seems to argue mistakenly.

Vanderschraff's mistake is to assume that justification implies compliance. Compliance and justification are subjects that overlap with one another. Nonetheless, they are different subjects. Compliance with a rule may be rational when the moral order is unjustified in the same manner that it may be irrational at some moments in the face of a justified moral order. In this chapter, we focus on the matter of compliance with the moral order. This subject is, in essence, the subject of stability.

1.2 \textbf{Rawls' Two Senses of the Idea of Stability}

There is an intrinsic connection between the matter of stability and that of compliance. To have a better grasp of the subject of stability, let us consult what is perhaps one of the best discussions of

\textsuperscript{220} See Hume, M 9.22, SBN, pp. 282-283.

Rawls writes about a 'question of stability'. Such question appears in the second stage of Rawls' two-stages presentation of his political conception of justice as fairness. The first stage concerns the formulation of the content of the conception, the principles of justice and the ideal of justice. This stage is the area where the original position appears and the arguments for his the two principles of justice are advanced. The second stage, in its turn, is about the stability of the conception formulated in the first stage. In Rawls' view, if the conception formulated at the first stage is found to be unstable in the second stage, it must be revised to assure that it will be stable. Stability is, in Rawls' discussion, a property of moral principles or set of moral principles. The principles or sets of principles that have the property of stability are those capable of eliciting compliance from individuals under them. We can verify whether a conception of justice is stable, then, by observing whether individuals are compliant with it or not. If they are, the conception is stable. This view of stability is, however, a broad conceptual formulation. Rawls argues that the idea of stability can be understood in two different senses.

In the first sense, the stability of a conception of justice is determined by its capacity of eliciting compliance from individuals. Under such perspective, the question of stability is twofold. It is a question of working out a conception of justice that is sound and reasonable to oneself and a question of finding ways of bringing others who reject the conception either to regard it as reasonable or to follow it. In the context of guaranteeing the compliance of others, the introduction of sanctions appears as a useful tool. Rawls argues that in this first sense, stability is a 'purely practical matter'. In this first sense, it is exclusively about compliance. For this reason, Rawls believes that if the conception of justice is not stable, in such first sense, then it is futile attempting to bring it about.

The second sense of stability that Rawls mentions is the one that he endorses. Stability in the second sense also refers to compliance. However, in its second sense, the stability of a conception of justice involves a new factor that is not present in the first sense. Rawls argues that stability in the second sense takes into account the nature of the forces that make the conception stable. In the second sense, it matters why individuals are compliant with the conception of justice, and not only that they are compliant with it. With regards to the nature of the forces that lead individuals to comply, Rawls stresses that two factors are essential. The first factor refers to the capacity of the conception of insufllatting a sense of justice in individuals. It matters for the stability of a

223 See Klosko (1994) for deeper discussion of the concept of stability in Rawls' work as a whole.
conception, Rawls believes, that persons who grow under just institutions – which are informed by the conception – develop a specific disposition, which he calls a sense of justice. The sense of justice is a disposition to comply with just institutions and to avoid injustices. The second factor involved in the stability of a conception, in the second sense, is the capacity of the conception of being the object of an overlapping consensus of reasonable comprehensive doctrines. An overlapping consensus is the affirmation of a single conception of justice – a political conception of justice – from the viewpoint of different comprehensive doctrines. When individuals that uphold different comprehensive doctrines reach a consensus about which conception of justice should determine their mutual relations with one another, then an overlapping consensus is established. Such a consensus over a conception of justice indicates that the conception can be justified from the perspective of different comprehensive doctrines.

The problem of stability is different depending on the sense of the idea of stability that we have in mind. In the first sense, the problem of stability is that of introducing appropriate sanctions to ensure compliance from everyone, including persons who reject the conception of justice. In the second sense, the problem is more complicated. Compliance is still what is being sought. However, what one seeks is compliance driven by specific factors, a sense of justice and the view that the conception is justified in light of the reasonable comprehensive doctrine to which one adheres. In Rawls' view, the idea of stability in the second sense involves assuring the support of every individual by demonstrating that reasons to endorse the conception of justice emerge from their own framework of beliefs.

Rawls develops the idea of stability in the second sense in the context of his discussion of justice in liberal democracies. The two factors that he inserts in the idea of stability have their source in the liberal political view. In this sense, they are requirements that citizens from liberal democracies already in possession of a set of certain moral convictions would require of any conception of justice governing their society. The liberal political view that citizens are supposed to share includes the convictions that persons are reasonable and rational, and also free and equal. The moral background present in the idea of stability in its second sense is evident when we consider the contraposition that Rawls makes between an overlapping consensus and a modus vivendi.

A modus vivendi is a pact between two or more parties grounded on the parties' interests. The pact is reached to solve or terminate a conflict between the parties. The terms and contents of the pact are such that violating the pact is profitable for none of the parties involved in it. The reason that leads the parties to the pact is self-interest, such as avoiding a mutually damaging conflict. So the reason the parties have to maintain the pact is that breaching it is not advantageous.

for them. However, as soon as the circumstances underlying the pact change and a violation of the pact becomes profitable, the parties lose the reason they have to uphold the pact, and thus they are driven to become non-compliant with the normative demands of the *modus vivendi*. A *modus vivendi*, then, lasts insofar as it is profitable for everyone involved. The stability of a *modus vivendi* is entirely dependent on a convergence of interests of everyone involved. A conception based on an overlapping consensus, on the other hand, is more stable, in Rawls' view. The reason for that is twofold. The conception of justice that is the object of the consensus is, in Rawls' discussion, a moral conception already. It is imbued with the set of moral convictions that citizens of liberal democracies have. Moreover, the conception is endorsed from moral grounds by the citizens. In an overlapping consensus, citizens endorse the conception from their own comprehensive doctrines' perspective. Such perspectives are already moral perspectives since comprehensive doctrines also involve moral beliefs. Because of these two factors, the endorsement from liberal democratic convictions and that from one's comprehensive doctrine, a conception endorsed by an overlapping consensus will be more stable than a *modus vivendi*. The allegiance to the rules that individuals have in such circumstances is based on moral grounds, and not on interest alone, as it is the case with a *modus vivendi*. For this reason, a shift in the circumstances that underlie the consensus on the conception will not affect the individuals' allegiance to it. Since they endorse the conception from a moral perspective, benefits and profits involved in respecting or violating the conception do not play a major role in defining the individuals' attitude toward the conception. The overlapping consensus, then, is a form of justification that uses moral factors in its justificational procedure.

Rawls' distinction between the two senses of the idea of stability is influenced by the nature of his discussion, namely the search for principles of justice for liberal democracies. Stability in the sense that Rawls endorses is defined through the perspective of citizens of such political systems. Although informed by a particular set of moral convictions, the concept of stability that Rawls endorses can be generalized beyond the set of liberal convictions. The second sense of stability suggested by Rawls involves a conception of stability that we may call *moral stability*. A conception of justice or moral principle can be said to be stable following the conception of moral stability if it can be endorsed by a prior moral principle or moral conviction. In Rawls' discussion, such prior moral factors are determined by the liberal democratic set of convictions. However, one must not restrict oneself to Rawls' set of convictions. We can conceive, for example, a view of moral stability that draws on Christian moral beliefs. For such a view, the stability of a principle or conception of justice is evaluated by whether the principle or conception can be endorsed from a Christian moral perspective. What stands at the core of the notion of moral stability is that the stability of principles

---

226 See Rawls (1993), pp. 147-149.
depends on moral convictions. Under such a conception, we expect individuals to comply with the principles because they feel morally committed to doing so. Think, for example, in a moral view of stability founded on nationalism. In this case, we evaluate the stability of principles of justice by their capacity of eliciting compliance from individuals of a specific nation. Individuals comply with the principle insofar as they believe that the principle promotes the interest of their nation. Their compliance is a by-product of their anterior commitment with their nations. They believe that the good of their nation is something morally valuable, i.e., something that they feel obliged to promote. If a principle of justice helps their nation somehow, their previous commitment with the good of the nation drives them to commit themselves with the principle.227

In the first sense discussed by Rawls, stability is already considerably general. Due to the minimalistic requirements that the notion sets to the stability of principles, we can name stability in the first sense rational stability. In this sense, a principle is stable if it is capable of eliciting compliance of persons. In more broad terms, a principle is stable in the sense of rational stability if persons regard compliance with it as rational. The factors involved in compliance are not important in the context of rational stability. Whether a person complies out of altruism, fear of punishment, an expectation of a benefit, or even a moral conviction in the principle is not relevant for rational stability. There are no other dimensions involved with it but the dimension of practical rationality. We can say, then, that a principle is stable in light of rational stability whenever such a principle can be the object of a modus vivendi. If the principle can sustain a modus vivendi, then it follows that those under the normative authority of the principle find it rational to comply with it.

Compliance is a fundamental factor in both the rational and the moral conceptions of stability. They differ concerning moral stability's demand to principles to conform themselves to certain moral convictions that persons have. It matters for moral stability the reason in particular that leads individuals to comply. If persons comply with a principle exclusively out of fear of punishment, for example, such a principle cannot be considered stable. However, requiring compliance with principles for specific reasons in order to consider them stable is a view that we, as contractarians cannot endorse. The contractarian view that informs our discussion does not allow us to introduce a view of stability that is concerned with the nature of the forces that guarantee the

227 If you recall our discussion about the split between contractualism and contractarianism performed in the first chapter, we can see that Rawls’ idea of moral stability is a typical contractualist view. As we argued, contractualists bring moral elements into their procedures of justification. They aim at justifying principles not only with reference to the advantages the principle affords but also with reference to the principle's adequacy to with another moral principle or conviction. Rawls' idea of moral stability fits perfectly into this contractualist approach. Take Rawls' view of the overlapping consensus as an example. Regarding an overlapping consensus, one tests the adequacy of a principle with two moral factors. First, the set of convictions that citizens of liberal societies have. Second, the particular moral convictions that individuals have in virtue of their comprehensive doctrines. It is only when the principle can be endorsed from both perspectives that it can be deemed stable.
stability of a moral principle or conception of justice. To argue that a specific principle is stable only insofar as this principle can be endorsed by a particular moral perspective involves a commitment with this prior moral perspective. The contractarian method requires such prior moral instance to be subjected to the justification procedure. In the contractarian view, every moral entity is assumed to draw its normative force from a hypothetical agreement. In this context, the normative force of moral principles is, in the contractarian procedure, always justified from non-moral grounds. If we desire to endorse a principle from a moral perspective, we must first investigate whether such moral perspective is itself justifiable from a non-moral perspective. Rawls' argument does not involve any form of justification of the moral factors involved in his conception of stability. He takes for granted the two moral factors involved in his view of moral stability. On the one side, the set of moral convictions of citizens of liberal democracies are assumed to inform the convictions of individuals in general. On the other side, the reasonable comprehensive doctrines are regarded as given facts, which together compose the fact of reasonable pluralism. Such factors are, from a contractarian perspective, altruistic factors and may not be introduced without previous justification. For this reason, the conception of stability that is akin to contractarianism is that of rational stability.

Beyond the problem of moral stability that we stressed above, we have an additional reason to reject Rawls' view of stability as presented in Political Liberalism. We are skeptical that an overlapping consensus is factually attainable. Rawls conceives the overlapping consensus as a tool to assure that his conception of justice can be endorsed from a diversity of moral views. In a plural society, then, a conception of justice can be deemed stable if it can enjoy the acceptance of conservatives, socialists, environmentalists, Kantians, utilitarians, Christians, Muslims, libertarians, and every other adherent of a comprehensive doctrine in that society. It is important, we must stress, that the acceptance of each of these persons is done from the perspective of their specific doctrine. It is important that the Christian accepts the principle not only because the principle affords him certain benefits, but also because Christianity itself endorses the principle. Such requirement is, we believe, too strong to have any empirical relevance. It is hard to imagine a moral principle or conception of justice that could be justified simultaneously from the perspectives of a libertarian, a utilitarian, and a Christian, for example. Of course that fundamental moral principles, such as the prohibition of murder or theft could be subject of an overlapping consensus. Nonetheless, it is unclear whether any principle with more substantive demands could also be the object of the consensus. In this respect, Rawls' overlapping consensus faces the same difficulty that contractarianism has in justificational endeavors, namely accounting for principles with substantive

228 For criticisms against the attainability of the Rawlsian overlapping consensus, see Perry (1989), section IV and also Gray (1995), p. 90.
normative requirements. In this respect, Rawls' view of stability does not offer, in practice, any meaningful advantage over the conception of rational stability.

We reject, therefore, Rawls' idea of moral stability. The two reasons above ground such a rejection. On the one side, the incompatibility of the idea with the contractarian method. On the other side, with respect to Rawls' view of stability, the apparent incapability of the conception to define any substantive principle or conception of justice as stable. We move, now, to a detailed examination of rational stability.

### 1.3 The Rational Conception of Stability

It is the view of rational stability that informs our discussion here. Such view of stability is relatively straightforward. A principle is stable when compliance with its commands is rational for those under its authority. The nature of the forces that lead rational agents to comply is, in this view, irrelevant. For this reason, the problem of stability under the view of rational stability is a question of ensuring that uncooperative individuals become cooperative. There is but one method to achieve such a goal. One must be able to make compliance with the principle rational for every individual under the principle's normative requirements. If subjected to a calculus of utility, compliance with the principle must afford individuals more utility than non-compliance. In this sense, if a particular individual has a non-compliant attitude toward the principle, the reason for such attitude is but that he has more advantages when he does not respect the principle than he would have otherwise. To change the attitude of a non-compliant individual, we must introduce a factor in the system that changes the whole incentive structure of the individual. The change must be such that, after it is done, it is more advantageous for the individual to comply than to not do it. The central measure at our disposal to perform such changes is the introduction of a sanctioning mechanism in the system, as we considered in chapter one.

In an essay entitled *The Social Contract as Ideology*, Gauthier considers the matter of stability in the sense of rational stability, as we have been calling such a conception so far. He defines stability as a property of specific cooperative outcomes. Gauthier conceives the participation in a moral agreement as participation in the selection of a join-strategy to select how to

---

229 See Gauthier (1977), p. 142. Although Gauthier refers to stability as a property of outcomes of a moral agreement, and not as a property of moral principles, as Rawls does, the underlying idea in Gauthier's view of stability is the same of Rawls'. In both cases, stability refers to the capacity of a moral principle (the moral agreement in Gauthier's case and a principles of justice in Rawls' case) to elicit compliance from those under it. Insofar as stability is a property of outcomes of moral agreements, it seems reasonable to claim that moral agreements that lead to unstable outcomes are themselves unstable moral agreements.
the utility surplus afforded by cooperation is to be divided.\textsuperscript{230} The selection of the strategy amounts to the selection of a specific outcome, which represents the specific division of the surplus that individuals agreed to.\textsuperscript{231} Regarding the selection of the joint strategy, either there is only one possible joint strategy, or there is more than one. If there is but one joint strategy available, that means that there is but one distributional baseline of the surplus that is acceptable for everyone. On the other hand, when there are several possible joint strategies, this multiplicity represents the existence of several baselines over which individuals can agree to cooperate. During the selection, the matter of stability arises. Gauthier argues that an outcome A will be stable if two conditions are fulfilled. The first condition is that there must not be an alternative outcome B in which someone is better off than in outcome A. The second condition is relevant only when the first condition is fulfilled. Then, if there is an individual for whom outcome B is better than A, the second condition states that this person must not be able to realize outcome B by unilaterally defecting from the agreed joint strategy that produced A. Instability is characterized by the conjunction between profitability of defection from moral cooperation and the capability of unilaterally being able to profit from the defection.\textsuperscript{232}

Gauthier uses this concept of stability to conceptualize three problems of interaction that appear at the moment of the selection of a joint strategy. These are the problems of coordination, bargaining, and compliance. The problem of coordination refers to the problem of adequately synchronizing each individuals' strategy so that a joint strategy is selected. The problem of bargaining refers to the selection of one joint strategy among several, in the case where individuals' standings in utility vary considerably from strategy to strategy. Although every joint strategy available leads to a cooperative outcome that is Pareto superior in comparison with the natural outcome, some individuals are considerably better off in some of the cooperative outcomes whereas others are considerably better off in other cooperative outcomes. In such circumstances, individuals must bargain to reach a cooperative outcome that is acceptable for everyone. The third problem is the problem of compliance. This problem is directly connected with the matter of stability.\textsuperscript{233}

\textsuperscript{230} In our analysis in chapter two, we saw that the topic of stability appears connected with Gauthier's argument for the Lockean proviso. Gauthier argues that the Lockean proviso assures the stability of the moral agreement because it ensures that the agreement is fair. The fairness of the agreement guarantees is stability, that is Gauthier's argument in \textit{Morals by Agreement}. The view that we expound here is different. In the earlier paper, \textit{The Social Contract as Ideology}, Gauthier discussion of stability does not involve any type of Lockean proviso nor any idea of fairness.

\textsuperscript{231} A joint strategy is basically a collective strategy where individuals decide collectively which strategy every individual will adopt privately. By coordinating their individual choices of strategy, individuals can then produce one of the particular outcomes may result from the combination of their strategies. In this sense, a game with two players, where every player has two possible strategies S1 and S2, is a game in which there are four possible joint strategies. The joint strategies possible are both players play S1, both players play S2, one of them plays S1 and the other S2, or vice verse. By selecting a joint strategy, players can accurately select which outcome they will reach.

\textsuperscript{232} See Gauthier (1977), p. 142.

\textsuperscript{233} See Gauthier (1977), pp. 142-144.
The situation involving the problem of compliance has the following characteristics. The problem of compliance appears when individuals have to select between joint strategies whose cooperative outcomes have three properties. They are all optimal.\(^{234}\) They are all Pareto superior to the natural outcome.\(^{235}\) Moreover, none of them is stable in the sense described above. They are such that whichever cooperative outcome is selected, there will always be at least one individual who is capable of shifting the whole system to another outcome through unilateral defection. Consider the example below to see how instability operates in the manner that Gauthier suggests.

Consider that two individuals are bargaining to divide a surplus of two utility units utilizing a joint strategy. There are then three possible cooperative outcomes that they can reach. Either one of them gets the two units, or the other gets the same two units, or they share the units equally among themselves. Now suppose that the individuals agree on the egalitarian cooperative outcome, and then introduce a joint strategy to produce the equal division of the utility units. If this outcome is unstable, that means that one of the individuals can shift the current distribution of utility to that of the outcome in which he has two units and, the other individual has zero by merely defecting from the joint strategy and adopting another strategy. Under these circumstances, it seems reasonable to think that the individual who can profit from defection will be heavily tempted to defect from the agreed joint strategy to improve his situation. The problem of compliance arises from such opportunities to profit from defection.

The solution to the problem of compliance, Gauthier argues, is the introduction of a constraining device.\(^{236}\) The device turns defection from cooperation unprofitable, and thus eliminates the incentive individuals have to be uncooperative. The connection between sanctions and stability is a traditional element of the Hobbesian tradition, as we argued in chapter one. Hobbes, Buchanan, Gauthier, and Stemmer, all emphasize in some measure the role of sanctions in moral cooperation. A sanction is a negative consequence attached to the performance of a specific action. Observe that a sanction represents a form of social power in the sense suggested by Downing, which we discussed in chapter three. Its core idea is to deliberately alter the incentive structure of persons in order to produce a particular outcome. The outcome may be hindering the person from performing a particular action, or forcing her to perform the action, depending on which is the moral rule in question. In both cases, one has to threaten with sanctions the occurrence of the opposite outcome that one desires.

In sum, in the rational view of stability, the introduction of a sanctioning mechanism is a

\(^{234}\) The optimality of these outcomes refer to the fact that each of these outcomes is such that the alternative outcomes are better for some person and at the same time worse for another person.

\(^{235}\) Pareto superiority regarding the natural outcome means that moving from the natural outcome to a cooperative makes no individual worse off.

\(^{236}\) See Gauthier (1977), pp. 143-144.
fundamental aspect of stability. Recall that in the rational stability view, it does not matter the reason driving the individual to comply. In this respect, compliance driven by fear of sanctions and compliance driven by an identification with the principles are equally valid reasons for compliance. In both cases, the principle in question can be deemed stable, since individuals comply with it. Nonetheless, the sanctioning mechanism can be deemed a more robust source of stability of principles because it appeals to reasons to which every individual is sensible. Avoiding punishment can be reasonably considered a fundamental general interest. For this reason, a sanctioning mechanism is always able to interfere with individuals' incentive structures.

Having explained rational stability and also considered the vital role that a sanctioning mechanism plays in it, we want to address another aspect of this view of stability. The introduction of a sanctioning mechanism is, as it was argued above, a fundamental aspect of the stability of a moral principle. It seems to us, however, that the contractarian tradition, with its strong emphasis on the role of sanctions, overlooks a crucial factor involved with the question of stability. To see what such factor is, let us return to Gauthier's analysis of the concept of stability.

In Gauthier's view, a moral principle is unstable if there is at least one individual in the social system for whom a violation of the principle is at the same time profitable and performable. An essential aspect of this view, one that can be easily overlooked, is that the rationale for defection still contains an acknowledgment of the importance of the moral principle. The alternative outcome that lures the individual to defect from cooperation is a cooperative outcome in the same manner that the current outcome in which the defector finds himself is. The alternative outcomes that can be produced through defection are all, in Gauthier's analysis, Pareto superior to the natural outcome. The defection from the joint strategy shifts individuals involved in the joint strategy to another outcome, which is still Pareto superior to the state of nature. In this vein, defection does not cause the breakdown of cooperation. It is the case, then, that the problem of compliance, or of stability, is, in Gauthier's analysis, the problem of convincing either the Fool or the Sensible Knave to not defect from cooperation. For both of them, defection from cooperation is rational insofar as it affords the individual more utility than cooperation itself, and insofar as it does not cause the breakdown of cooperation. This rationale for defection can be described as that of a cheater in a game.

The cheater is the individual who violates the rules of the game to acquire an advantage in the game itself. The cheater does not want to quit the game or even change the rules of the game. His general goal is still to play the game. If the cheater knew that cheating would drive other players to stop playing with him, he would undoubtedly refrain from cheating. For the cheater, cheating is profitable only insofar as the other players remain playing. That is a requirement for cheating to be rational. We will call from now on the form of uncooperativeness of the Fool and the
Sensible Knave as *cheating* and the individual who defects from cooperation under this rationale as the *cheater*.

The introduction of sanctions is an appropriate mechanism to eliminate the threat to stability posed by the cheater for two reasons. The first reason is that the sanction is sufficient to alter the benefits and losses connected with defection. It makes defection costlier, and consequently, cooperation becomes more profitable. The second reason is that the cheater himself accepts the principle guaranteed by the sanction. He does not deny that the moral order is advantageous to himself. Since he accepts the principle, it is reasonable to assume that the cheater also accepts the introduction of sanctions to ensure the stability of the principle. Of course that it does not follow necessarily from the justification of the principle that the sanction associated with the principle is also justified. The acceptance of a principle does not entail the acceptance of *any* sanction connected with the principle. An individual can endorse a moral principle that commands the performance of charity but reject the introduction of a sanction that threatens to confiscate goods from those that fail to act following the principle. His rejection does not concern the principle or the proposal to introduce sanctions to stabilize the principle. Instead, his rejection refers to the sanction that was chosen in particular. What follows from the acceptance of a principle is that *some* sanctions are acceptable, but not that every sanction is acceptable. Hence, the cheater himself agrees that sanctions must be introduced to assure that individuals will adhere to the principle. Cheating, however, is not the only sort of rationale for defection. We think that there is another sort of rationale that moves individuals toward defection. Moreover, the introduction of sanctions is not an effective means of dealing with individuals acting under this second type of uncooperative rationale. It is the observance of this second drive toward non-compliance that we think that the contractarian tradition has overlooked.

Albeit his tendency to defect from cooperation, the cheater remains committed with the moral order. If cheating resulted in the breakdown of cooperation and the reinstallment of the natural condition, the performance of cheating would be irrational from the outset. In Gauthier's analysis, the alternative outcome that lures the cheater out of the joint strategy is always Pareto superior to the natural outcome for every individual in the system. A second case that may lead an individual to defect from cooperation is a preference for a shift back to the natural outcome rather than a shift to another cooperative outcome. If an individual has such a preference, we can imply that he prefers to be outside the moral order rather than inside it. Regardless of the reason behind this preference, it can have only one source. Its source is the fact that the individual thinks to be better off in a natural outcome rather than in any of the cooperative outcomes. The advantages that the individual reaps in the state of nature outweigh those he gets in the moral state.
The rationale for non-compliance from an individual who wants to return to the state of nature is different from that of the cheater. Whereas such individual is worse off with the moral principle in question rather than without it, the reverse is true for the cheater. As a consequence, whereas the cheater accepts the introduction of sanctioning mechanisms, such individual refuses it. The attitude of this individual toward the moral principle is entirely different from that of the cheater. The individual does not accept the principle nor the sanctions connected with it. If he can, he does not comply with the principle and he resists its sanctions. If the sanctioning mechanism is more powerful than he is, he conforms his behavior with the principle but attempts to undermine the sanctioning mechanism. In a different manner from the cheater, this individual does not cease to have destabilizing effects for the principle once sanctions are introduced. The introduction of sanctions is indeed a factor that reduces the destabilizing effect that the individual has. However, because the individual rejects the principle, and consequently the sanctions behind it, he will not comply when he has a chance to avoid compliance. Moreover, he also works to undermine the sanctioning mechanism that assures the stability of the principle since it forces such an individual to act by a principle that he does not accept. Such individual behaves as a rebel. His attitudes toward the principle and the sanctioning mechanism are of a rebellion.\textsuperscript{237}

What we believe that the contractarian tradition has overlooked is the role of the rebel in the matter of stability. As we argue, the tradition puts a strong emphasis on the matter of sanctions. Such emphasis indicates that the central concern has been to ensure the stability of principles against cheaters, such as Hobbes' Fool and Hume's Sensible Knave. However, a sanctioning mechanism is not entirely efficient in a confrontation with rebels. The relation between the mechanism and rebels is the next subject that we examine. To see the effects of the rebel over the stability of a moral principle, we must abandon an abstract and formal view of the sanctioning mechanism and adopt a more material view. We perform this task in the next section.

2 STABILITY AND REBELLION

2.1 STABILITY AS A PUBLIC GOOD

\textsuperscript{237}In chapter one, we argue that normativity can be either justified or unjustified. Rebellion is a reaction of rational agents against unjustified normativity. The rebel is the individual who resists the authority and the force of a normative entity that he deems unjustified.
To see the effects of the rebel over the stability of a moral principle, we must abandon an abstract and formal view of the sanctioning mechanism and adopt a more material view. In the formal view, sanctions are negative consequences connected with the realization of some state of affairs. Their introduction is made to assure that uncooperative behavior will be much costlier than the cooperative behavior. The view is not concerned, however, with how exactly sanctions are introduced, sustained over time, and how they materially work to assure the stability of moral principles. To address such matters, we must provide a more material account of the mechanism. James Buchanan's discussion of sanctions gives us a hint to how we may shift the perspective from a formal to a more material one.  

Buchanan argues that compliance with the moral principles depends on a calculus made by individuals that take into account the efficiency of the sanctioning mechanism and the harshness of its sanctions. These two factors are, Buchanan argues, determined by the commitment of the whole moral community with the sanctioning mechanism itself. Buchanan's discussion indicates two essential factors that concern the sanctions. The first is that sanctions do not necessarily produce a compliant attitude toward moral principles from rational agents. A sanction must possess a certain degree of efficiency and harshness to engender such an attitude. Efficiency refers to the likelihood that the sanctioning mechanism has of successfully administering the sanction to the defector. Harshness, on the other side, alludes to the severity of the sanction itself, to how harmful is the consequence imposed on the defector. The second factor indicated by Buchanan's discussion is that sanctions are social products. They are built, sustained, and employed by the community.

With Buchanan's view of sanctions, it is possible to see why we must consider the sanctioning mechanism from a more material perspective in order to discuss the effects of rebels over stability. Since individuals comply with the principles based on a calculus of utility, what one must do to ensure that compliance is always rational is to introduce a sanction that always makes defection costlier than cooperation. If the sanction is threatening enough, no individual, irrespective of the circumstances in which he finds himself in, will find it rational to defect from cooperation. Consider the case in which persons have been convinced that they will be sent to Hell for eternity in case they violate any of the moral principles in their society. If they are indeed convinced that there is such a sanction, and if they are rational agents, they will never commit any transgression whatsoever. Defecting means, for these persons, to be subject to the loss of an infinite amount of utility. There is no advantage or benefit that a violation of a moral rule can grant them that will

239 From here on it becomes clear that our contractarian approach is nearer to Hobbes' and Buchanan's than Stemmer's. We treat the sanctioning mechanism here as a social construction that is sustained by a collective investment of the community In this sense, the mechanism we have in mind here is nearer the state than the system of informal sanctions that Stemmer writes about.
outweigh such cost. A rebel in this system will always comply with the moral rules, and even if he is irrational enough to attempt to destabilize the system, he has no means of ever doing so. The rebel cannot affect the cosmological structure in which he is inserted in. Once the possibility of a sanction that is that threatening is allowed, the distinction from a rationale toward rebellion and one toward cheating is not relevant. Such sanctions, however, are but hypothetical. In pragmatic terms, there is no way to introduce such a powerful sanction in the system. The reason for that is that sanctions are social constructs, as Buchanan indicates. For this reason, they are inherently subjected to a limitation with regards to how threatening they can be. In practice, as we argue ahead, a sanctioning mechanism is inherently inefficient and cannot eliminate the rationale for defection entirely. It is in because defection always remains a rational option in certain circumstances that the rationale for rebellion must be stressed, since it offers a risk for the stability of the moral order different from that offered by the cheater.

The efficiency and harshness associated with sanctions are not hard to comprehend. The higher the harshness and the higher the efficiency of the sanctions, then the higher the incentive to not defect from cooperation. Conversely, inefficient and mild sanctions are not a strong incentive to abstain from defection. With regards to the incentive they give to cooperation, the harshness and the efficiency of sanctions are intertwined in a particular relation. Let us assume that the sanctioning mechanism is perfectly efficient, i.e., that it never fails to constrain defectors. In this case, in order to assure the stability of the moral principle, the harshness of the sanction must be such that the cost it imposes on defectors is only slightly higher than the benefits he would acquire from the defection. There is no need to increase the harshness in such cases since doing so causes no alteration in the stability of the principle. Now consider that the constraining mechanism is not perfectly efficient. It sometimes fails to administer the sanction to some defectors. In such cases, some individuals will be tempted to defect because the defector has a chance of avoiding the sanction. The sanction must become harsher in comparison to how it was under the perfectly efficient mechanism in order to eliminate chances of defection. Otherwise, the principle loses the stability it had. The increased harshness will then make the risk of defecting and being caught by the constraining mechanism irrational for potential defectors. We can see that there is, then, a relation between efficiency and harshness concerning the stability of principles. The lower the efficiency of the constraining mechanisms, then the harsher must the sanction be. Conversely, the harsher the sanction is, then the less efficient must the mechanism be. The stability of a principle, then, is a product of the efficiency of the constraining mechanism behind the principle and the harshness of the sanctions associated with it.

The combination of the efficiency of the sanctioning mechanism and the harshness of the
sanctions it administers indicates the amount of power, in the sense suggested by Robert Dahl that we examined in chapter three, that the mechanism possesses. It indicates how capable the mechanism is of ensuring that recalcitrant individuals comply with the moral principle. Now, the stability of a principle is entirely dependent on whether individuals find it rational to comply with it or not. The introduction of sanctions in the system is done precisely with the purpose ensuring that individuals find it rational to act following the principle. We may claim, then, that the stability of the moral principle is dependent on the power of the mechanism since it is the mechanism that supplies each individual with the rationale to comply. The rationale for compliance, it is interesting to note, can only be said to appear if the sanctioning mechanism is indeed able to sanction the individual. Otherwise, in a case in which the mechanism is either weaker or as strong as the individual, the mechanism is of no help to produce compliance. Compliance through sanctions is only assured when the sanctioning mechanism has ascendancy over every individual and coalition in the social system. The mechanism must have considerably high odds of successfully constraining individuals and coalitions into compliance with the principle. If such a condition does not obtain, then there are persons or coalitions in the system that cannot be adequately constrained by the mechanism. As a consequence, such persons or coalitions can defect from cooperation and avoid punishment for that. Such is a circumstance in which the moral principle is unstable. One should expect here moral cooperation to break and the state of nature to be reinstalled.

That a sanctioning mechanism has ascendancy does not signify that the mechanism is always able to sanction every uncooperative individual successfully. Instead, It means that the mechanism always has higher odds of successfully sanctioning non-cooperative individuals. For this reason, although ascendancy is a property that the mechanism acquires once it reaches a certain degree of power, different ascendant mechanisms can have different amounts of power. Ascendancy is acquired when a certain level of inequality in power is reached. After this inequality is established, mechanisms can have different amounts of power, and thus different odds of coercing individuals with success, even though all mechanisms have ascendancy after such inequality level is achieved.

The possibility of significant variations in power between ascendant mechanisms allows us to conceive of a spectrum of stability. When the mechanism acquires ascendancy, the moral principle guaranteed by it becomes stable. However, the higher the power of the mechanism, then the higher the odds it has of successfully constraining individuals are, and thus the higher the stability of the principle it sustains is. We may imagine here a perfectly stable moral principle. In

---

240 It is both Downing's and Dahl's concepts of power that underlie our discussion here, as it is the case in chapter three.
such a case, every action of every individual is always done in conformity with the principle. For this purpose, the sanctioning mechanism must be so powerful that it would never fail to sanction a defector.241 Because individuals know of the overwhelming power of the mechanism, defection from cooperation is always irrational for them, regardless of the circumstances in which they find themselves. On the other side, we may imagine a highly unstable moral principle. In such a case, the constraining mechanism still has ascendancy over the social scheme. However, the power of the mechanism is such that its ascendancy is based on a not so considerable inequality of power between it and individuals in the social scheme. A principle whose stability is assured by such mechanism can still be said to be stable. Despite that, the frailty of the mechanism is such that there is a series of circumstances in which it is rational to defect from cooperation due to the not overwhelming gap of power between individuals or coalitions and the mechanism.

In sum, the first element that allows us to give some materiality to the sanctions is the idea of the amount of power of a sanctioning mechanism. The mechanism must have ascendancy over individuals to force them into compliance. Otherwise, it is likely that individuals will violate the moral order the mechanism is built to preserve. The next element which assists us in shifting the formal view of sanctions into a more material view concerns how the social system determines the amount of power that the mechanism has.

The sanctioning mechanism and the sanctions themselves are, as Buchanan indicates, constructions of the social scheme. In this vein, the power of the mechanism is directly determined by the community itself. The determination must involve at least two aspects, a formal and a material one. The formal aspect refers to the structure and the amount of power that the mechanism is supposed to have.242 In determining such aspect, individuals are concerned with a single purpose. They want to create a powerful mechanism, a mechanism that has ascendancy over them. The formal aspect refers, in short, to how powerful individuals want the mechanism to be.243 The material aspect, on the other hand, is about determining how the formal aspect is to be brought about in the social system. The discussion here concerns the financing and maintenance of the mechanism. It is a discussion about allocating resources available in the social scheme to the mechanism so that it can function properly. To visualize the difference between the two aspects, consider the following case.

All individuals accept a moral rule that prohibits murder. As a consequence, they also accept

241 That is the case of a mechanism that involves the threat of sending persons to Hell eternally, that we considered previously.
242 In the next chapter, we consider a political dimension of the sanctioning mechanism. Such a dimension is a crucial part of the formal aspect of the sanctioning mechanism.
243 In the social contract tradition, this aspect represents the creation of the state, the Leviathan, by the multiplicity of individuals giving up their strength and power on behalf of a sovereign. See Hobbes (1998), p. 114.
the introduction of a sanctioning mechanism to ensure the stability of the rule. A discussion about
the formal aspect of the principle is a discussion about whether the community should introduce a
police force or a surveillance mechanism that alerts the entire community as soon as someone
attempts a murder, for example. It is a discussion about how, by which means, the stability of the
rule is to be assured. As for the material aspect of the mechanism, the discussion revolves around,
for example, deciding the taxation system that will collect the resources to fund the sanctioning
mechanism. The material aspect is about the community mobilizing goods and resources to
materialize the mechanism the community wants to introduce.

The amount of power of the mechanism guarantees the stability of principles. If the
sanctioning mechanism is a product of the social system, then the functioning of the mechanism
depends on the investment of resources and goods by the community. As a consequence, we can
argue that the power of the mechanism is entirely dependent on the goods and resources available in
the system. The maximum level of stability that a moral principle in a given community may enjoy
is determined by the maximum quantity of goods and resources available for the community. In this
context, we may imagine a community in which every resource and good is allocated to assure the
stability of the moral principles that govern it. Such a community is, we may imagine, a community
where every individual is a policeman and all the resources are directed to sustain the policing
activity. A community of police officers is, of course, a counter-factual abstraction. Although
individuals generally desire to be under stable moral principles, it is unreasonable to suppose that
they are willing to invest all their goods and resources into the sanctioning mechanism to assure the
stability of the principles. The more they invest, the less they are left with to pursue their plans and
desires. We may suppose, then, that there is a threshold in each's share of goods and resources that
he is not willing to trespass in his investments in the sanctioning mechanism. Such a threshold is
crossed when the individual's investment in the mechanism is so high that he would be better off
without the mechanism than with it. He may be made worse off by the mechanism if it draws from
him so many resources and goods that he becomes less able to pursue his interests and desires than
he would be in a state of nature. Sustaining the mechanism, hence, becomes irrational if the
threshold is crossed. Conversely, whenever the contribution is below the threshold, individuals are
willing to sustain the mechanism.

The overall power that a constraining mechanism may have, then, is limited by the overall
amount of goods and resources that the community is willing to invest in it. That is the first relevant
factor of the material aspect of the sanctioning mechanism. Sanctions cannot be created freely and
in complete accordance with what individuals desire. Instead, individuals are constrained by the
goods and resources available in their system and their own willingness to spend these goods and
There is a second relevant aspect that we come to consider now. The sanctioning mechanism is a device that is collectively sustained. It is through the goods and resources of the whole community that the power of the mechanism is built and sustained. It is such power that creates the stability of the moral principles for the social system. Since the mechanism creates the stability, we may consider it a good that the community creates utilizing the mechanism. In this respect, stability can be classified as a public good, and as every public good, it is subjected to the free-rider problem.

Two characteristics define public goods. Public goods are nonexcludable and nonrivalrous. To be nonexcludable means that it is either too costly or impossible to exclude individuals from enjoying the good once the good is available. To be nonrivalrous means that when a person uses the good, she does not create a hindrance or difficulty for others to use it. The paradigmatic example of a public good is a lighthouse. A lighthouse in the coast marking a dangerous coastline helps every ship navigating in its proximity. The lighthouse is nonrivalrous. When a ship uses the lighthouse to avoid dangerous points of the coastline, the lighthouse remains the same it was before the ship made use of it. If a ship uses the lighthouse, its light does not become weaker nor does it stop shinning. It shines equally strong for every ship navigating nearby. The lighthouse is also nonexcludable. If we imagine that the lighthouse has an owner, a person who invested goods and resources in building and maintaining the lighthouse, it is hard to think how such owner could offer the service of the lighthouse for some but deny it for others. Whenever the lighthouse is lit, every ship in the vicinities can see it, whether the owner authorizes it to use the lighthouse or not.

The stability provided by the constraining mechanism is a public good in the same manner that a lighthouse is. Stability of moral principles assures that every individual in the community enjoys the benefits afforded by social cooperation. Stability is a nonexcludable good. A member of the moral community cannot be excluded from the enjoyment of the stability of the moral principles that govern the community. When a principle is stable, that is so because the sanctioning mechanism guarantees that every individual will comply with the principle. To prevent someone from enjoying the benefits of stability implies that the mechanism does not prohibit persons from violating the principle to affect the excluded person. In order to exclude someone from the benefits of stability, one must exclude the person from the moral community. Since the mechanism is a device to make moral principles stable, an exclusion of a group or individual can only be done with the reinstallation of a state of nature between the group or individual and all the rest. If the

244 See Buchanan (1975), pp. 47-50.
mechanism were to do it, it would defeat the purpose for which the mechanism itself was introduced, i.e., to guarantee stability. Excluding someone from the moral community would only increase the overall level of instability in the community. Like the lighthouse, stability is also nonrivalrous. Everyone in the system can enjoy stability regardless of the benefits it affords each in particular. The enjoyment that a single individual draws from it does not hinder others from enjoying it.

As a public good, stability suffers from the problem of free-riding. That is a problem of efficiency that, in the case of stability, drives the community to underproduce the good. The problem is the following. The public good is produced through a joint investment of goods and resources from members of the community. Once produced, everyone can equally enjoy the good and cannot be hindered from enjoying it. Because the good is nonexcludable and nonrivalrous, individuals notice that their enjoyment of the good will not be harmed if they either reduce or stop their particular contributions to sustain the production of the good. A reduction or cessation of one individuals' contribution has no considerable effects over the production of the good since a multiplicity of persons produces the public good. If a single person ceases to contribute, the overall amount of resources pouring into the production of the good remains large enough to sustain the production without any considerable change. The individual who either ceases or reduces his contribution but keeps enjoying the public good is the free-rider. To be a free-rider is to enjoy a public good without adequately contributing to the production of the good.

Stability is affected by the problem of free-riding insofar as stability is a public good created by the sanctioning mechanism, and the sanctioning mechanism itself is subjected to this problem. As argued before, the amount of power of the mechanism is determined by the societal investment of resources and goods into the mechanism. Depending on how much investment in the mechanism is made, the principle can be stabler or less stable. For this reason, the investment in the mechanism can be reduced without necessarily making the principle unstable. If the reduction does not threaten the ascendancy of the mechanism over the social scheme, the principle remains stable, although less stable than it was before. Given this possibility, a reduction of investment in the mechanism, from a single individual's perspective, is not outright irrational, since it does not necessarily eliminate the stability of the principle. It is in virtue of such possibility that the free-rider problem connected with the sanctioning mechanism affects the stability of moral principles. Private reduction in investment in the mechanism is not irrational because the reduction does not necessarily make the principle unstable. Consequently, there is an incentive to free-ride in the production of stability. Such incentive lures individuals to reduce their contributions. Such incentive is, in essence, an incentive for the individual to reduce his private investment of goods and resources in the maintenance of the
constraining mechanism, and thus to reduce his contribution in the production of stability for the moral order. With a reduction of investment, it follows that the principle becomes less stable. The free-rider problem connected with stability indicates that the level of stability in the system is likely to decrease as time advances.

We have now, the two relevant factors that materially determine the sanctioning mechanism. They are the limitation of goods and resources available to invest in the mechanism and the free-rider problem associated with such investment. Both factors, and in particular the free-rider problem, indicate that the mechanism will always be somewhat inefficient. It cannot be as powerful as individuals would like it to be. Firstly, because the quantity of goods and resources individuals have to use to erect and sustain the mechanism is not only limited, but the amount they are willing to cede to the mechanism is also limited. Secondly, because stability is a public good, individuals have an incentive to free-ride on its production. Consequently, the investment done in the sanctioning mechanism is supposed to reduce as time advances. Such a reduction leads invariably to a decrease of the mechanism's power. The less power the mechanism has, then the lower is the level of stability it can sustain.

It is important to notice, however, that the inefficiency that marks the sanctioning mechanisms does not, by itself, results in unstable moral principles. The limitations in the resources and goods and the free-rider problem conjoined, may lead to a gradual decrease in the overall level of stability of the moral principle. Such a decrease, however, cannot reach a point where the sanctioning mechanism loses its ascendancy over the social scheme. If that happens, the moral principle becomes indeed unstable. Ascendancy, nonetheless is always preserved because as the power of the mechanism is reduced, and therefore the stability of the principle, the incentive to free-ride decreases and the willingness to invest in the mechanism augments. Rational agents want to be under a stable moral principle. The less stable the principle becomes, the more uncertain the enjoyment of the benefits afforded by the principle becomes, and thus the higher the incentive to invest resources in the mechanism. There is a 'tug-of-war' between the desire to retain as many goods and resources as possible for oneself and the desire to enjoy the benefits afforded by the moral principle. As the stability of the principle decreases, the desire to enjoy the benefits of morality becomes stronger than the desire of retaining resources for oneself. Conversely, the more stable the principle is, the stronger is the desire to free-ride on the maintenance of stability. Such relation, we believe, assures that the principle is not driven into total instability because of free-riding. The possibility of a return to the state of nature, we may think, works as a natural sanction that motivates individuals to maintain the stability of the moral principle.

We have now a more material view of the sanctioning mechanism. We know that the
mechanism is inherently inefficient. Although it manages to make the moral principle stable, the degree of stability of the principle tends to diminish with the flow of time. Such less formal view of the mechanism allows us to discuss the effects that the rebel has over the stability of moral principles more appropriately. That is the subject to which we turn in the next topic.

2.2 The Rebel

Our conclusion in the previous discussion is that the sanctioning mechanism is always inefficient. Such inefficiency affects the stability of the moral principle assured by the mechanism. The inefficiency of the mechanism indicates that rational agents will have chances of avoiding its sanctions in cases of defection. For this reason, there are circumstances in which cheaters and rebels will find it rational to defect. Nonetheless, the damage that each type of defector has over the stability of the principles is different.

The sanctioning mechanism affects the incentives to defect that both the cheater and the rebel have. However, the rebel always prefers not to comply rather than to comply. In a different manner from the cheater, who may defect from cooperation but will refrain from doing so if the defection puts at risk the stability of the moral framework, the rebel will always defect, except if he is forced to comply by sanctions. The inefficiency of the constraining mechanism may offer the cheater some opportunities to defect, that is true. However, the cheater's defections are not a significant threat to the stability of the moral principles because the cheater acknowledges the benefits of moral cooperation and, for this reason, remains committed to the principles and their stability in general.\textsuperscript{245} The rebel, on the other hand, is not committed to the principles nor with their stability. What he desires is to destabilize the principles and ultimately to remove the moral principles from the social system. In virtue of the inefficiency of the sanctioning mechanism, the rebel always has some margin to carry on his destabilizing activities.

We argued in the last topic that the individual's incentive to free-ride on the maintenance of the sanctioning mechanism reduces insofar as the mechanism becomes inefficient. In other words, the less stable the principle becomes, the less intense the incentive to free-ride becomes, and the stronger the incentive to invest in the mechanism becomes. Such incentive structure, however,

\textsuperscript{245} That does not mean that cheating has no effect over the stability of a moral order whatsoever. Cheating affects the stability insofar as individuals will refrain to cooperate with one another out of fear of being cheated. If cheating is widespread and rampant, individuals will likely assume that they are better off not cooperating rather than cooperating. We return to this subject in the last chapter, when we consider the profits that entail from cooperation and how such profits are connected with the stability of the moral principles.
applies to the cheater, but not to the rebel. The idea of 'free-riding' onto the mechanism does not apply properly to the rebel. The free-rider is someone who is benefiting from the use of a public good and who does not contribute to create or sustain. The rebel, in his turn, does not draw any benefit from the sanctioning mechanism nor the stability of the moral principles. The rebel wants the destabilization of the principle and, ultimately he wants the mechanism to lose the ascendancy it has over him. Due to this desire, the effect that a reduction of the stability of the principle has over the rebel's attitude toward investing into the mechanism has the opposite effect that it has over the cheater. The more inefficient the mechanism becomes, and thus the less stable the moral principle becomes, the higher is the rebel's incentive to avoid contributing. For this reason, the rebel offers a threat to the stability of the principle that the cheater does not. As the stability of the moral principle sinks, the incentive of rebels is to push the levels of stability to even lower grades until the principle finally becomes unstable.

The options that rebels have to undermine the mechanism and the stability of the system are an interesting question to be pursued. We do not, however, pursue such question thoroughly here. The reason for that is that such a question is highly empirical. The means the rebel has for undermining the mechanism are highly dependent on the empirical environment that the social scheme finds itself. Nonetheless, we may sketch a broad and formal account of how such undermining of the mechanism may be carried out by the rebel. The rebel must undermine the structure of investment in the sanctioning mechanism to undermine the stability of the moral principle. The rebel must, in short, reduce the amount of goods and resources that pour into the mechanism. As the investment reduces, the power of the mechanism reduces as well, and thus the level of stability of the principle sustained by the mechanism decreases. The overall objective is to reduce the investment to a margin where the mechanism does not have ascendancy over the rebel anymore. The reduction of investment in the system can be made in two manners. Either by reducing the investment that the rebel himself is forced to make or by reducing the investment that other individuals make. With the first manner, the rebel can reduce his investments insofar as he can avoid the sanctions he would receive for doing so. In the second case, the rebel works to reduce the investment in the mechanism that other individuals perform. This objective may be achieved in a myriad of ways. The rebel may, for example, force other individuals to make unproductive transfers of goods to himself, and then refuse to transfer resources to the mechanism. Another method would be to convert other individuals into rebels, i.e., to start a coalition of rebels inside the moral framework. There might be other manners to undermine the mechanism, but notice that in all cases, especially the two we considered, the success of the strategy is heavily dependent on the chances of the rebel to avoid the mechanism. Such strategies are only possible because the sanctioning
mechanism is inherently inefficient, and the rebel always has opportunities to avoid sanctions.

We can conclude, then, that the rebel poses a formidable threat to the stability of the moral order. The rebel aims at making the moral order unstable. The sanctioning mechanism has some capacity of assuring the rebel's compliance, but because it is inherently inefficient, it always leaves some room for the rebel to perform his undermining activities. Moreover, differently from the cheater, a reduction in the amount of power of the mechanism, which we should expect to happen in virtue of the free-rider problem that afflicts the mechanism, always works to reinforces the rebel's non-compliant attitude. The rebel indicates that sanctions are not enough to assure the stability of moral principles. Sanctions cannot adequately assure the stability of a principle because they are not wholly effective against rebels. Although sanctions can make rebels compliant in some level, they are entirely unable to change the attitude of the rebels with regards to the principle it guarantees. Regardless of the amount of power of the sanctioning mechanism, the rebel regards the principle as unacceptable, and for this reason, he always maintains a non-compliant attitude toward the principle. To handle the threat to stability posed by the rebel, one must approach the matter from another perspective than that of sanctions. To see what perspective that is, we must investigate how an individual becomes a rebel. So far we considered only the attitude of the rebel toward the principle and the mechanism, but not how such attitude emerges.

What we said to be the defining trait of the rebel is a preference for the state of nature over the cooperative outcome. Such preference implies a factor. It implies that the holder of the preference is better off in the state of nature rather than in the moral framework. If a move to the state of nature makes the rebel better off, then it must be the case that the moral framework, as it is arranged, either does not afford him any considerable benefit or the benefits it affords are outweighed by the benefits that the state of nature affords the rebel. In any case, the moral framework cannot be deemed profitable nor useful for the rebel. If that is so, we can conclude that the moral framework, as it is arranged, is not justified or justifiable for the rebel. As we argued in chapter one, the contractarian method refers to the usefulness or profitability of moral principles to justify them. It is solely because an individual is better off under a principle rather than without it that the principle is justified to him. We may conclude, then, that the rebel is an individual for whom the moral principle is not justified and that is the reason why he wants to destabilize it and undermine the sanctioning mechanism that underlies the principle.

The advantages with which a principle provides individuals determine its justifiability. In the last chapter, we argued that three essential circumstances make moral principles in general justifiable for rational agents, the circumstances of moral justification. It seems reasonable to suppose, then, that if the rebel's defection has as its reason the unjustifiability of a moral principle,
such reason is in some manner connected with alterations in the circumstances. Notice, however, that the attitude of rebel must necessarily encompass the whole moral framework and not necessarily a particular rule. If the rebel wishes to reinstall the state of nature, then the moral order as a whole makes him worse off. If an individual dislikes a particular norm but still sees the whole framework as advantageous because of the other norms, such an individual cannot be a rebel. The rebel is the individual who is made worse off by the moral framework as a whole.\textsuperscript{246} It may be the case that the rebel objects directly to just a single rule, and not to the others. However, if he is a rebel, that means that the normative requirements of the rejected rule give him more disadvantages than the sum of advantages of the other norms in the framework. Regardless of whether the rebel is displeased with a single rule, a group of rules, or every rule in the moral framework, the displeasure is such that he prefers to reject the whole moral framework. For this reason, what the rebel regards as unjustified is not a single rule or a group of rules, but the \textit{whole} moral order.

The rebel attitude towards the moral order indicates a particular interpretation of the contractarian procedure of moral justification. The contractarian procedure is often formulated as a procedure to justify specific rules and not bundles of rules. When discussing the justification of moral rules, Stemmer, for example, argues that the decisive factor to determine whether a given rule is justified is to examine whether the rule in question collides with a norm that prohibits oppression.\textsuperscript{247} Such norm is, in a first moment, an average norm in the same manner that the prohibition of murder or theft are. Individuals have a fundamental general interest in being free from oppression. From a state of nature scenario, then, the best manner to promote such interest is to sign a moral agreement on a rule that forbids oppression. However, Stemmer's view is that this norm also works in a second instance as a \textit{meta-rule} to evaluate whether particular moral rules are justified. A rule always involves some normative constraints for those under its authority. If the normative authority of the rule is not justified, the constraints it imposes should be perceived as a form of oppression, since the unjustified rule interferes with the freedom of individuals without their endorsement. With the meta-norm, we can evaluate the justifiability of each particular moral rule in the moral order. Stemmer's view of meta-rule, however, is problematic in one aspect. It has a

\textsuperscript{246} Another important factor that subsumes this attitude of the rebel is that the sanctioning mechanism cannot be weakened in relation to its capacity to assure the stability of certain moral rules but not of others. The power of the mechanism is used to ensure the stability of the whole moral order, with all the rules it includes. Consequently, when rebel undermines the power of the mechanism, he affects negatively the stability of every moral principle in the system. To perform such action, i.e., to be a rebel, the individual must prefer to be outside of the moral order rather than in it. If the individual dislikes a particular norm, but accepts all the others, and the others rules provide him more advantages than the single norm he dislikes disadvantages him, then it is irrational for him to confront the sanctioning mechanism. For doing so would result in the destabilization of the whole system, including the norms the rebel accepts.

\textsuperscript{247} See Stemmer (2013), pp. 118-121. The norm that Stemmer holds to be the one with which all the others are compared is the \textit{Unterdrückungsverbotnorm}. 

156
tone that is too particularistic. It seems to miss the point that although some rules cannot be justified when considered in isolation, they can be justified when considered along with other rules. Consider the following example. An agent A may come to reject the norm R₁, whereas agent B endorses such rule. Because of A's rejection, the rule is deemed unjustified. In another moment, we may consider a norm R₂ that A accepts but that B rejects. For Stemmer's meta-norm, both norms are not justified. If norm R₁ is installed, it oppresses A whereas the installment of norm R₂ oppresses B. Despite that, a proposal to install both R₁ and R₂ can have a different outcome regarding A's and B's attitudes toward the rules. To see that, we have to imagine a case in which the installment of rule R₂ affords A a certain amount of benefits that overweight the losses that R₁ imposes on him, whereas the benefits produced by R₁ to B outweigh the losses he has with R₂. In such case, if R₁ and R₂ are simultaneously installed, they can be endorsed by both A and B. With the joined endorsement to both rules, then it becomes hard to see how any of the rules can be considered oppressive. It is correct that A would prefer to be in a situation in which only R₂ had authority over him than the situation where both rules are present. However, A still prefers to be under both R₁ and R₂ than in a state of nature, even though R₁, in isolation, displeases him.

The view of justification that we suggest above should not be understood as a criterion different from Stemmer's meta-rule. Instead, it is an interpretation of the meta-rule. The interpretation indicates that the fact that an individual rejects the normative constraints of a specific norm when the norm is considered in isolation does not warrant us to conclude immediately that such a norm is unjustified. We must consider individuals' attitudes toward the norm against the whole moral framework in which it is inserted. It may be the case that, as we argued, a rule displeases him in particular, but that his attitude toward such rule changes when the rule is examined against the whole moral framework. Our interpretation of the meta-norm is that the transition from the state of nature to the moral state is not done utilizing a series of moral agreements, each pertaining particular rules. Instead, the transition is done through a single moral agreement that contains several rules. In this sense, the moral agreement should be understood as a 'package' of rules. Every single rule in the package is justifiable if the individual, in the end, believes that the package with all the normative constraints it brings with it is acceptable.

The package interpretation of the moral agreement is vital to understand the attitude of the rebel toward the moral order. The rebel is someone who rejects the moral order as a whole and acts to destabilize the moral order because the final sum of advantages and disadvantages that every moral rule affords him is negative. It may be the case that some of the rules in the package are justifiable for him in light of Stemmer's particularistic meta-norm. Nonetheless, the framework contains at least one norm that rebel regards as so disadvantageous that it turns the moral order as a
whole as something prejudicial for him. It is because the moral order is, for the rebel, unjustified in this sense that he behaves as we discussed above.

In sum, the rebel is a representation of a particular rationale to non-compliance with moral rules. Different from the cheater, the rebel is non-compliant because he regards the moral order that he finds himself in as unjustified. The moral order makes him worse off in comparison to how he would be in a state of nature. As a consequence, the rebel prefers the state of nature over the cooperative state. The non-compliant attitude of the rebel can be controlled in some instance utilizing the sanctioning mechanism. The mechanism can force the rebel to comply by threatening him with sanctions. However, the mechanism is an inherently inefficient system. Its inefficiency is grounded on two factors. Firstly, the amount of goods and resources available in the system and the willingness of individuals to cede them to fund the mechanism. Secondly, the tendency that individuals have to free-ride on the production of stability that results in a reduction of the power of the mechanism. The inefficiency of the mechanism indicates that there are always opportunities for non-compliant behavior. It is in the context of such opportunities that the rebel emerges as a serious challenge to the stability of a moral principle. The rebel is not only non-compliant with the principle, but he also attempts to destabilize it. The rebel wants to destroy the framework that attempts to regulate his actions by means of the sanctioning mechanism. For this reason, he is determined to employ efforts and resources to undermine the power of the sanctioning mechanism in order to destabilize the moral order. The inefficiency of the mechanism indicates that rebel will invariably have chances to carry on destabilizing activities and thus to harm the stability of the moral order. That is why rational stability cannot be considered a matter of introducing appropriate sanctions exclusively. Sanctions are unable to contain the destabilizing effects of rebels fully.

3 REBELLION AND THE CIRCUMSTANCES OF MORAL JUSTIFICATION

One should not be carried away by the metaphor of the moral agreement and miss what the actual relevance of the rationale for rebellion is. So far we have argued that the rebel is an individual who prefers to abandon the moral state and return to the natural state. Of course that an intent to return to the state of nature is a metaphor, it does not refer to an actual intent that persons have. The state of nature is, after all, a mere hypothetical construction. In more realistic terms, the rebel's preference to return to the state of nature is a representation of a desire to renegotiate the terms that govern the
moral framework. Edward McClenen argues that two tendencies create problems of stability for a moral order. A tendency to defect from the normative requirements of the moral order and a tendency to press for renegotiation of the terms of moral cooperation. The first tendency, when isolated, refers to the case of the cheater. The second tendency, when conjoined with the first, refers to the case of the rebel. What stands at the center of the rebel's drive to non-compliance is a desire to renegotiate the terms of cooperation.

3.1 Renegotiation Expectations

James Buchanan suggested the idea of a renegotiation of the moral agreement. Buchanan argues that evaluating the justifiability of an actual cooperative framework can only be done by imagining that the actual framework – the status quo – was a product of an agreement. To determine an individual's attitude toward the status quo, we must perform a comparison. We must compare how advantageous for the individual in question the terms of cooperation are in the status quo with how they would be in a hypothetical framework resulting from an also hypothetical natural struggle. Whenever it is the case for someone that the difference between the advantages afforded by the hypothetical terms of cooperation outweighs those afforded by the status quo, this person is drawn to press for a renegotiation of the terms of cooperation.

The idea is that if the status quo was abolished, and individuals were thrown back into the state of nature, the cooperative framework that would emerge from the natural competition would be more advantageous than that which constitutes the status quo. Buchanan's view is, then, that an evaluation of the status quo is done through individuals' perceptions of how good or bad they would be in an imagined framework resulting from a state of nature. The difference between the total sum of advantages afforded by the individual's position in the status quo and that of the imagined framework determined what Buchanan terms 'renegotiation expectations'. The rebel is the individual for whom the renegotiation expectations are considerably large. For him, the difference between the advantages afforded by the terms of cooperation in the status quo and those in the ideal framework has a massive magnitude.

249 See Buchanan (1975), ch. 5.
250 There is an important reason for why a rebel appears only in case of substantive renegotiation expectations. Remember that we share Hayek's pessimism regarding individuals' capacities to understand the social phenomena that surrounds them. Given such pessimism, we must assume that individuals' knowledge about their chances in a renewed state of nature is at best precarious. To imagine how well one would fare after the state of nature is reinstalled would demand a volume of knowledge that individuals cannot grasp due to their cognitive limitations. In other words, it is highly uncertain what he outcome of a renewed natural struggle would be. Under such uncertainty,
by the terms as they are in the *status quo*. As a consequence, he presses to renegotiate the terms. Faced with a rejection from the other parties to renegotiate, the individual starts to act like a rebel in the way we considered above. He becomes non-compliant toward the cooperative framework and works to destabilize it in an attempt to force a renegotiation of the terms. The goal of the renegotiation is to reduce the difference between the advantages afforded by the hypothetical framework and those that he has in the *status quo*.

The rationale for rebellion is grounded on the existence of considerable renegotiation expectations. To properly understand the rebel and its challenge to the stability of the moral order, then, we must examine the factors that determine the size of the renegotiation expectations. Whichever such factors are, we already know that they change the individuals' perception about the justifiability of the moral framework. When the renegotiation expectations are too high, that means that the individual feels highly disadvantaged by the terms of cooperation in the *status quo*. The expectations, then, are connected with the justifiability of the framework itself. If it is correct that there is such a connection, then the emergence of considerable renegotiation expectations is connected with what we have called the *circumstances of moral justification*. If an individual who views the framework as justified changes his view and starts to see it as unjustified, it must be the case that some change took place among circumstances of moral justification. The rebel, we argued, is an individual for whom the moral order cannot be justified. In this vein, if a rebel appears in the social scheme, we must look for the cause of the appearance in the circumstances of moral justification. Such circumstances are the key factors that determine whether the moral order can be justified. However, despite their importance to the rationality of cooperation, they are not immutable, as we argue ahead. Alterations in the circumstances are possible, and they are the cause for the appearance of rebels in the social system.

We analyze now how alterations in the circumstances cause the appearance of rebels in the social system. We leave the first condition, however, outside of the analysis. Unsocial sociability is the circumstance that defines characteristics that individuals have. The circumstance may, in theory, suffer alterations. We can imagine that through the development of new technologies, such as those associated with human and moral enhancement, some of the elements of unsocial sociability could be altered. The introduction of moral enhancement technologies, for example, could reduce the individuals will always refuse to reinstall the state of nature, unless the benefits they expect to acquire from it are so massive that it becomes rational for them to take the gamble of reinstalling the state of nature. For this reason, an individual only becomes a rebel if his renegotiation expectations are considerable, so that the moral order he expects to emerge from the renewed state of nature is much more advantageous to him than the *status quo*. Otherwise, in case the expectations are not so substantive, it is hard to see how it can be rational to press for a return to the state of nature, in which it is unclear which outcome will follow.

---

251 We examine below how individuals react to a renegotiation proposal.
natural partiality that individuals have toward themselves and augment the role that altruism plays in persons' daily decisions. Such possibilities, however, are still somewhat farfetched. For this reason, we leave them out the discussion, and as a consequence, we also leave the first circumstance out. Our discussion, then, focus on the second and third circumstances, the social balance of power and the absence of widespread poverty.

3.2 The Social Balance of Power

Moral cooperation is only rational for everyone if the powers of each agent and coalition are in balance with that of all the others. Such balance is, we argued in the last chapter, one of the circumstances of moral justification. To understand the connection between the social balance of power and the emergence of a rebel in the social system, recall the idea of ascendancy. The balance of power is a representation of a distribution of power in which no individual or group has ascendancy over the others. The central alteration that the circumstances of moral justification might suffer is a shift in the distribution of power that creates an ascendant individual. Such an alteration has immediate consequences for the stability of the moral framework. The individual who becomes ascendant gains enough power to be able to exact his interests over all the other individuals and groups in the social system. The power he acquires augments drastically his ability to affect the incentive structures of the others with high chances of success while it reduces the power of the others to do the same thing toward him. Consequently, the moral framework loses its usefulness for the one who acquired ascendancy. For the individual or group with ascendancy, the moral framework becomes an unjustified hindrance on the pursuit of their interests. If we were to apply Stemmer's meta-rule to examine a framework that contains an ascendant individual, we would find that such individual regards the system as oppressive. He collects no benefit from the framework, whereas he has his freedom to promote his interests restrained by framework’s normative requirements.

Besides the change in attitude toward the framework's justifiability that the emergence of ascendancy causes, another crucial factor is that such change concerns the relation between an ascendant individual and the sanctioning mechanism. An individual with ascendancy over the whole social scheme is necessarily an individual with ascendancy over the social scheme's mechanism. Recall that the mechanism is a social product. It draws its power from investments of resources and effort that the social community makes. The mechanism, then, can never be more powerful than all those individuals who support it combined. With this restriction, it is easy to see why ascendancy
over the community implies ascendancy over the mechanism. An individual who is ascendant over
the community, must necessarily be ascendant over every other individual, either alone or
combined. For this reason, a constraining mechanism funded by such individuals cannot match the
power of the ascendant individual. It should be clear, then, that when an individual acquires
ascendancy over the community, he immediately becomes a rebel. He becomes a rebel because the
moral framework becomes an unjustified constraint on the promotion of his interests. The fact that
it becomes unjustified leads the ascendant individual to become non-compliant toward the
framework and ultimately destroy it utilizing his overwhelming power.

The appearance of an ascendant individual or group is, however, the most extreme form of
alteration that the balance of power can undergo, and which creates a rebel. A shift in the
distribution of power that creates an ascendant individual produces the immediate collapse of the
moral order. However, in order for a rebel to appear, the shift in the distribution must not be so
drastic. As we argued above, the rebel has a renegotiation expectation. When such an expectation is
considerably large, he is drawn to be non-compliant, to destabilize the cooperative framework, and
force its renegotiation so that the terms of cooperation become more favorable to him. In the
context of the balance of power, renegotiation expectations emerge from shifts in the distribution of
power. As it was argued before, in the contractarian thought-experiment, we imagine that the
cooperative framework emerges from the conflict in the state of nature. When a balance of power
emerges between individuals in the state of nature, moral cooperation becomes rational for
everyone. However, although cooperation becomes rational for all, such a change happens only
provided that the terms of cooperation are defined by the distribution of power that underlies the
balance. The terms of cooperation must be shaped by the inequalities in power that are present
below the balance.252

Consider a state of nature with three groups, the Julii, the Brutii, and the Scipii, in which the
distribution of power is 1:2:3 in terms of power units.253 There are several configurations of this
system that can produce a balance of power. There are several arrangements that the system may
possess, given such distribution of power. In several of them, a balance of power emerges. That
happens if the Julii and the Brutii join up in a coalition against the Scipii. Also in the case where the
Julii and the Scipii join against the Brutii. A balance of power is also present when no coalitions
between the three are formed. It is only when the Brutii and the Scipii join forces against the Julii
that a balance of power does not emerge in the system as a whole, although it appears between the
Brutii and the Scipii. Despite the several possibilities, imagine that a balance of power appears in

252 The balance depicts the absence of ascendancy, but not the absence of inequalities in power.
253 Assume again that an advantage of three power points is required to have ascendancy.
the form of a system without any coalitions between the three groups. In this case, the distribution of power that underlies the balance is 1:2:3. From such a situation of balance, the three groups can start moral cooperation. The cooperative framework that appears, nonetheless, must be shaped by the power inequalities between the groups. For this reason, we can imagine that the framework that appears is such that the advantages individuals gain with the framework, regarding their utility, have the pattern of 10:15:20, for the Julii, Brutii, and Scipii respectively.

Imagine now that time passes after the cooperative framework is established. As time went by, the distribution of power between the three gentes changed. With time, the power of the Julii increases. They become more powerful than they were before but remain very far from gaining ascendancy in the social system. Nonetheless, the increase in power changes the Julii’s attitude toward the moral framework in which they are inserted. With the newly acquired power, the Julii start to believe that if the state of nature between them and the other two groups is resumed, they would be able to secure a better position in the balance of power and demand terms of cooperation more advantageous for themselves than those they currently have. Let us say that their power augments from one power unit to three so that they now have the same amount of power that the Scipii have and the one more unit than the Brutii have. The new pattern becomes 3:2:3. With the power acquired, let us say that the Julii would, in a state of nature with the other two groups, be able to obtain terms of cooperation that would award them twenty utility instead of the ten they currently have. Such a change creates a renegotiation expectation for the Julii. They start to press to renegotiate the terms of cooperation and to alter it in order to make it produce for themselves advantages similar to those they would expect to acquire in an ideal framework.

The change in the distribution of power, it is important to notice, does not transform the Julii into rebels immediately. What they aim is to reshape the terms of cooperation observing the new distribution of power. To achieve such an objective, they must not necessarily become rebels and work to bring the whole system back to the state of nature. The most reasonable option is to demand the Brutii and the Scipii for a renegotiation of the terms of cooperation. The object of the renegotiation is the improvement of the Julii’s position in the framework. It is only if such proposal for renegotiation is rejected that the Julii become rebels. If the proposal is rejected, the current moral framework becomes a hindrance to the interest of the Julii of acquiring the terms of cooperation that they see fit to the new distribution of power.

254 Perhaps it is easier to imagine the distribution of advantages of moral cooperation regarding inequalities in power if we imagine that the natural competition was a competition for goods. When the balance of power is established, individuals have different shares of goods in virtue of their differences in power. The more powerful hold more goods than the weak. When the framework is established, the powerful maintain the larger share of the goods they collected in the natural competition. In this context, it is perhaps more appropriate to refer to an ‘initial imputation of rights’ over the goods instead of terms of cooperation.
The attitude of the Brutii and the Scipii toward the renegotiation proposal of the Julii is always negative. Granted that the renegotiation improves the standing of the Julii with regards to the terms of cooperation, it implies that it worsens that of the Brutii and the Scipii. The Julii's request of an alteration of the terms of cooperation is a request to alter the structure of rights and obligations based on the normative requirements of the framework. Such alteration implies necessarily one of two things. On the one side, that the Julii acquire rights that they did not have before, and thus that the Scipii and the Brutii are given certain obligations, which correspond to the Julii's new rights. On the other side, it can imply that the Julii are freed from certain obligations. The gained freedom necessarily implies that the Brutii and Scipii lost certain rights that they had over the Julii. In either case, the Julii's acquisition of a right or their liberation of an obligation, the Brutii and the Scipii have their conditions worsened. The two gentes do not want to worsen their situation to improve that of the Julii. For this reason, their willingness to renegotiate depends on a single factor. It depends on whether they are confident that the sanctioning mechanism can coerce the Julii into compliance, should they reject the proposal for renegotiation. The refusal means that the Julii convert themselves into rebels and become non-compliant with the moral order. Insofar as the Brutii and the Scipii believe that the sanctioning mechanism can assure the stability of the moral order, they reject any renegotiation proposal.

With regards to the capability of the sanctioning mechanism to cope with the increase of power of the Julii, we may differentiate three scenarios. The first case is the one we examined at the start of this section. It is the case where the power acquired by the Julii makes them ascendant over the community. In this case, the shift of power that takes place is so drastic that the Julii become much more powerful than the Brutii and the Scipii, either isolated or together. As a consequence, the Julii also become much more powerful than the sanctioning mechanism in the system. In such a scenario, the moral order collapses entirely.

The second case represents an alteration in which the Julii do not become more powerful than the mechanism, but the mechanism loses the ascendancy it had over the Julii. In this case, it is likely that the state of nature is reinstalled and the natural struggle resumed. In this case, the Julii must not be as powerful as the Scipii and the Brutii together. Remember that the power of the mechanism is not necessarily the same as the total sum of the powers of every individual that supports it. The mechanism is inherently inefficient. The power it has is not a depiction of the sum of the power of every individual that supports it. For this reason, if the mechanism loses ascendancy over the Julii, that does not mean necessarily that the Julii have become the strongest unit in the system. Differently from the case above, in this case, individuals resume the natural competition from which the terms of cooperation, which the Julii expect to be superior to those they had before,
The third case appears when two conditions obtain. The first condition is that a shift of power happens that makes the Julii considerably stronger than they were before. The newly acquired power leads the Julii to press for a renegotiation of the terms of cooperation. The second condition is that although the shift changes the Julii's amount of power considerably, the change is not drastic enough to eliminate the ascendancy that the sanctioning mechanism has over the Julii. The ascendancy of the mechanism indicates that the Brutii and the Scipii will not be open to the Julii's renegotiation proposal. They will expect the mechanism to keep the Julii compliant with the moral order. In this case, the Julii become rebels as we have been discussing throughout the last section. They become individuals who are non-compliant with the moral order and who act to destabilize the cooperative framework in order to force the renegotiation they desire. In the other two cases, the alteration in the distribution of power is so drastic that it produces an immediate collapse of the moral framework. In such contexts, it seems meaningless to comment on the effects of the Julii over the moral order because the order is immediately destabilized and collapses.

The third case is the only case in which the Brutii and the Scipii would refuse the renegotiation proposal of the Julii. In the first case, the proposal becomes irrelevant because the Julii can obtain whichever terms of cooperation they desire using their sheer power. In the second case, the renegotiation proposal presents itself as an alternative to a return to the state of nature. Since in this case the sanctioning mechanism has lost ascendancy over the Julii, the Julii become fully non-compliant with the moral framework, and they restart the natural struggle with the Brutii and the Scipii. The alternative the Brutii and Scipii have to avoid a return to the state of nature is to accept the renegotiation proposal. It is only in the third case that the Brutii and the Scipii reject to renegotiate outright. Since the mechanism still holds ascendancy over the Julii, they remain confident that the mechanism can protect their standing in the current moral framework that the Julii wants to worsen to increase their own.

The mechanism, thus, whose operation causes the emergence of rebels in the social system is an alteration in the distribution of power that underlies the balance of power and the moral order. We can imagine that every moral order is the product of a particular distribution of power. Insofar as the actual distribution of power between individuals remains similar to that of the imagined balance of power, the moral order is justified for everyone. However, the more the actual distribution shifts away from the imagined 'original' distribution, the more considerable will be individuals' renegotiation expectations. The more considerable the shift, the more individuals start to think that they could assure better terms of cooperation if the social scheme returned to the state of nature in order to produce a new framework from a renewed natural struggle. Such individuals
will start to press for a renegotiation of the current terms of cooperation. The renegotiation, however, implies that some individuals will become worse off than they are in the current framework. These individuals will undoubtedly refuse to renegotiate. As long as they believe that the sanctioning mechanism can assure the stability of the moral order against those that become non-compliant because the renegotiation proposal is refused, they will refuse the proposal. The consequence, then, is that the individuals who could improve their standing in a new state of nature become rebels.

3.3 ABSENCE OF WIDESPREAD POVERTY

We come to consider now how rebels may appear through changes in the third circumstance of moral justification, the absence of widespread poverty. As argued in the last chapter, a moral order can only be justified for everyone if there is a distributional baseline of goods that all individuals may agree to use as a baseline for moral cooperation. When there is no such baseline, its absence indicates that the amount of goods available is such that the majority of individuals lack freedom from poverty. In such circumstances, the moral framework faces problems of justification from the perspective of the poor. The view that poverty gives rise to rebels and problems of stability has already been suggested by Thomas Hobbes. Hobbes argued that needy men are likely to 'continue the causes of war' and 'to stir up trouble and sedition.' We develop this Hobbesian argument further here.

We argued in the last chapter that, in the context of poverty, morality should be understood as the imposition of restrictions over the means of power that individuals may employ to acquire goods. Restricting one's means of power to acquire goods results necessarily in a restriction of one's amount of power to acquire goods. The poor individual is the one who is unable to command the goods he requires to fulfill his basic needs inside the moral framework. However, assuming that there are goods available in the social system, it is important to consider the reason why the poor are unable to acquire the good. In this respect, there are two different scenarios, natural scarcity and artificial scarcity.

Natural scarcity is a scenario in which the amount of goods available in the social system is such that it cannot sustain any distributional baseline capable of eliciting a positive attitude toward moral cooperation from everyone. In natural scarcity, it does not matter how the resources are

256 In the case of JNL, such depiction of morality is particularly evident. In it, justice is defined by means of the rules that govern the means that individuals are allowed to use to acquire goods.
allocated, a majority of individuals will always be poor. In this context, whenever an individual acquires the goods he requires to be free from poverty, another individual necessarily becomes poor. Individuals are trapped in what can be depicted as a zero-sum game where the good over which they dispute is freedom from poverty. Moral cooperation is impossible in an environment of natural scarcity for there are no possible distributional baselines that can be widely accepted.

The most straightforward manner to imagine how natural scarcity can occur is to conceive of a natural disaster that affects the quantity of goods available in the system. A drought that incapacitates the agricultural production of a region is a good example. The drought reduces the amount of food available in the system drastically. Famine is expected to ensue. Those that can acquire the food will not negotiate it with the others. Those that lack the food will attempt to acquire it to the best of their capacities. In this context, the moral order serves as a hindrance to the poor and as a protection to those that have what they need. For this reason, although some, those in possession of foods, accept the normative requirements of the moral order gladly, others, the poor, see it as a hindrance for their fundamental general interest of being free from poverty. As a consequence, the moral order cannot be justified for the poor.

The second scenario we have in mind is that of artificial scarcity. This form of scarcity differentiates itself from natural scarcity due to one reason. In a scenario that depicts artificial scarcity, the amount of goods available in the system is such that the set of possible distributional baselines in which the goods can be distributed includes at least one baseline capable of eliciting acceptance from everyone. In a different manner from artificial scarcity, natural scarcity does not refer to a real lack of goods. Instead, it refers to an arrangement of goods that leaves individuals in poverty. A social scheme can be said to contain artificial scarcity if two conditions are met. The first condition is that the set of baselines the system can sustain due to the quantity of goods it possesses must contain at least one baseline in which virtually no individual is poor. The second condition is that the actual baseline of the system must be such that there are poor individuals in the system.

In a situation of artificial scarcity, the poor have the same attitude toward the moral framework that they have in the case of natural scarcity. The framework hinders the poor from fulfilling a fundamental general interest and, as a consequence, it cannot be justified to them.

---

257 Jared Diamond has examined how the emergence of scarcity of resources caused by environmental causes lead to the decline and downfall of civilizations. Diamond’s analysis refers to the phenomenon that we have in mind here. See Diamond (2004).

258 Amartya Sen's analysis in *Poverty and Famines* concerns what we term here artificial scarcity. Sen analysis famines that occurrence in the past (the Great Bengal Famine, the Ethiopian Famine, drought and famine in the Sahel region, and famine in Bangladesh) concludes that such famines were not caused by lack of food, but by people's lack of access to the food available. See Sen (1981), ch. 10.
However, different from natural scarcity, the case of artificial scarcity allow us to consider that what underlies the poor's non-compliance are renegotiation expectations. In natural scarcity, it does not matter which baseline is in place since there are always individuals forcing a shift to another baseline. There is never stability in the system.\(^{259}\) Such a characteristic does not appear in a system plagued by artificial scarcity. Since in such case there is at least one baseline that can elicit moral cooperation from everyone, there are chances for a renegotiation of the terms of cooperation.\(^{260}\) Whether a renegotiation takes place or not depends, as we argued in the last section, on the confidence that those that are disadvantaged by the renegotiation have in the sanctioning mechanism. Insofar as they believe that the mechanism can control the poor, they refuse to renegotiate with them.

Both artificial and natural scarcity can destabilize a moral order. The poor are, as we argued, likely to be non-compliant with the moral order. What determines the effect the poor have over the stability of the moral order is the amount of power they agglutinate. The more powerful are the poor, the less stable is the moral order. In the case of natural scarcity, the power of the poor threatens the existence of moral cooperation in general. If they manage to resist the sanctioning mechanism effectively, the moral order collapses and the whole system is thrown back to the chaos of the state of nature. However, since the quantity of goods available is such that every possible baseline contains a considerable number of poor people, individuals become trapped in the state of nature without a way out. They remain engaged in a perpetual struggle for resources.\(^{261}\) In the case of artificial scarcity, the gap in power between the sanctioning mechanism and the poor is also the central factor for the stability of the moral order. Insofar as the gap is such that the mechanism is ascendant over the poor, the system maintains some degree of stability and the non-poor refuse to renegotiate the terms of cooperation. In both cases, the power of the poor depends on how widespread is poverty in the social scheme. The more individuals are poor, then the more significant is their power to confront the sanctioning mechanism. Although a single poor individual behaves as a rebel, a single individual will not have any relevant effect on the stability of the moral order.\(^{262}\)

\(^{259}\) A clear difference between artificial and natural scarcity is that redistribution of goods is a measure that can restore stability to the system only in the case of artificial scarcity.

\(^{260}\) In the case of poverty, the terms of cooperation refer to the rights and obligations related to freedom from poverty. When artificial scarcity occurs, the renegotiation proposal that the poor advance is an alteration to include rights or remove obligations such that they are free from poverty after the alteration.

\(^{261}\) If we allow for the facts that individuals are removed from the system when they fail to supply their basic needs for a certain time, the consequence is evidently different. They remain locked in the state of nature until enough persons have perished so that a baseline able to sustain cooperation emerges.

\(^{262}\) In the next chapter, we examine the case of rebellion from a more material perspective. We introduce an economic and political dimension which underlie the rationale to rebellion. We see that although it is possible to conceive a powerful single rebel capable of causing major problems of stability for the moral order, such case is not possible in the case of rebellion caused by poverty. In the case of poverty, the central factor that determines the threat to the stability of the moral order is the number of poor individuals in the system.
The effects are entirely different, however, if we imagine that the poor refer to half of the individuals in the system. In this case, the amount of power directed against the sanctioning mechanism is such that significant problems of stability arise.

4 CONCLUSION OF THE CHAPTER

Our purpose in this chapter was to present the idea of stability. With references to Rawls and Gauthier, we presented what we termed the conception of rational stability. In this conception, the stability of a moral principle or moral order refers to the capacity of the principle or the order of engendering what Rawls has termed a *modus vivendi*. We saw that the creation of a *modus vivendi* depends exclusively on the compliance of individuals to the normative requirements of the principle or the order. Insofar as it is rational for them to comply, there is a *modus vivendi* in place. We moved them, to analyze the elements involved in the rationale for non-compliance and we suggested that non-compliance can be differentiated into two types, cheating and rebellion. Cheating refers to the rationale for defection present in the cases of Hobbes' Fool and Hume's Sensible Knave. The drive here is to defect provided defection affords advantages, and it does not compromise the maintenance of the moral order. The Fool and the Sensible Knave do not question the justification of the principle or order whose normative requirement they violate. The second rationale for defection that we suggested was that of rebellion. The rebel is the individual who is non-compliant with the normative requirements of the principle or order because he regards them as unjustified, i.e., as a form of oppression. The non-compliance of the rebel is more dangerous to the maintenance of a *modus vivendi* than that of the cheater because the rebel's end is to eliminate the principle or the order. His objective is to be free of their normative requirements. For this reason, a sanctioning mechanism is not as efficient to deal with the rebel as it is in the case of the cheater. The cheater sees the principle or the order as justified, and he does not want their destabilization. Cheaters contribute to and want to see the sanctioning mechanism working correctly. Rebels, on the other hand, act to undermine the sanctioning mechanism. To be free of the normative constraints they do not accept, they work to reduce the gap of power between them and the mechanism. Such an attitude has dire consequences for the maintenance of a *modus vivendi*. Although the mechanism can thwart the rebel in some measure, it suffers from inherent inefficiencies that indicate that the rebel always has opportunities to be non-compliant and act toward the destabilization of the whole system. We also inquired into the mechanisms responsible for the appearance of rebels in the social
system. For this purpose, we resumed our discussion of the circumstances of moral justification. We argued that what prompts the appearance of rebels are alterations and changes that pertain two of the circumstances, the social balance of power and the absence of widespread poverty. Rebels appear in the system when, on the one side, drastic shifts in the distribution of power occur throughout time and also when poor individuals appear in relevant numbers in the social system. In both cases, individuals who desire a renegotiation of the terms of cooperation appear. When their renegotiation pretensions are not fulfilled, they become rebels.

In the next chapter, the discussion remains on the topic of stability. However, our objective is to examine the stability of a moral order governed by JNL, the conception of justice we suggested in chapter two. Our objective is to argue that a moral order with JNL cannot be considered stable since it is entirely unable to deal with the problem posed by rebels. To perform such an analysis, we increase the level of materiality of the discussion and refer to economic and political aspects connected with the maintenance of a modus vivendi. Our conclusion in the next chapter opens up the way to our central argument, which is, that the stability of the moral order requires the inclusion of a rule of distributive justice with more substantive redistributive demands than JNL.
CHAPTER FIVE: THE INSTABILITY OF JUSTICE AS NATURAL LIBERTY

OVERVIEW OF THE CHAPTER

Our goal in this chapter is to fulfill a similar task to what Rawls has termed the 'second stage' of his contractualist method. Our task is to examine the stability of Justice as Natural Liberty (JNL), the conception of justice that we formulated in chapter two with the contractarian method. The task at hand is to examine whether a social system informed by JNL can be considered stable. In previous chapters, we investigated the relevant aspects of both the notions of justification and stability, and now we bring the conceptual framework presented in previous chapters to examine JNL.

Our argument in this chapter is, in some sense, an elaboration and expansion of a view espoused by Hobbes. Hobbes argues that there are several doctrines that either weaken the commonwealth or that lead to its dissolution.263 Hobbes lists several of such doctrines, and, among them, he mentions the attribution of absolute propriety over one's own goods to citizens.264 Such doctrine leads to the dissolution of the commonwealth, Hobbes believes. The reason he has in mind is that whenever individuals have absolute right over what they own, i.e., a right that excludes the state from interfering with their private property, the state is thwarted from fulfilling his tasks correctly. These are the tasks that individuals themselves have conferred the state, namely to protect them from their common injuries and from foreign enemies. Hobbes does not argue in detail why the attribution of absolute property thwarts the state from reaching its goals. That is our task in this chapter. We develop an argument to show how the attribution of absolute propriety weakens the moral order and leads it to its dissolution. We do so by arguing that conceptions of justice – JNL in

263 See Hobbes (1998), ch. 29 for all the doctrines.
the case of our argument – that cannot interfere with individuals propriety to perform redistributive efforts contribute to create instability.

The goal of the analysis here is to examine whether the normative demands of JNL have destabilizing consequences for the moral order as a whole. We examine whether what JNL allows and prohibits individuals to do can lead to alterations in the circumstances of moral justification. If JNL's requirements produce such alterations, it is likely that JNL itself is a factor that destabilizes the moral order. Consequently, JNL could not be considered a stable conception of justice, since its very normative requirements would lead to the destabilization of the moral order in which it is inserted. The analysis is done in two different contexts. In the first place, we examine the effects of JNL on the social balance of power. In this respect, we stress a connection between JNL and two crucial types of social power involved in the social balance of power, economic power and political power. In the second place, we examine JNL in connection with the absence of widespread poverty. Here, we examine whether JNL's requirements produce alterations in such circumstance of moral justification, i.e., we examine whether JNL augments the number of poor individuals in the system and whether it can contain their destabilizing effects.

1 PRELIMINARY REMARKS

We start with a short recapitulation of JNL. In the second chapter, we suggested JNL based on James Buchanan's discussion about the emergence of private property from a state of nature. We argued that the contractarian justification of private property entails a particular conception of distributive justice which we named justice as natural liberty. Distributive justice is, as we considered with Aristotle's view in the matter, about distributing goods in proportion to individuals' entitlements. A just distribution is the one in which there is an equal proportion between every individual's share of goods and the entitlement of each to his share. In our discussion, we suggested that the contractarian view can justify and endorse a specific standard to determine what entitles persons to goods, what we called the principle of natural liberty. The standard states that a person is entitled to a good if she was able to acquire that good in observance of the three rules of private property: acquisition, transfer, and rectification. A person is entitled to a certain good, then, if two conditions are met. The first condition is that the person must have taken the good for herself, she must hold some form of possession over the good. The second condition is that her possession of the good must be a result of a process that does not contain violations of the rules of private
property. When the two conditions are met, the person is entitled to the good she possesses. We combined this standard with Aristotle's concept of distributive justice. JNL is the outcome of the combination. JNL states that the just distribution can be defined as that in which each individual's share of goods is a product of an act of acquisition done in accordance with the rules of private property. As a consequence, whichever distribution emerges in observance of private property is a distribution that ought not to be interfered with, since it is just. It is in this sense that JNL represents the doctrine of absolute right over one's property to which Hobbes refers.

The view of stability that informs our discussion here is that of rational stability that we discussed before. For such a view, two factors are fundamental for the stability of a moral order. The factors are, firstly, the justifiability of the principle for those under its normative requirements and, secondly, the presence of appropriate sanctions. Sanctions play an essential role in thwarting non-compliant behavior, but they are not sufficient to ensure the stability of the moral order. The mechanism responsible for administering the sanctions is always inefficient and, as a consequence, there are always opportunities to violate the moral rules and avoid the sanctions. In virtue of such inefficiency, it is of fundamental importance that individuals remain committed with the moral order, i.e., that they remain regarding the moral order as justified. Otherwise, if individuals regard the principle as unjustified, they become rebels and start to work to destabilize the social system. The analysis we perform in this chapter is focused on the threat of rebels in the context of a moral order governed by JNL.

When we discussed JNL before, we argued that such conception of justice is justifiable by the contractarian method. Referring to Buchanan's discussion of private property, we argued that rational agents in a state of nature would agree to take part in a moral agreement with JNL as its content. However, one should not conclude from the argument for the justifiability of JNL that the stability of JNL follows. Justifiability does not imply stability directly. The stability of a moral principle involves its justifiability insofar as the normative requirements of the principle do not undermine its own stability. When we discussed the justification of JNL before, we considered solely whether individuals would agree on JNL as a rule of distributive justice to govern their society. However, we did not cover the matter of whether the normative requirements of JNL can alter the circumstances of moral justification. Such a question is still open, and it is with it that we are concerned in this chapter.

From the view of rational stability, the factors that govern the maintenance of the stability of a principle and its destabilization are the circumstances of moral justification. These circumstances make moral principles beneficial to everyone. JNL is justified in this respect. However, such circumstances are not unchangeable. They describe empirical states of affairs that can be changed.
by individuals themselves. From the fact that the circumstances can be altered, we can infer that
principles that were justified for everyone can become unjustified. With changes in the
circumstances taking place, the attitude of some persons' toward the moral principles that govern
their society may change. Some persons may start to see such principles as unjustified. These
persons may start to think that they would be better off without such principles than with them. The
matter of stability is intrinsically connected with the possibility of change in the circumstances of
moral justification, as we argued in the last chapter. To observe such connection, consider the
following example. Imagine a principle X and that this principle is justified under the circumstances
of moral justification. However, suppose also that the normative requirements of X are such that
individuals are driven, directly or indirectly, by its requirements to perform a series of interactions
with one another that alter one of the circumstances. Consider also that such alterations are drastic
enough to create substantive demands for a renegotiation of the terms of cooperation. A
consequence, then, of the interactions is that when they are concluded, some individuals will regard
the principle X as an unjustified principle and will, thus, become rebels and attempt to destabilize
the principle. Principle X is a principle that cannot be considered stable, even though it could be
justified under the circumstances of moral justification. This example indicates that the effects that
moral principles have over the circumstances of moral justification are a vital part of their stability.
Moral principles must be able to cope with changes in the circumstances in order to be considered
stable.

To analyze whether a social system governed by JNL remains stable throughout time, we
must decrease the level of formality of our discussion and make it more material in certain respects.
The reason for that can be apprehended from the paragraph above. The analysis of the stability of a
principle involves accounting for the effects that the normative requirements of the principle have
over the circumstances of moral justification. The requirements, nonetheless, are directed to the
members of the community. The principle demands them to act or not to act in certain manners. The
circumstances of moral justification are affected by the principle through the actions and
interactions of individuals under the framework established by the principle. The discussion must
consider the actions and interactions of individuals as well as the circumstances themselves from a
more material perspective in order to examine the effects that such actions and interactions have
over the circumstances. The principle requires individuals to do or refrain from doing specific
things. It is the effect of such specific things over the circumstances that determine whether the
principle is stable or not. In the case of JNL, it is a conception of distributive justice. It determines
how goods ought to be distributed and how they ought not to be distributed. The stability of JNL,
then, is intrinsically connected to how its normative requirements over distributions affect the
circumstances of moral justification.

The need for a more material consideration of the circumstances and the effects of the principles over them is mainly present in the case of the social balance of power. So far in our discussion, power was considered from a formal perspective exclusively. In chapter three, we discussed power in the context of the works from Downing and Dahl, but the discussion was considerably formal. We observed Downing's differentiation between outcome power and social power, as well as Dahl's four factors of power, but we did not address the matter of what empirical factors constitute power and what is it that makes different persons have different amounts of power.\textsuperscript{265} However, with such a formal view, we cannot conduct the discussion we plan to do here. That JNL is destabilized when the distribution of power that underlies the social system governed by it shifts drastically is a clear fact. However, such a statement is correct for every moral principle. The statement says nothing about the connection between JNL and shifts in the distribution of power. It may be the case that such shifts never come to happen in a system under JNL, for example. In such a case, JNL could be considered stable because it can cope with the threats to stability that changes in the social balance of power produce. However, to reach such conclusion, we must investigate how JNL manages to prevent or cope with shifts in the distribution of power, and for this end, we must consider power and its dynamics from a more material perspective.

Considering that JNL's normative requirements refer to the distribution of goods in society exclusively, it is then reasonable to assume that if JNL’s effects over the social balance of power and its ability to cope with changes in it are somehow connected with JNL’s normative requirements regarding distributions. Such a connection can only be investigated by associating the concept of power with that of the goods about which JNL is concerned. We perform such task here, and it occupies us for most of the chapter. That is the task of accounting for power in a material form, a form by which we can connect the normative requirements of JNL with shifts in power.

The need for a more material approach is not so prominent in the case of the third circumstance, the absence of widespread poverty, than it is for the social balance of power. We have already considered such a subject from a reasonably empirical perspective in chapters three and four. For this reason, our discussion about the capability of JNL to cope with the appearance of poor individuals in the social system is considerably shorter than the one concerning the balance of power.

\textsuperscript{265} We referred to Dahl's concept of power to argue that the amount of power that an individual has is determined by the basis and the means of power that belong to the individual in question. In this sense, these two factors determine the person's amount of power. However, that is not the question that we are raising here. Instead, it is a question about the material factors involved in determining the amount of power that someone has. The question can be seen as a question about the materiality of the basis and the means of power, a question concerning the actual objects that constitute these two factors of power.
We move, then, to the first and longest part of the chapter, the examination of the stability of JNL with regards to the social balance of power. The length of this analysis results from the development of a more material account of power, an account that can be connected with JNL. We argue that inside a social scheme governed by a moral order, two forms of power dominate the dynamics of power connected with the social balance of power, economic power and political power. We suggest an account of what each of these two forms of powers is about, how rebels may arise from shifts in their distribution, and finally whether JNL is stable with regards to the possibility of such shifts. We move then to the much shorter analysis of the stability of JNL in the context of poverty.

2 JUSTICE AS NATURAL LIBERTY AND THE SOCIAL BALANCE OF POWER

2.1 A MATERIAL ACCOUNT OF POWER

In chapter three, we covered Downing's and Dahl's discussions regarding the concept of power. We noticed that the idea of the balance of power refers primarily to social power, as Downing conceives it. Social power refers to the power of deliberately altering the incentive structure of another person in order to produce a specific outcome. It refers to the power that individuals have over other individuals. The balance requires that the distribution of social power be such that there are no individuals with ascendancy in the social system. However, insofar as social power is formulated in these terms, the balance of power can refer to wildly different empirical scenarios. We can speak of ascendancy in the context of a group of swordsmen, social classes, or even nations. Social power, as Downing suggests it, is highly abstract. Nonetheless, to examine whether JNL is stable or not, we need a more material view of power for the reasons we discussed above. JNL is a normative principle that refers to specific empirical states of affairs, i.e., distribution of external goods. To examine how the normative requirements of JNL affect the circumstances of moral justification, we must examine how JNL's requirements over the distribution of goods affect the social balance of power. For this reason, we must connect the idea of social power with that of external goods.

Dahl divides the concept of power into four factors. A more material account of power involves a more material approach to two factors in particular, the basis of power and the means of
power, what we may call the primary factors of power. These are the two factors responsible for determining individuals' amounts of power, which is the central concept in the idea of the social balance of power. Concerning the material form of the primary factors of power, we propose to focus the idea of a social balance of power in two particular forms of social power, economic power and political power. The relevance of these two forms of power arises from two factors. The first factor concerns that both of them allow us to connect the idea of power to that of goods and distribution of goods, which are the objects of JNL. As we argue ahead, goods are at the source of both forms of power. The second reason is that both of them are forms of power whose distribution can suffer considerable alterations. In a different manner from, for example, physical power, we can conceive, as we argue ahead, cases of ascendancy caused by shifts in political power and economic power without much difficulty. On the other side, it is hard to see how an individual might acquire ascendancy through physical power in any relevant sense. In sum, to conceptualize social power in the form of economic and political power helps our discussion because it allows us to connect the normative requirements of JNL to fluctuations in the distribution of power. As individuals interact, the distribution of several types of power is affected. Changes in the distribution of both these forms of power are crucial for producing the effects that lead to the rise of rebels. We consider first economic power and then political power.

2.2 Economic Power

2.2.1 What is Economic Power

Considered as a form of outcome power, economic power refers to the capacity an individual has of maximizing his utility in a market. It refers to the power individuals' have of commanding the goods they desire in the market. The central factors that are pertinent to this sort of power, i.e., the primary factors of economic power, are income and wealth. Wealth refers to the individuals' net assets. It refers to the accumulated goods and resources an individual has. Income, on the other hand, refers to resources and goods the individual can use without reducing his wealth. It is the possession of wealth and income that determines an individual's amount of economic power.\textsuperscript{266}

\textsuperscript{266} Income and wealth, the central factors of economic power, are functions of the share of goods and resources individuals have. However, it is important to notice that what determines the amount of economic power that an individual has is not the quantity of goods and resources that he possesses. Instead, it is the total value of the goods and resources possessed that determine the amount of economic power. The higher the value, the higher is his ability to command goods and resources in the market. Value is, in this context, market value. The value of a good
As a form of social power, economic power can be employed to deliberately alter the incentive structure of other individuals to produce a particular outcome. Wealth and income are powerful tools to promote one's interests against those of other persons. In the dynamics of economic power that permeate society, economic power appears as the power to build coalitions. By altering the incentive structure of an individual to produce a specific outcome utilizing the individual who was influenced, what the holder of economic power is doing is enlisting or using the individuals' primary factors of power in his, the holder of economic power, favor. Economic power allows its holder to have access to the basis and means of power of other individuals. It works as a tool to build coalitions for certain purposes. Consider the historical example of the Roman politician Marcus Licinius Crassus. Crassus managed to amass so much wealth that at a point in Roman history he was able to raise, train, and equip entire Roman legions at his own expenses. The legions were, in the end, loyal to him and not to Rome, what gave him enormous power over the citizens of Rome. Utilizing his economic power, Crassus was able to mobilize the primary factors of power of countless other individuals with the purpose of producing outcomes that he desired. Economic power allows one to easily make use of other individuals' primary factors of power in one's favor. A more common example of this sort of power is that of the banker who hires a thug to force a debtor to pay his debt to the banker. The economic power of the banker allows him to use the thug's muscular power and threatening appearance against the debtor. With his economic power, the banker creates a coalition with the thug to produce a specific outcome, namely forcing the debtor to pay his debts.

A coalition is, as we argued before, a cooperative system that includes a fraction of the social system. The power of a coalition is determined by the conjunction of the primary factors of power of its members. The advantage it offers its members is, firstly, that it establishes some form of security between them and, secondly, that it allows the use of the whole group's primary factors of power to shared goals and purposes. Economic power allows their owners to form coalitions because such power allows them to offer other individuals goods in exchange for the use of their primary factors of power. The logic that underpins a coalition formed by economic power is that of a mercenary army. Those that join the coalition join it in virtue of the economic advantages that they receive from its founder. On the other side, the one that forms the coalition does it to achieve some goal that he cannot achieve using his non-economic powers.

One should not assume that coalitions formed employing economic power are weaker or less stable than coalitions formed for other reasons. In any form of coalition, we assume individuals to be rational. Consequently, their participation in the coalition is determined by the fact that such or resource is determined by the supply and demand of the good or resource in the market.
participation is more advantageous for them in comparison with staying outside of the coalition. A coalition formed for defensive purposes, for example, can maintain itself as long as individuals feel that the coalition improves their chances of defending themselves against the others. However, as soon as such perception changes, their incentive to participate in the coalition changes as well. In the same manner, a coalition formed through an individual's or group's economic power remains stable and functional insofar as its founder or founders can supply the goods its participants require to remain in the coalition. A coalition formed through economic power is likely to be more efficient than other forms of coalitions. Since the exclusive interest of the members of the coalitions is, except for the founders, to receive some form of transfer from the founders, it is likely that few or even no problems of coordination regarding the realization of actions to attain the goals of the coalition will appear. Those in charge of the coalition will have much more autonomy to use the power of the coalition for the goals they desire.

Economic power is also connected with political power. However, we explore this connection ahead when we come to consider political power itself. For now, let us consider the matter of shifts in the distribution of economic power.

2.2.2 Shifts in the Distribution of Economic Power

Now that we have a clear grasp of what economic power is about, we come to consider how such power is distributed in the social system and how its distribution changes. The first question is simple to answer, we believe. The distribution of economic power is the distribution of wealth and income. The wealthier an individual is, the more powerful he is. The second question is more complex, and it involves subtleties that are important for our discussion. This section concerns the examination of such a question.

It is useful to recall the contractarian thought-experiment in order to understand the mechanisms that govern the shifts in the distribution of economic power. After a balance of power

---

267 Another essential aspect to keep in mind is that the power of coalitions is not determined necessarily by the number of persons in the coalition. The amount of power of the coalition is a product of the conjoined primary factors of power of its members. Naturally, a coalition with a large number of individuals can be considered powerful in virtue of the mere sum of each individual's physical power. Despite the importance that physical power can have in the determining the power of coalitions, other important aspects determine the power of coalitions. Cohesion and technology available are examples of such aspects. History is abundant with examples that display the importance of factors other than numbers in determining the power of coalitions. The Greek victories against the Persian empire; the Spanish conquest of the Americas against the Aztec and the Inca; the imposition of the so-called Century of Humiliation on the Chinese Empire by the hands of European powers, all these cases are examples of the importance of factors such as cohesion, organization, and, mainly, technology, in the determination of a coalition's power.
is installed in the state of nature, individuals convene and agree on a series of rules to establish a cooperative framework and a sanctioning mechanism for their social system. At the initial point of cooperation, the distribution of wealth and income, and thus of economic power, is such that there is a balance of power between individuals. At the initial point, that is the only characteristic that the distribution of economic power must necessarily have. However, the creation of the cooperative framework creates the conditions that will provoke shifts in the initial distribution of economic power. The two mechanisms that drive alterations in the distribution of wealth, and thus economic power, are trade and production. In the case of an imagined recently founded moral community, these two factors cause the initial distribution of goods, or the initial balance of power, to shift. Trade in the market and production are activities that necessarily imply alterations in the share of goods that individuals have. Such alterations necessarily affect the individuals' wealth and thus end up also altering the distribution of economic power. In strict terms, when the moral framework is established, the initial distribution of economic power is altered as soon as the first market exchange or productive activity takes place.

The shifts that trade and production cause have an interesting characteristic. A specific shift in the distributional baseline of economic power is caused by the sum of every trade and productive activity that took place during a specific period. Every single trade or productive activity realized is performed for the sake of a clear goal, the utility-maximization of those involved in such activities. Despite this fact, it would be wrong to affirm that those involved in the shift of the societal baseline intend its alteration. What each has in mind when he performs a trade, or a productive activity is an immediate benefit that the operation affords him. The emergence of a societal pattern from the sum of trades and productive activities is not part of the objectives that lead individuals to perform such activities. Individuals are causally responsible for the emergence of the baseline, that is true. Without their actions, the shift toward the new baseline could not occur. However, the creation of

---

268 The establishment of the framework and the mechanism prompt the emergence of a market in the system. The fact that individuals have different preferences, plans of life, and tastes indicates that they can improve their positions through trade. In the market, they can engage in commercial interactions with each other to acquire goods that they want in exchange for goods that they have but do not value as much as those that they want. The market is a consequence of the security in possessions created by the moral framework. Such security is not only responsible for the emergence of a market, but also for an increase, or appearance, of productive activities in society. A productivity activity consists in the employment of economic resources – labor, capital, or land – to create new goods or to introduce new goods into the market. Such form of activity is responsible for augmenting the overall amount of goods available in society, and consequently the overall wealth of society.

269 Trade and production always affect the distribution of economic power. Through trade, individuals exchange different goods with one another. The exchange necessarily results in a change in the overall wealth of individuals' involved in the exchange. Production has the same effect over individuals' wealth. Once a good is produced, someone will have its ownership. Maybe this person is the one that produced it or another individual, as in the case of a commissioned production. In both cases, production augments someone's share of goods, thus altering the distribution of wealth and economic power. In sum, an individual's economic power is set by the individuals' wealth. This factor, in its turn, is determined by the overall value of the goods that the individual possesses.
the new pattern is not a goal intended by those involved in the social system. What individuals seek through trade or productive activities is their immediate utility-maximization and only that. Individuals would not alter their behavior if, for example, the pattern that resulted from the sum of all activities was different, what indicates that they are indifferent to the final societal pattern that emerges. If those involved in shifts in the distribution of economic power do not intend them, there is an important thing that we can conclude about such shifts. They are not consequences of collective coordination. To think that the shifts involve some form of coordination of those involved in the distribution requires assuming that the shift is in some sense intentional. That cannot be the case.

Shifts in economic power are also connected with another factor that we considered in previous chapters, epistemological ignorance. When individuals participate in the market, they know the possible short-term outcomes of their interaction for themselves and those with whom they interact. With regards to future consequences of their interactions, individuals are most likely ignorant. A person that pays for a ticket to watch a football match knows that she will give money to the club, the club will give her the ticket, and she will be able to enjoy the game with the ticket. This person is also aware that the money she pays for the ticket goes to the club, which uses it to pay the stadium's staff and maintenance as well as the salaries of its team members. Nonetheless, the person is almost certainly ignorant about how each person in the staff and the team spends their salaries. As the chain of exchanges distances itself from the trade or productive activity performed,

---

270 In some sense, that is Adam Smith's influential idea of the Invisible Hand to which we are referring here. See Smith (1976), p. 456.

271 Robert Nozick famous Wilt Chamberlain case illustrates this point very well. Nozick's example starts with the consideration of a society whose resources are organized following the distributional pattern D1. He then adds that this society contains a very talented basketball player, Wilt Chamberlain. Wilt Chamberlain is so talented that persons crowd stadiums to watch his games. Chamberlain's capacity of attracting huge crowds to stadiums is a massive incentive for Basketball clubs to hire him to play in their team. The revenue that clubs expect to have if Chamberlain plays for them is so large that clubs are willing to pay him a fortune in order to sign him up. The clubs' intentions to sign Chamberlain drive them to compete to see which of them can offer him the most attractive contract. The more money Chamberlain expects to acquire with a contract, the more attractive it is. The club that manages to hire Chamberlain, Nozick asks us to consider, is the one that offers him the highest margin of the value of each ticket sold for his matches. Under these terms, and because Chamberlain's talents ensure that the stadiums are always full, the distributional pattern in society shifts to a pattern D2. In D2, Chamberlain has an immense fortune that he did not have in D1. The shift from D1 to D2 is caused almost exclusively by one single factor, the interest that persons' have in watching Chamberlain playing. It is such interest that results in crowded stadiums, large revenue for the clubs, and the creation of Chamberlain's fortune. Such outcomes, nonetheless, are not intended by those that attend the games. Individuals buy tickets for Chamberlain's games with a single purpose, to see his display of skill as a basketball player. The creation of affluence for the club or Chamberlain does not figure among the reasons that move people to buy the tickets. The shift from D1 to D2, concerning the emergence of fortunes for both the club and Chamberlain, is an unintended consequence from the part of those that want to watch the games. Consider that the government decides to tax 80% of the revenue that basketball clubs acquire from tickets sales. In such a case, if individuals continue to behave as they do in the shift from D1 to D2, D1 will not shift to D2, but to a pattern D3. In D3, the clubs are not wealthy as they are in D2. Despite this difference, it is clear that individuals would behave exactly in the same manner that drove D1 to D2, with the difference that the shift would lead D1 to D3 this time. To make the club wealthy is not part of the goals of those attending the stadiums. See Nozick (1974), pp. 160-162.
its effects for other individuals and the whole distributional baseline become extremely hard, perhaps impossible, to predict or anticipate. Let us say that one of the team members decides to use his wealth to create a charity organization. He then installs such charity organization in the neighborhood of someone that went to one of his games, although the player is entirely ignorant about who are the persons that live there. Consider, then, that the charity organization starts to assist the person there that has been to one of the player's matches. In such a case, the person that went to the match was a causal contributor of the installment of such organization in her neighborhood. Despite that, she could not have predicted such an event only from the trade she performed, when she exchanged money for a ticket to the match.

Epistemological ignorance and the unintentional nature of shifts in the distribution of economic power indicate that such shifts are either entirely outside of individuals' control or that they can only have minimal control over it. The shift is produced by the conjunction of many interactions between an enormous amount of individuals. To force a shift toward a specific pattern, one would have to coordinate or coerce a considerable amount of individuals to perform a series of actions in the market to produce the desired shift. However, even if we assume that such coordination or coercion could be done, the epistemological ignorance of individuals still poses an insurmountable problem to the success of such an endeavor. An attempt to produce a certain distributional baseline of economic power requires an intricate and detailed knowledge of the causal chain that must be followed in order for the baseline to emerge. The 'planner' must know exactly which actions each individual in the system must perform to achieve the desired objective. Such knowledge, it is reasonable to assume, is unattainable. Consequently, we can conclude that shifts in the distribution of economic power are both impossible to control and to predict at a societal level.

In sum, the distributional baseline of economic power is altered through individuals' interactions in the market. Such interactions appear either in the form of trade or productive activities. In both cases of market interaction, the distribution of wealth and income is altered, and thus also the distribution of economic power. We also stressed that such alterations, from a societal perspective, are unintended. Although every single interaction is done for a particular reason or purpose, the baseline that emerges from the sum of all interactions is not intended by those that participated in the particular interactions. Another important factor is that because of the epistemological ignorance that defines the individuals' cognitive capacities, the shifts in the distribution are either extremely hard or impossible to predict. The sheer volume of information involved in the causal chain leading one distribution to another makes it unreasonable to think that individuals could grasp the actual intricacies of the shifts.
We come now to an essential part of the discussion in the chapter. We consider the stability of a social system under JNL in the context of shifts in the distribution of economic power. We examine first how such shifts create rebels in the system and then whether a system under JNL would be able to cope with such a threat to its stability.

As the economic power of an individual grows, the gap of power between him and the sanctioning mechanism that protects the moral order diminishes. The reduction of such gap indicates that the individual increases his chances of avoiding being sanctioned by the mechanism. Higher chances of avoiding sanctions are incentives to defect from cooperation. Such incentives turn the individual either into a cheater or a rebel. The cheating opportunities that appear through economic power are intrinsically connected with political power. We shall examine such connection in detail ahead when we come to the consideration of political power. For now, we examine the case of individuals who become rebels in virtue of shifts in the distribution of economic power.

We follow Niccolo Machiavelli in the belief that the threat to stability that power offers come much more from those that already have significant amounts of power rather than those that want to have some power. It is intuitive to think that those that already have power would be a force acting toward stabilizing the status quo, whereas the seekers of power would act as a destabilizing force. According to this intuition, what the powerful would seek is the consolidation of their position, which implies some form of stability, whereas those that want power would want to break the status quo to become themselves powerful. However, we side with Machiavelli in thinking that such intuition is wrong. Those that seek power are indeed a destabilizing force, that is correct, but what the intuition gets wrong is that the powerful are not a force that acts toward the stability of the system. The mistake lies in the belief that the powerful cease to seek more power because they are already powerful. If that were true, they would indeed be a stabilizing force in the system. The problem is that the powerful are likely to behave in the same manner that those that are less powerful than them and desire more power. The powerful also seek to increase their amount of power. The source of this behavior is the same anticipation logic that drives the war of all against all in Hobbes' state of nature. As Machiavelli argues, the powerful fear to lose their power. The powerful fear the reduction of the gap of power between them and the others. The fact that they are powerful is not enough to make them feel secure. For them, the goal of 'consolidating' power is an unattainable goal. In this sense, their perception is that they are only able to maintain their position.

---

regarding power if they remain seeking more power.\textsuperscript{274} Both aspects of Machiavelli's views, that the powerful are a more destabilizing force than those that are not powerful and seek power, and that the powerful remain attempting to increase their power despite the superiority they already possess are present in the context of economic power.

Concerning economic power, those who have higher amounts of power are the ones that can become rebels. An individual thinks that the moral order is justified insofar as he thinks that it is beneficial for him. However, the more power an individual has, the higher is the likelihood that the moral order will become prejudicial for him. The reason for this change in attitude lies in the fact that the more power an individual has, the abler he is to exact his interests upon other individuals whilst the others become less able to the same to him. In this sense, his power to exploit the others augments, and his vulnerability to exploitation diminishes. With the reduction of the vulnerability, the moral order starts to look more like something that thwarts the promotion of one's own interests rather than something that helps the promotion of such interests.

In the contractarian metaphor, every moral order is based on a hypothetical initial balance of power. That is the hypothetical balance from which the moral order emerged. Insofar as the actual distribution of power remains similar to that of such hypothetical balance, the moral order remains justified for every person. However, when the distribution of power underneath the balance shifts toward a new baseline, the attitudes of some individuals toward the moral order may change as well. Such changes may turn individuals into rebels, i.e., individuals who regard the moral order as unjustified and who act to destabilize it with the end of forcing the adoption of a moral order more favorable to them. They are individuals with substantive renegotiation expectations regarding the terms of cooperation. In the previous chapter, we argued that shifts in the distribution of power cause the appearance of rebels in the system in three different manners, depending on how drastic is the shift. Let us consider these three manners in the case of economic power so that we have a better grasp of how precisely economic power leads to rebellion. Each case refers to a specific difference in the amount of power between an individual or group and the sanctioning mechanism. The first case is that of a shift in the distribution of economic power resulting in an individual or group becoming ascendant over the sanctioning mechanism. The second case is characterized by a shift toward a distribution of power in which the sanctioning mechanism loses its ascendancy over an individual or group. Lastly, the third case happens when the shift causes certain individuals to have

\textsuperscript{274} In chapter three, when we discussed the social balance of power, we suggested an example involving four individuals Frodo, Sam, Merry, and Pippin. We stressed in the example that individuals have two general attitudes toward power. If they have the upper hand in the distribution of power, they want to increase the gap of power. On the other side, if they are the worst-off in terms of power, they want to diminish the gap. The powerful want to increase the gap, as Machiavelli argues. Increasing the gap is the sole manner they have of ensuring their favorable position in the distribution of power.
considerable renegotiation expectations, what leads them to regard the moral order as unjustified but they remain compliant with the moral order because the sanctioning mechanism remains ascendant over them.

The key to understanding how any of the three cases may come to happen in the context of economic power is to recall that economic power is, in our current discussion, the power to form coalitions. The more economically powerful an individual is, the more powerful is the coalition that he can form. It is the power of the coalition that the economically powerful individual can create that determines whether he becomes a rebel or not and which of the three cases we mentioned above occurs. Two factors connected with the power of the coalition one can assemble utilizing his economic power decide whether such an individual is a rebel or not. The first factor is that in the hypothetical case of a return to the state of nature, the individual must be able to fend for himself using the coalition. For this purpose, the coalition must able to defend the individual from those outside of it successfully, and it must be able to maintain itself cohesive and stable. In this sense, the coalition must be powerful enough to protect its head, and it must be able to establish a stable form of security internally. That is the first condition that must be met for an economically powerful individual to become a rebel. The second condition is that the coalition one's economic power allows oneself to raise must be powerful enough to, in the context of a competition in the state of nature, secure for its founder a position in the balance of power that affords him better terms of cooperation than the those he has in the \textit{status quo}. The coalition must be able to secure him a better position in a hypothetical new moral order than the one he currently has. If both conditions are fulfilled, the economically powerful individual becomes a rebel.\footnote{Recall that, as we argued in chapter four, an individual in such situation presses to renegotiate the terms of cooperation firstly. He becomes a rebel only when the renegotiation proposal is refused by the others. This difference, however, is not that relevant since, as we argued also in chapter four, the others are highly motivated to refuse the renegotiation proposal.} In such circumstance, one of the three cases that we mentioned above occurs. What determines which case occurs is the amount of power of the coalition that the economically powerful individual can create. The first case occurs when the coalition is powerful enough to have ascendancy over the sanctioning mechanism. In this case, the individual and his coalition completely overpower the sanctioning mechanism and those outside the coalition. Such power allows the economically powerful individual to force the institution of a moral order under new terms, and in which he is much better off than he was in the \textit{status quo}. The second case happens when the coalition is as powerful as the mechanism is. Such a case represents the scenario of a \textit{civil war}. The coalition starts to interact with individuals outside of it as if they were in a state of nature. Its goal is to seek a balance of power from which they, and in particular its founder, can negotiate better terms of cooperation for themselves. The third case
occurs when an individual with economic power can build a coalition powerful enough to improve his position in an imagined state of nature, but the coalition remains less powerful than the sanctioning mechanism. In this case, the moral order does not collapse immediately. The economically powerful individual is not powerful enough here to challenge the sanctioning mechanism. However, because he is a rebel, he will become non-compliant with moral order. Moreover, he will also work to reduce the power of the mechanism. Both tendencies have highly destabilizing effects on the social scheme. As we argued, the rebel will explore the inefficiency from which the sanctioning mechanism suffers to weaken it and, through this process, to destabilize the social system.

A critic of the position we expounded above could argue that it is wrong to think that economic power can turn an individual into a rebel. The reason for that, the critic would say, is that the usefulness of economic power depends entirely on a cooperative framework. It is only in a context in which there are property rights that one can adequately exercise economic power. For this reason, an economically powerful individual cannot possibly have the intention to return to the state of nature or to destabilize the moral order, since his power, in such a condition, would become useless. The critic's argument is, in essence, that economic power cannot be transferred from a cooperative condition to a state of nature. The collapse of the cooperative framework would drag down with it the institutional network that is vital for the employment of economic power, thus making it a useless form of power. That is an interesting objection. However, the objection misses an essential point about economic power, and it also misses the point of our discussion.

The objection relies on an implicit assumption that we have considered and dismissed in chapter three. That is the assumption that coalitions are not possible in a Hobbesian state of nature. It is true that economic power requires some form of cooperative framework to work. It requires some recognition of property rights to work correctly, and where such a framework is absent, economic power cannot, indeed, play any relevant role. Such a framework is indeed absent in a state of nature. However, it is necessarily absent only in the context of the whole social system. In the context of small groups, coalitions, such cooperative frameworks may be present. Recall James Buchanan's suggestion that the natural struggle may involve the emergence of several limited constitutional contracts until a 'final' contract takes place and unifies the whole social scheme under a single cooperative framework. Cooperative frameworks that sustain individuals' economic powers can exist in the state of nature through such limited constitutional contracts, i.e., coalitions. Part of

276 Bertrand Russell, in his commentary on economic power, calls attention for this fact. He argues that economic power is not a 'primary' form of power, but instead a derivative form. Russell argues that economic power depends on the willingness of the state to enforce the rights that individuals claim over their goods and resources. In this sense, economic power can employed only insofar as state authority is willing to enforce the whole set of institutional structures that constitute one's economic power. See Russell (1957), ch. 8.
the benefits that coalitions afford their participants is the security they establish between those that are part of them. In this sense, insofar as coalitions that contain moral frameworks are possible in a state of nature, as we argued before that they are, economic power can be transferred from a cooperative outcome to a state of nature.

The second problem with the objection is that seems to take too literally the contractarian metaphor. To say that economically powerful individuals prefer to return to the state of nature is but a metaphor. It is a metaphor for a desire to renegotiate the basic terms of moral cooperation, as we discussed in a previous chapter. The economically powerful individual feels that the cooperative framework is not justified. He thinks that certain restrictions it imposes on his freedom are but an exercise of sheer power through the sanctioning mechanism. His real goal is not to destroy the social system and the mechanism, but instead to alter the rules and institutions that he believes to make him worse off. For this purpose, economic power is an incredible tool in virtue of the capacity it gives its holder to mobilize support in favor of the changes he desires. In the contractarian metaphor, the coalition may sound like an armed band or a small army, i.e., a group that engages in physical defense of the economically powerful individual. However, that is a metaphor for a group that the individual can mobilize in support of the changes he wants to see in the moral framework he is inserted.

Shifts in the distribution of economic power have destabilizing effects insofar as they produce rebels. When the shifts cause the emergence of individuals that are economically powerful on the levels that we discussed above, such shifts have destabilizing consequences for the social system. We come now to the treatment of such destabilizing shifts in the context of a social system governed by JNL. Our purpose is to investigate whether such cooperative framework can withstand the destabilizing effects that such shifts cause. If it is, then we may consider it a stable cooperative framework.

The distribution of economic power in society shifts in accordance with two factors, namely trade and production. Regarding such activities, JNL has no restraints or normative requirements, except for the rules of private property. The sole normative requirements of JNL is that trade and production be done in accordance with the rules of private property. If the rules are respected, then the distribution that emerges from such activities is just. If there are no requirements regarding the distribution of goods, except respect for property, then we can also conclude that there is no requirement to shift the distribution of wealth in society to particular patterns. Provided that the activities that shift the distribution were performed in conformity with the rules of private property, the distribution that emerges is just. Such a lack of requirements over the circulation and distribution of wealth allows us to conclude, then, that JNL has no normative requirements directed
at shifts in the distribution of economic power.

In a social system governed by JNL, as long as individuals engage in trade and production observing the rules of private property, all the shifts in the distribution of economic power, and the patterns that emerge thereof are just. Such conclusion indicates that a social system under JNL cannot act to prevent the emergence of distributional patterns of wealth that create rebels. If a rebel appears in the social system through a series of trades and productive activities that were done in conformity with the rules of private property, he is entitled to the wealth he possesses, and preventing him from acquiring such wealth by some form of intervention would be an injustice from the perspective of JNL. A moral order under JNL, then, cannot deal with the threat of destabilization posed by rebels utilizing preventive efforts. To prevent the appearance of a rebel, one would have to employ measures that cannot be endorsed by JNL. Such preventive efforts could take one of two possible forms, either some form of confiscation or some form of prohibition to engage in trade or production. However, the employment of both forms of prevention is not allowed by JNL. Consider the example below.

Let us recall the case of the Roman Marcus Licinius Crassus again. For the sake of argument, imagine that the Roman Republic was a social system governed by JNL. Consider that at some point before Crassus reached the peak of his fortune, Roman politicians notice the risk that Crassus' ascending wealth poses for the stability of the republic. They know that if Crassus' wealth reaches a certain threshold, he will be in a position to exert his will over the republic and design the rules that define the moral framework of the republic as he prefers. They consider how to deal with this situation without violating the rules of their social system, JNL in particular. A suggestion that quickly appears is to confiscate a large chunk of Crassus' wealth. Although pragmatic, such an idea cannot be endorsed by JNL. Each talent of gold that enters into Crassus' coffers does so under the rules of the social system and JNL. To confiscate Crassus' wealth would be, then, an action that not only is not demanded by JNL but an action that is forbidden by JNL. Confiscation would amount to an injustice against Crassus. A different approach would be to stipulate that Crassus' wealth cannot pass a certain threshold, under pain of sanctions. Such measure, nonetheless, could not be properly applied in a social system under JNL. The purpose of such measure is to prevent the emergence of specific patterns, patterns in which Crassus' wealth offers considerable threats to the republic. However, JNL offers the politicians no rule or principle that they can use to endorse such restriction. The patterns in which Crassus has a tremendous amount of wealth are just distributions, provided they arise in conformity with the rules of private property. To prevent the emergence of such patterns should be, then, considered also an injustice, an arbitrary exercise of power that cannot be endorsed by the moral order.
Crassus' example indicates that a social system governed by JNL cannot prevent the occurrence of shifts in the distributional baseline of economic power that create rebels in society. Even though rebels threaten the stability of the social system, the normative requirements of JNL hinder any action to prevent shifts in economic power that create rebels, since the performance of such actions is unjust. Such restriction leaves moral orders that include JNL with two options to deal with the threat posed by rebels, to use the sanctioning mechanism to contain them or to consider their proposals to renegotiate the terms of the moral cooperation.

We have already considered in chapter four the use of the sanctioning mechanism to assure the rebel's compliance with the social system. The mechanism is indeed able to provide some degree of stability by threatening the rebel with sanctions in case of defection. Despite that, the mechanism is an inefficient tool to fight the rebel. It suffers from problems that the rebel can explore to decrease its power in relation to his. Moreover, the mechanism is only capable of assuring some degree of compliance from the rebel while it still has ascendancy over the rebel. As soon as the ascendancy is lost, the mechanism becomes useless to assure the stability of the social system. In sum, the sanctioning mechanism is far from able to assure the stability of the social system against the rebel. The problems that the sanctioning mechanism faces to constrain the rebel to comply are examined in more detail ahead when come to consider political power.

The last resort that a social system under JNL has to deal with the destabilizing effects caused by rebels is considering their renegotiation proposals. What rebels seek is, in essence, to alter terms that govern moral cooperation. They have renegotiation expectations that make them regard the framework as unjustified. Consequently, they want to alter the framework to accommodate such expectations. A measure, then, to end the destabilizing effects that rebels have on the social system is to accept the renegotiation proposals they desire. As it is the case of the sanctioning mechanism, accepting to renegotiate is not an adequate solution.

A first and clear problem with this approach is that as long as individuals trust the sanctioning mechanism's ability to constrain the rebel, they will refuse to renegotiate. We considered this topic in chapter four. Renegotiation, as we argued before, necessarily makes the other individuals in the system worse off. For this reason, it is likely that every person in the system but the rebels will be recalcitrant toward the renegotiation proposals. In matters of renegotiation, we are dealing with a case that can be described as a zero-sum game. In any outcome in which an individual has a gain in comparison with the status quo, at least another individual is made worse off. It is likely that those expected to suffer the loss will prefer to augment their investments in the sanctioning mechanism, in an attempt to make it more efficient against the rebels, rather than ceding to the renegotiation proposals.
The second problem is that a proposal for renegotiating the terms of cooperation may end up costing JNL itself. The element of the cooperative framework that the rebels desire to reform or remove from the framework may be the view of distributive justice of JNL. They may find that a framework shaped by JNL is unjustified and that a different conception of justice should orient the mutual interactions in the social system. In this case, the weakness of JNL is strongly in evidence. For such case demonstrates that JNL creates the conditions for its own destabilization and collapse. Individuals interacting under JNL may arrive at certain circumstances that lead them to reject JNL and to favor another principle of distributive justice.

The third problem is that accepting to renegotiate the terms of cooperation may solve specific threats to stability, but it does not solve the general problem of stability posed by rebels. If a renegotiation proposal is accepted, the cooperative framework that shapes the social system is reformed, and individuals resume their cooperative activities inside the new framework. However, as individuals engage in trade and production, the distributional baseline of economic power begins to shift again. The process that leads to the stability problem, which was solved by the renegotiation of the terms of cooperation, starts to operate again. It is reasonable to expect, then, that at some future point in time shifts in the distribution of power will produce individuals with substantive renegotiation expectations, i.e., potential rebels again. When these individuals appear, the social system under JNL becomes unstable again, and the social system must decide again whether they acquiesce to the new renegotiation proposal. If they refuse it, the problem of stability remains. If they accept it, the problem is solved, but only until a new considerable shift in the distribution of economic power happens. One could adopt the principle of always accepting renegotiation proposals. Such a principle would undoubtedly ensure the stability of social cooperation. This solution, however, is extremely inappropriate to address the matter of stability. Adopting such a solution involves adopting a principle of justice in the line 'To each according to his power'. That is the standard of power that we examined in chapter two. As we saw with Stemmer's discussion of such a standard, the standard of power cannot ground a conception of justice to govern the social scheme. Such a standard opens up the possibility of continuous renegotiation of the moral agreement, which is a problematic possibility, as Edward McClennen argues. Continuous renegotiation is likely to make individuals question the legitimacy of the moral agreement. What the

277 Imagine a situation in which the less talented members of society have managed to amass a large amount of economic power. In this scenario, such individuals may push a renegotiation proposal to alter the conception of justice that informs their society. They may demand the removal of JNL and the inclusion of a conception of justice based on a standard of entitlement similar to that suggested by Engels and Marx, which we considered in chapter two. If the proposal is successful, JNL is removed from the system since the new conception explicitly rejects the standard of entitlement that underpins JNL.

solution does is to condition changes in the terms of cooperation with changes in the distribution of power. When the whole structure of cooperation can be so 'easily' and frequently altered, it starts to lose the value it has as a set of rules to guide social cooperation. The value of such norms lies in the fact that they make interactions with other individuals predictable. They indicate which terms orient mutual interactions between persons. Individuals get from them a social system in which they can live their lives. If such order is bound to be altered as soon as a group emerges with a request for renegotiation, then the predictability they offer is sharply reduced.\(^{279}\)

Hobbes' intuition seems to be correct. Absolute propriety over one's own goods weakens and can even lead to the dissolution of the social order. In our analysis of JNL's stability, we verified why that is the case. A social system informed by JNL is, we can then conclude, not capable of solving the problem of stability that arises in connection with the distribution of economic power. On the one side, it cannot prevent the emergence of rebels. Since JNL associates justice with private property, it cannot prevent individuals from accumulating large amounts of wealth that will turn them into rebels. It must then deal with the rebels using the sanctioning mechanism or through renegotiation of the terms of cooperation. Sanctions, we have argued, are not enough to solve the problem. They are capable of forcing the rebels to be somewhat compliant with the moral order, but they do not change their view that the moral order is not justified. For this reason, and since the sanctioning mechanism is inherently inefficient, rebels have opportunities to destabilize the social system. The alternative left to solve the problem of stability, then, is to concede to the renegotiation requests that arise from changes in the distributional baseline of economic power. Such alternative, however, does not solve the problem adequately. It solves the problem temporarily until a new shift in the baseline occurs. As soon as such shift happens, the problem reappears. It does not matter how many times the terms of cooperation are renegotiated, as long as there is a market in the community, the distributional baseline of economic power will shift at some point, and new demands for renegotiation will appear. To adopt a principle to always cede to the renegotiation demands would solve the problem in general, yes. However, adopting such a principle runs against the very idea of establishing a moral framework to orient social cooperation.

To conclude, the stability of a social system under JNL is highly vulnerable to shifts in the distribution of economic power. Such shifts create rebels. Rebels are a powerful destabilizing force for the social system, and in the case of an order under JNL, society has no appropriate means to

---

\(^{279}\) Adopting the posture to always renegotiate the agreement has an additional problem. To always renegotiate the agreement in accordance with substantive renegotiation expectations can be interpreted as a conception of justice itself. A standard of entitlement similar to the standard of power that we considered in chapter two underpins such posture. The view is that the terms of cooperation should always be shaped by the distribution of power and that they should also change in accordance with the changes in the distribution. Adopting such posture means, in essence, to reject JNL.
deal with the risk presented by rebels. Hence, concerning economic power, JNL cannot be deemed a stable conception of justice.

**2.3 Political Power**

**2.3.1 What is Political Power**

We come now to the analysis of the stability of JNL with respect to the second type of social power that we connect with the balance of power, political power. The proceeding is the same of our analysis of economic power. We first advance an overview of what political power is about, then we proceed to examine how shifts in political power occur, and, lastly, we move to examine in how far a social scheme governed by JNL can cope with threats to its stability created by shifts in the distribution of political power.

In our discussion, political power is power over the state. An individual has political power if he can mobilize the state's power to produce outcomes he desires. For our discussion, the state is but part of the sanctioning mechanism. In the first chapter, we indicated that our contractarian outlook is nearer to that espoused by Hobbes and Buchanan than Stemmer's or Gauthier's. Our affinity with the Hobbesian contractarian outlook refers precisely to identifying the state as the central entity that makes up the sanctioning mechanism. As part of the sanctioning mechanism, the central task of the state is the assurance of the stability of the moral order. The political structure of a society is, in essence, the organizational and institutional frameworks that compose the state-part of the sanctioning mechanism. The mechanism requires a series of institutions and organizations to work correctly. It requires a body to decide which sanctions it shall apply to which violations, it requires a body to determine how the sanctions selected will be applied, also a body to carry out the sanctions themselves, a body to determine how the mechanism will be financed, a body to collect the financing to the mechanism, a body to allocate the finance in order for the mechanism to function correctly, and other sorts of bodies. Each of these bodies or institutions attend to a particular need that must be fulfilled for the correct functioning of the mechanism. The totality of those bodies composes the political structure of society, the state.\(^{280}\) From now on we use the terms

---

\(^{280}\) Our definition of political power draws from John Locke's definition of the same concept, which we briefly considered in chapter one. Locke identifies political power with the power of making laws with sanctions to regulate private property, enforcement of laws, and defense of the community against external threats. Locke also argues that all such goals of political power are pursue only for the public good. It is a matter of interpretation what 'public good' refers to. One can understand it as the 'good of the majority', the 'good of every single individual', or
'state' and 'sanctioning mechanism' interchangeably.

The central factor of political power is the political office. Possession of a political office determines the amount of political power that individuals have. Holding a political office grants specific control over the power of the state to employ it for specific purposes. What determines the purposes to which the state's power may be employed is the political office itself. Consider the examples of an absolute monarch and a tax collector. The absolute monarch is someone who possesses a massive amount of political power. He holds a political office that enables him to use a large share of the state's power for whichever purposes he desires. The fact that the amount of power of the monarch is large means that his political office enables him to use not only a large share of the state's power but also in a wide variety of cases. The tax collector, on the other hand, also has some political power, but a milder amount of power in comparison with that held by the monarch. Like the monarch, the tax collector can use the state's power. However, different from the monarch, he can use only a fraction of the state's power and only for a specified purpose, collecting taxes. Different political offices afford different amounts of power regarding how much of the power of the sanctioning mechanism their holder has access to and how often he uses it.

Although political offices are designed for particular purposes, we cannot infer that the holders of such offices always use the state's power following the office's goals. The holder of the office has, naturally, some form of obligation or duty to use the office exclusively for the purpose it was designed. However, as it is the case with every form of normative requirement in the contractarian perspective, actual compliance with the requirement is assured only through sanctions. The administration of sanctions, however, suffers from the problems we discussed in the last chapter. The case of a holder of a political office involves an additional complication. Such an individual is part of the sanctioning mechanism itself. To effectively sanction him, other parts of the mechanism must be mobilized to do so. Moreover, the higher the political power of the individual is, the harder it becomes to sanction him. The absolute monarch represents the extreme case. The political office he occupies affords him so much political power and control over the sanctioning mechanism that the mechanism has no chance or power to sanction him. We explore the topic of sanctioning holders of political offices ahead. In any case, because of the inefficiency of the sanctioning mechanism, there is a margin for holders of political office to instrumentalize their office in search benefits for themselves. Such form of instrumentalization can appear in two different forms, corruption and rent-seeking.

Footnote: the 'good of a particular class'. In any case, our discussion of political power does not touch this problem of interpretation. We focus of our discussion lies in the first part of Locke's definition, i.e., in what activities are involved in political power. In this sense, political power is the power to set up laws backed by sanctions. See Locke (2003), p. 101.
Corruption is characterized by the abuse of political power to obtain benefits for oneself. Such form of instrumentalization of political offices can be performed only by the holder of the political office. As argued before, a political office gives its holder the power to use the state to fulfill some purpose connected with the goals of the state in general. A political office always has specific goals to fulfill. It is to attain such goals that the holder of the political office responsible for a particular goal is given the capacity to use the state's power. Corruption appears when the holder of a political office abuses such power for his own gains.

Rent-seeking is the second form of instrumentalization of political offices. In the economic literature, rent-seeking refers to efforts and investment of resources of private agents to obtain advantages from the state. Rent-seeking is a phenomenon that often appears in connection with corruption. Through rent-seeking, individuals seek the creation of 'artificial advantages' utilizing some form of governmental intervention in market exchanges, i.e., in people’s interactions. Such advantages are artificial because they would not have emerged 'naturally' in standard market interaction under the moral framework. They appear solely because the sanctioning mechanism intervenes in some form in individuals' interactions. Rent-seeking revolves primarily around an agent that does not have a political office, i.e., it is centered in someone that does not have any significant amount of political power. Rent-seeking occurs when a person without political power convinces someone with political power to abuse such power in her favor. It is only through political power that the rent-seeking agent can obtain the advantage he desires to have. The central distinction between corruption and rent-seeking is the following. With the practice of corruption, what the holder of the political office aims at is abusing his political power to obtain gains for himself. On the other side, in rent-seeking, the political office is abused not for the sake of the holder of the political else but someone else.

State organizations and institutions are somewhat impersonal entities regarding the benefits such organizations and institutions provide. The state composes part of the sanctioning mechanism. Its purpose is to assure the stability of the cooperative framework in which individuals interact with one another. The benefits that the state create are, ideally, benefits for every member of the social scheme. That is why we stress that the state is impersonal. It does not exist to favor certain individuals or groups, but to favor everyone. The two forms of instrumentalization of political offices that we stressed, corruption and rent-seeking, constitute attempts to reduce the level of the impersonality of the state, i.e., to transform the state into an instrument for the promotion of the interests of particulars or groups. Francis Fukuyama refers to such phenomenon, the use of the state in one's own favor or in favor of others in contrariety with the impersonal status of political offices,
as *repatriominalization*. Repatriominalization refers to the phenomenon of social groups taking control of the state apparatus and instrumentalizing it for their own benefit. We return to this topic ahead, when we start considering stability in connection with political power.

Efforts to repatriominalize state institutions and organizations are, it is critical to stress, efforts that, when successful, always make at least one individual in the community worse off. The state is sustained by a collective investment of resources done by the community. In this sense, its primary factors of power spring from contributions that come from each individual. Such individuals invest their resources in the mechanism in the expectation that the mechanism works to assure the stability of the social system. Every resource invested is invested for this end. For this reason, whenever a holder of a political office uses it in a repatriominalization effort, he is making use of the investment of the community not for the goal set by the community, but instead for his own goal, that the community does not share. Consequently, repatriominalization activities make those that invested in the mechanism, and are not benefited by the activities, worse off.

In sum, political power is the power to use the state, either part of it or its whole. This form of power is exercised by means of political offices. The holders of political offices are entitled to use the sanctioning mechanism for the goals assigned to their offices. However, because there is a margin to instrumentalize the office to one's favor, individuals with political power engage in corruption or rent-seeking. Opportunities for rent-seeking attract individuals with high levels of economic power. An individual can use of his economic power to enlist holders of political offices in the pursuit of his own interests, i.e., to perform rent-seeking for him at the expenses of other individuals in the community, who fund the mechanism with their own resources. We have now a clear view of what political power is about. We proceed to the analysis of how shifts in the distribution of political power occur.

2.3.2 Shifts in the Distribution of Political Power

---

281 See Fukuyama (2011), p. 453 “The second form of political decay is repatriominalization. The favoring of family or friends with whom one has exchanged reciprocal favors is a natural form of sociability and is a default manner of human interaction. The most universal form of human political interaction is a patron-client relationship in which a leader exchanges favors in return for support from a group of followers. In certain stages of political development, this constituted the only form of political organization. But as institutions evolved, new rules were put in place to recruit on the basis of function or talent the Mandarin examination system, the devshirme in Turkey, the celibacy of the Catholic priesthood, or contemporary legislation outlawing nepotism in hiring. But there is constant pressure to repatriominalize the system. Individuals initially recruited into an institution on impersonal grounds nonetheless often try to pass on their positions to their children or friends. When institutions come under stress, leaders often find they have to give in to these pressures in order to secure political primacy or meet fiscal needs.” See also, pp. 81, 230, 312, 453.
We come to consider now how shifts in the distribution of political power occur. To perform this task, we proceed as we did in the analysis of economic power. We refer to the contractarian metaphor. It is vital to stress beforehand, however, that the dynamics that govern the distributions of political power are different from those that rule over the distributions of economic power. Political offices determine political power whereas economic power is defined by wealth and income. In the case of economic power, the market and productive activities are the central mechanisms that drive shifts in the distribution of power. Such mechanisms are connected with political power, as we shall see ahead, but only indirectly. The core mechanism responsible for the shifts is a struggle that happens inside the community for political offices. However, before considering such mechanism, let us see what the general principle behind the distribution of political offices is.

Recall the contractarian thought-experiment. Individuals convene in the state of nature to establish a cooperative framework. The moral agreement establishes the framework. Such an agreement has two key features. The first is the bundle of rules and norms that shape the cooperative framework. The second feature is the design of a sanctioning mechanism and the proceedings to establish it. The mechanism is one of the objects of the moral agreement. In chapter four, we argued that the deliberation over the sanctioning mechanism that takes place in the state of nature has two aspects, a formal aspect and a material aspect. The formal aspect concerns the structure and power of the mechanism. It is about finding the structural configuration for the mechanism that assures its ascendancy over the community and thus the stability of the social system. The material aspect, in its turn, concerns the proceedings to realize the sanctioning mechanism materially as it was designed in its formal aspect. Such aspect concerns how the community's resources should be mobilized and allocated to the mechanism in order to assure its planned and proper functioning. In the context of such a division between formal and material aspects, the matter of which political offices should be created and how they should be distributed falls in the formal aspect of the sanctioning mechanism. The answer to both questions determines, together, what is the political system that will rule the social system. In this sense, part of the moral agreement involves determining how political power is to be distributed in the community. As it is the case with the sanctioning mechanism as a whole, such distribution is determined by one single purpose, to assure maximum efficiency possible for the mechanism.

The efficiency of the mechanism is defined by the level of stability it can proportionate for the social system. It is this purpose that guides individuals in a state of nature in the design of the political system. Political offices are to be created and distributed in the manner that assures the maximum possible level of stability for the social system. Such a goal is achieved when the design of the political system assures that the mechanism has ascendancy over every single individual or
group in the community. To design a mechanism that has ascendancy over individuals that do not hold political offices, and thus do not have political power, is not too complicated. In this case, it is enough that the mechanism is more powerful than the individual. However, the establishment of ascendancy over individuals that hold political offices is a more complex task.

Holders of political offices are assumed to be rational utility-maximizers in the same manner that every other individual in the procedure. Such individuals will instrumentalize their political offices through corruption or rent-seeking if they find opportunities to do so. Sanctions must be introduced in order to hinder individuals from performing acts to repatrimonialize the state. However, holders of political offices are part of the sanctioning mechanism. They are in a better position to escape or resist the administration of sanctions than individuals without political power. To ensure that holders of political offices comply with the rules, one must appeal to an individual or group that holds more political power than the first individual. A higher amount of political power guarantees that those with a lower amount of political power remain compliant with the social system. The problem is that each new level of political power that we bring into the scene to assure compliance of individuals with lower amounts of political power also requires a higher level of political power to remain compliant with the moral order. They are also rational utility-maximizers and will instrumentalize their political offices if the sanctioning mechanism allows them to do so. Another individual or group with even more political power must be brought into the scene in order to thwart them. This problem repeats itself until we run out individuals or groups to constrain holders of political offices that want to perform repatrimonialization. Such a problem must be solved in order to assure the creation of a sanctioning mechanism able to assure the stability of the social system. It is a problem of constitutional design.

Machiavelli suggested the solution to this problem. Machiavelli has explicitly argued that the stability of a constitution depends on the presence of checks and balances on the exercise of political power.\textsuperscript{282} Political offices should be distributed in the manner that diminishes to the maximum possible the chances that holders of political offices have of engaging in acts of repatrimonialization or of avoiding sanctions. The distribution that accomplishes this goal is that in which the power of each holder of a political office is in balance with that of the others. When such balance is in place, no holder of a political office is capable of avoiding the sanctioning mechanism

\textsuperscript{282} See Machiavelli (1994), p. 94. “Those who know how to construct constitutions wisely have identified this problem and have avoided each one of these types of constitution in its pure form, constructing a constitution with elements of each. They have been convinced such a constitution would be more solid and stable, would be preserved by checks and balances, there being present in the one city a monarch, an aristocracy, and a democracy” and “But let us turn to the other particular characteristics of that city. I maintain those who criticize the clashes between the nobility and the populace attack what was, I would argue, the primary factor making for Rome’s continuing freedom.” For Machiavelli’s historical role as a proponent of constitutional design, see Ingersoll (1970).
when he defects from cooperation. The reason for that is that political offices are distributed in such a way that for every office there is always at least another office powerful enough to sanction violations of the holder of the first office. Political offices should be distributed following how well their distribution assures the stability of the social system. However, even if we assume the case of an ideal distribution of political offices, such distribution is likely to be changed as time advances. Let us return to the contractarian metaphor to see the reason for that and the mechanism that drives such changes.

The transition from the state of nature into the cooperative state involves the distribution of political power in such a manner that there is a balance of power, or checks and balances, between the holders of political offices. The presence of such a balance marks the hypothetical initial distribution of political power. However, as soon as the political order is established, a mechanism starts to operate that leads to changes in the distribution of power. The mechanism is a social struggle for political power. As the cooperative condition is established, individuals notice that through political power, they can obtain more advantages for themselves. The distribution of political offices is done with the stability of the social system in view. They are distributed in the manner that assures the highest level of stability that can be acquired. Ideally, we can imagine a community in which the distribution obtains a perfect amount of stability. Such distribution, nonetheless, is not realistic. As we argued before, the sanctioning mechanism is an inherently inefficient system, and such inefficiency creates possibilities for uncooperative behavior. That is also true in the case of political power. The inefficiency of the mechanism opens up opportunities for corruption and rent-seeking. Such opportunities work as a perverse incentive to individuals. Although the cooperative framework condemns corruption and rent-seeking, they are, in the crude perspective of utility-maximization, methods to improve one's overall standing regarding utility. For this reason, a social struggle for political offices starts. Individuals are driven to take part in the struggle either by a desire to profit by exploiting the sanctioning mechanism and therefore the community, of rent-seeking and corruption or to avoid being a victim of such exploitation. The logic underneath the struggle for power is similar to the one that governs the Hobbesian state of nature. It is a form of the anticipation logic. Even individuals that do not desire to profit from the abuse of political power are driven to seek political power to avoid suffering from being exploited by others that acquire political powers.

It is in the struggle for political power that we find a strong connection between economic power and political power. The social struggle for political offices involves a series of factors, but a

283 Fukuyama also indicates that the presence of a balance of power in the political realm is importance for the stability of the social system. See Fukuyama (2011), pp. 325, 374, 377, 422, 424, 431.
prominent factor among them is economic power. As a power whose scope involves the cooptation of individuals or groups to serve one's interest, economic power is a fundamental aspect of the struggle for political power. We follow Hume in the belief that political power tends to shift to where economic power is. Economic power gives its holder a considerable advantage in the social struggle for political power over those with less economic power. Economic power produces political power for its holder in two different manners. The first is rent-seeking. In this sense, the individual with economic power uses this power to enlist an individual who holds a political office to abuse his political power for the benefit of the one that enlisted him. In the second sense, economic power is converted into political power. In this case, the individual with economic power uses it in the social struggle to obtain a political office for himself. The exact manner in which economic power can be converted into political power depends on the empirical reality of the society in question. In this aspect, both the purchase of title of nobility in an aristocratic political system and the use of economic power to finance one's political campaign to the office of senator in an electoral democracy can be used as examples of conversion of economic power into political power.

The struggle for political power can shift the distribution in two different directions. It can shift it to a distribution that increases the level of stability of the social system or to one that decreases it. The typical case of a shift that fragilizes the social system is one in which an already powerful group manages to increase its power. On the other side, an example of a shift that reinforces the stability of the social system is one in which a weak group acquires more power to rival the more powerful groups in society. Let us move not to see how such shifts affect the stability of a social system governed by JNL.

2.3.3 Political Power and Stability

We have now a clear picture of what political power is about and how its distribution in the social system works. We can, then, examine how political power and its distribution affect the stability of a social system informed by JNL. Such examination involves considering how rebels appear in the...
social system in the context of the distribution of political power and how a social system under JNL is capable of coping with the threat to stability that they pose.285

Political power creates defectors in potential. It can create both cheaters and rebels. However, in both cases, the rise of defectors happens differently than in the context of economic power. For political power has an important trait that differs it from economic power. Political power is a type of power that is intrinsically connected with the sanctioning mechanism, the state in particular. It is a form of power that exists solely inside a social framework. Political power is not conceivable in the state of nature since there are no states in such condition. Possession and exercise of political power depend on the existence of a cooperative framework. For this reason, it is reasonable to believe that holders of political offices, i.e., individuals with political power, cannot intent any regress to the state of nature. Different from economic power, political power is not transferable to a state of nature condition.286 If social cooperation collapses, so does the sanctioning mechanism and with it the political power that was associated with it. Holders of political power cannot be rebels in the sense we have been discussing so far, consequently. Rebels, as we argued, are individuals that regard the social system as unjustified and prefer a return to the situation of natural competition. Political power cannot give individuals incentive to do so because there is no form of political power in the state of nature. Political power can, nonetheless, destabilize social cooperation. We argued before that rebels appear in the social system in three different circumstances. Either the rebel has ascendancy over the social system, or the sanctioning mechanism has lost ascendancy over the rebel, or the mechanism maintains its ascendancy, but the framework is still regarded as unjustified. In the context of political power, we can conceive of cases similar to all three, but not identical. In the first two cases, we can refer to the figure of a

285 It is important to keep in mind that the political order is not determined by the moral agreement entirely. The actual political system, norms, and organizations that compose the political realm of a society are determined much more by historical and social contingencies than by moral requirements. The sole moral requirement that conditions the political world is that the state must be able to maintain the stability of moral order in some measure. In this sense, what the moral agreement demands is the creation of a state and the distribution of political offices in a pattern in which political power is balanced. For this reason, changes in the political world do not constitute changes in terms of moral cooperation necessarily. Changes in the political world may cause changes in the cooperative framework, but they are not, by themselves, changes in the framework. A society that changes from a military dictatorship to an electoral democracy does not undergo a change in the moral framework that governs it. However, a society that changes from an absolute monarchy to a dictatorial theocracy may undergo changes in the framework if the regime attempts to remove specific moral rules and introduce new ones. Our discussion here concerns this second type of change in the political world, i.e., those that have effects over the moral order.

286 We could argue that political power is transferable to the state of nature in some sense. As we argued before, coalitions in the state of nature work as primitive forms of cooperative frameworks, i.e., the limited constitutional contracts that Buchanan writes about. In order to work properly, the coalitions must possess some sort of political structure to ensure the stability of its terms of cooperation. Nonetheless, the political power that exists in the state of nature is extremely limited and it does not confer any substantive advantages for their holders. It is a power applicable only inside the coalition and a power that should be exercised for specified purposes. Due to the lack of advantages that political power confers in the state of nature, it is reasonable to assume that it cannot be transferred to the state of nature.
tyrant whereas in the third we can refer to repatrimonialization. Let us start with the third case.

Previously we referred to the third case as a scenario in which the individual regards the social system as unjustified, but the sanctioning mechanism retains ascendancy over him. In this case, the individual becomes a rebel and works to destabilize the mechanism, to gradually reduce its power over him. In the context of political power, such a case is not possible. A holder of a political office cannot become a rebel through political power. The reason for that is that a holder of a political office cannot rationally act to destabilize the social system in the manner that the rebel does. The rebel destabilizes the social system to force a return to the state of nature. Political power cannot lure an individual to this goal because if the goal is attained, the individual loses the political power he has. However, there is a similar case that is possible in the context of political power, namely the performance of repatrimonialization of the state.

Repatrimonialization of the state involves the practices of corruption and rent-seeking, as we argued before. Such practices are forms of uncooperative behavior. They amount to an abuse of political offices for one's own gains or the gains of another individual. Corrupt holders of political officers and rent-seekers cannot be rebels because the benefits they acquire with their political power depend on the existence and function of the sanctioning mechanism. What the political power lures them to become are not rebels, but cheaters. Recall that cheaters are individuals who defect from cooperation because they expect to obtain some benefit that they cannot obtain if they observe the rules of the cooperative framework. The defection of the cheater is not connected with whether the moral order is justified or not. What motivates his defection is the opportunity of obtaining a specific benefit by violating the cooperative framework. For the cheater, the cooperative framework is justified and profitable. The defection of the cheater is conditioned to the stability and maintenance of the social system. The defection that he is lured to perform is rational for him only insofar as it does not cause a significant effect on the stability of the social system. The holder of a political office has a cheater's attitude toward abusing his political power. In this respect, the practices that repatrimonialize the state are forms of cheating. Corrupt holders of political offices and rent-seekers can only pursue and attain their uncooperative goals inside a cooperative framework sustained by a functional sanctioning mechanism. It is only through the framework and the mechanism that they can obtain the advantage they seek. Both forms of repatrimonialization require the presence of a sanctioning mechanism that has ascendancy over the community in general. Holders of political offices can be considered to be a force that attempts to maintain the stability of the status quo when we consider them to be cheaters. Nonetheless, corrupt holders of political offices and rent-seekers can destabilize the social system in two manners.

The first manner refers to the individuals that are made worse off by corruption and rent-
seeking. As we argued before, both these activities make at least another individual in the social system worse off. The reason for that is that corruption and rent-seeking involve the instrumentalization of political offices for purposes that are not those for which the office was designed and maintained. Since the office is maintained and financed by resources from the collectivity, the instrumentalization of the office for personal gains amounts to use of collective resources for purposes other than those for which resources were invested. Those that performed the investment, then, are made worse off. Destabilization may arise from such individuals. As victims of corruption or rent-seeking, it is reasonable to expect that the attitude of such individuals toward the cooperative framework and the sanctioning mechanism deteriorates when they see the mechanism failing to redress such uncooperative activities. If their perception is that the holders of political offices are acquiring improper benefits for themselves that they would not be able to maintain in a situation of natural competition, such exploited individuals might become rebels. In such a case, they start to work to destabilize the system. A historical case of rent-seeking leading to destabilization is noticed by Fukuyama. He notices the case of pre-revolutionary France, where a rent-seeking elite managed to acquired tax privileges that ended up making the peasantry of the country strongly worse off.\footnote{Fukuyama (2011), p. 342. “The French fiscal system that developed by the late seventeenth century was highly regressive, taxing the poor in order to support the rich and powerful. Virtually every elite group, from high aristocrats to guild members to bourgeois towns, had succeeded in securing for itself a tax exemption, leaving the greatest burden to fall on the peasantry. This naturally provoked a long series of peasant uprisings and revolts. Tax increases to support the wars of Louis XIV were met with revolts in 1661, 1662, 1663, 1664, 1665, 1670, 1673, and 1675, the last being the large and serious uprising of the Bonnets Rouges. All were violently suppressed; for example, in the tax revolt of 1662, government troops took 584 rebels captive. Those over seventy and under twenty years of age were pardoned; the rest were condemned to the galleys.”} The result was a series of peasant uprisings against the state. That is the first manner in which acts to repatrimonialize the state destabilize the social system.

The second manner does not destabilize the social system directly, but it leads to the two scenarios in which an individual might become a rebel-like, or tyrant as we argued ahead, individual. That is the case when individuals use their political power to acquire more political power. It does not matter here whether political power is exercised in observance of the cooperative framework or through practices of corruption and rent-seeking. What is important is that political power is being used to acquire more political power. Recall that social struggles for power govern the shifts in the distribution of political power. In such struggle, possession of some level of political power is itself a factor that contributes to one's success in the struggle. As an individual accumulates political power, his power to perform corruption, rent-seeking, and avoid sanctions increases. Nonetheless, that does not change his attitude toward the social system, i.e., it does not change his behavior from that of a cheater to that of a rebel. Such alteration happens when a threshold of political power is reached.

287 Fukuyama (2011), p. 342. “The French fiscal system that developed by the late seventeenth century was highly regressive, taxing the poor in order to support the rich and powerful. Virtually every elite group, from high aristocrats to guild members to bourgeois towns, had succeeded in securing for itself a tax exemption, leaving the greatest burden to fall on the peasantry. This naturally provoked a long series of peasant uprisings and revolts. Tax increases to support the wars of Louis XIV were met with revolts in 1661, 1662, 1663, 1664, 1665, 1670, 1673, and 1675, the last being the large and serious uprising of the Bonnets Rouges. All were violently suppressed; for example, in the tax revolt of 1662, government troops took 584 rebels captive. Those over seventy and under twenty years of age were pardoned; the rest were condemned to the galleys.”

202
The threshold that changes the attitude of the individual toward the social system is that in which his political power reaches such a level that the sanctioning mechanism loses its ascendancy over him. An individual that has a high amount of political power is an individual able to use a large share of the state's power. If the share is large enough, it may come to happen that the individual has such tremendous control over the state's power that he can use it as if it were part of his own primary factors of power. In such circumstance, the mechanism's capacity of sanctioning such individual for abuse of his political office is small or nonexistent. His political power grows so large that there are no more checks or balances to keep him compliant with the cooperative framework. In such circumstance, the individual is not only able to perform acts of corruption or rent-seeking with few hindrances, but also of turning the sanctioning mechanism against the framework whose stability it is designed to assure. Such an attitude is similar to that of the rebel. Nonetheless, it is wrong to consider such an individual a rebel. What the individual with a high amount of political power desires is to change the cooperative framework in his favor and to force other individuals through the sanctioning mechanism to remain compliant with the new framework. His objective is not to resist or confront the sanctioning mechanism, but to use it against those dissatisfied with the alteration he performed in the framework. The proper nomenclature for such an individual is not rebel, but tyrant instead. He is an individual with so much political power that he takes control of the sanctioning mechanism and turns it against the cooperative framework to alter it in his favor and use it against the rest of the community to assure that the new framework enjoys some level of stability.288

The instability that tyrants create should be easy to notice. Tyrants have substantive renegotiation expectations and also a considerable power to impose them. They alter the cooperative framework to favor themselves. In other words, they change the terms of cooperation utilizing their political power. Such change occurs without the agreement of other members of the community. Although it is conceivable that the change will make some individuals better off, the likelihood is that a considerable number of individuals will be made worse off. The individuals made worse off are likely to become rebels, for the alterations carried by the tyrant were not only performed without their approval, but they also made them worse off.

The central difference between the holder of a political office that performs acts of

288 Shifts that create tyrants can happen in two ways. Either they create an individual whose political power is so vast that allows him to control the mechanism and to maintain the mechanism's ascendancy over the social system or the shift produces an individual that is powerful enough to control a share of the mechanism whose power is large enough to rival that of the community. Historically, the first case is observable in Napoleon Bonaparte's ascension as Emperor that ended the first French republic and the Meiji Restoration in Japan that ended the Tokugawa Shogunate and consolidated the power of the Japanese emperor Meiji over the Japanese political system. Historical examples of the second case can be observed in the Final War of the Roman Republic between Octavian and Mark Anthony and also in the American Civil War.
repatrimonialization and the tyrant lies in how each wants to improve their standing in society. The corrupt official and the rent-seeker seek to redistribute the cooperative surplus in their favor. They do not want to alter the cooperative framework itself, but only to obtain a favorable redistribution of the cooperative surplus employing their political power. The tyrant, on the other hand, seeks to change the cooperative framework itself. He aims at new terms of cooperation that are more favorable for him. Now that we have examined the effects of shifts in the distribution of political power over the moral order, we can analyze the relationship between political power and the stability of a social system under JNL.

The distribution of political offices in a society is a matter of historical and social contingency. The provision that the cooperative framework contains in this respect is that the distribution of political offices should be such that the stability of the social system is maximized. We may imagine an ideal distribution for each type of society, a distribution that would assure the maximum amount of stability for that society given the empirical and social circumstances in which it finds itself. Such ideal distribution offers us a standard to evaluate shifts in the distribution of political power in the society. Whenever a shift approaches the current distribution to the ideal distribution, such shift increases the level of stability in the society, such shift has, then, stabilizing effects. On the other side, whenever a shift distances the present distribution from the ideal one, it destabilizes the social system. The destabilizing shifts are those that facilitate the performance of corruption and rent-seeking and also augment the chances of a tyrant appearing in the system. Destabilization caused by shifts in political power happens either because of the appearance of a tyrant or rebels in response to acts of repatrimonialization done by holders of political offices. In the first case, an individual acquires control of a large share of the state's power. In the second case, holders of political offices, utilizing corruption and rent-seeking, make individuals worse off and such individuals become rebels. The capacity that a moral order has to cope with such shifts in the distribution of political power can be examined by how able such moral order is of either hindering the occurrence of destabilizing shifts or nullifying the adverse effects such shifts have over the stability. If the social system is capable of performing one of these two tasks, then it can be considered stable concerning the destabilizing oscillations in the distribution of political power. The focus of our analysis is in the capacity of a social system under JNL to cope with shifts that decrease its stability.

A moral order has only one option to cope with the destabilizing effect that the emergence of a tyrant has over the social system. It must prevent the emergence of the tyrant. For once the tyrant emerges, the very mechanism that is in place to assure the compliance of individuals to the cooperative framework comes under the control of the tyrant. That means that there is no possibility
of coercing the tyrant into compliance with the framework once he appears, since the social system's mechanism for such end is under the tyrant's control.

The social system must hinder individuals or groups from accumulating political power in order to prevent the emergence of tyrants. A tyrant is an individual that has an amount of political power that allows him almost full control over the power of the state. He acquires such power through the social struggle for political offices that takes place in the community. The tyrant is, one can argue, the ultimate 'winner' of the struggle. To prevent his appearance, then, the social system must ensure that no individual is in place to 'win' the struggle. For this end, the social system must impose some limitation in the power that individuals bring to bear in the struggle. A 'victory' that creates a tyrant is the manifestation of the superiority regarding the power of an individual over the others. The types of power involved in the struggle are, as we already stressed, diverse and many. However, as we argued before, in the social context two forms of power are fundamental and dominant over the others, political power and economic power. Both these two types of power play essential roles in the struggle for political offices. Due to their importance, it is reasonable to argue, then, that the social system's capability of hindering the appearance of tyrants depends on its capacity of controlling the amount of political power and economic power that individuals employ in the social struggle for political power. If the social system is successful in such an enterprise, then it assures that there is never an individual involved in the struggle that is powerful enough to 'win' it.

The first option we consider is the restriction of political power. Political power is restricted by limiting the individuals' access to political offices. This option is, however, unattainable. Political offices are already limited in some form. They give their holder some power over the sanctioning mechanism to fulfill certain specific purposes, such as, for example, administering sanctions to defectors, collecting resources to maintain the system, or allocating the resources in the system. If an individual uses his office to obtain more political power, he does so against the purpose of the office itself. In other words, the use of a political office to increase one's political power is a form of corruption. It is already a form of defection from cooperation, i.e., an action that violates the cooperative framework. If an individual managed to increase his amount of political power utilizing political office, he did so avoiding the sanctioning mechanism. To use a political office to increase one's political power is already something prohibited by the moral order. If an individual manages to do such a thing, the social system has already failed to prevent him from doing so, and thus there is nothing else it can do.

The second option available to contain the emergence of a tyrant is to restrict individuals' economic power. The idea here is to prevent individuals from succeeding in the social struggle for
power through discrepancies in economic power. Economic power is, as we argued before, the power to form coalitions. It is the power of enlisting other individuals to pursue the interests of the holder of economic power. In the struggle for political power, economic power acts by mobilizing other individuals taking part in the struggle to either help or favor the economically powerful individual. Again, the higher is one's economic power, then the higher is the number of persons he can mobilize to help himself. The higher the number of persons he has on his side, the higher are his chances of succeeding in the struggle for power. The moral order must either limit the economic power individuals can have or the amount of such power that they can employ in the struggle in order to prevent the emergence of a tyrant.

Formally, limiting the use of economic power in the social struggle for political power can solve the risk of emerging tyrants. With a limitation in the employment of economic power in place, economically powerful individuals would not be able to form the coalitions that would give them the upper hand in the struggle to obtain more political power. It is hard to see, however, how such a proposal can be materially implemented. The social system must prohibit the individual from interacting with the others with the purpose of 'recruiting' them employing his economic power in order to forbid the use of economic power in the struggle. If interactions are hindered, the economically powerful individual cannot perform a transfer to the others to gain their support. The problem, however, is what such prohibition implies materially. The prohibition is not that individual A is not allowed to support individual B in the struggle. Instead, it is that individual B is not allowed to support A in exchange for some of A's wealth. Coalitions in the struggle for power are not something whose occurrence one attempts to hinder here, but only those resulting from economic power. The difficulty lies in the fact that, in the end, the only measure of assuring that A's economic power will not be used to cooptate B in a coalition is to prevent every form of economic interaction between A and B from occurring. Unless we have a system that can pierce into B's mind and verify whether he took part in an economic transaction with A to exchange his support to A in the struggle for some of A's wealth, unless such farfetched possibility is available, there is no manner to hinder A from exercising political power.

The alternative is to limit individuals' economic power. Here, the objective is to set some cap to the amount of economic power that individuals are allowed to possess. To acquire wealth beyond the threshold established becomes an infraction, a violation of the cooperative framework. Such a solution would solve the problem posed by the tyrant. With such a threshold instituted, individuals never have the amount of economic power required to 'win' the social struggle and become tyrants. The restriction 'locks' individuals in the struggle for political power and assures that no individual ever becomes able to 'win'. The limitation does not forbid the existence of economic inequalities,
but it forbids that the inequalities augment to extents that allow individuals to become tyrants. The struggle for power does not cease with the imposition of a limitation in economic power. Since shifts in economic power continue to happen through individuals interacting with one another, individuals' positions in the struggle tend to change as such shifts happen.

Although a limitation in individuals' economic power does solve the threat of the tyrant, a social system under JNL cannot endorse such a solution. The reason for that is the same that we considered during the analysis of economic power and JNL. JNL forbids interference with distributional patterns of wealth if they are just. To set a cap on the amount of wealth that an individual may have is to set a standard of entitlement different from the one that informs JNL. Such new standard would be such that beyond a certain amount of wealth, individuals are not entitled to any wealth anymore, regardless of whether they came to possess such wealth under the rules of property or not. Such rule conflicts with JNL undoubtedly and therefore could not be introduced into the social system.

In the case of repatrimonialization, the threat to stability appears from individuals that are made worse off by holders of political offices abusing their political power. In this case, the threat can be dealt with either by assuring the compliance of the rebels or by eliminating the opportunities for corruption and rent-seeking. We believe that both cases cannot be attained. The case of the rebel should be clear. The only solution one could appeal to contain the rebel is to bring the sanctioning mechanism into the scene. Such a solution, we have already argued, is insufficient. The mechanism is indeed capable of assuring some level of compliance with the framework from the rebel, but it does not change his attitude toward the whole framework, which the rebel deems unjustified. The elimination of opportunities for corruption and rent-seeking is also farfetched. Through constitutional design, one can elaborate a political system in which each holder of a political office is put at check by another political power and thus prevented from performing corruption or rent-seeking. Nonetheless, even if such design is implemented in the social system successfully, it is bound to be changed at some point because of a shift in the distribution of political power. The reason for that lies in the connection between economic power and political power, and such connection displays the inability of JNL also to cope with destabilization risks posed by shifts in the distribution of political power.

JNL is a conception of distributive justice that concerns the distribution of external goods. It is, in essence, a principle that concerns the distribution of economic power, and economic power is, as we argued, closely connected with political power. Economic power is one of the central factors involved in the social struggle for political power. Moreover, as Hume argues, political power tends to follow economic power where it goes. Because of this connection, we can infer the effects that
the normative requirements of JNL have over the distribution of political power. It is important to notice, nonetheless, that others factors beyond economic power are involved in the social struggle for power. In this respect, although economic power may be a major, perhaps the critical factor in the struggle, it is relevant to notice that there is not a perfect correspondence between economic power and political power with regards to shifts. Although shifts in economic power create enormous pressure for shifts in political power, such shifts may not happen or may take some considerable time to happen.

Our previous analysis of a social system under JNL and shifts in the distribution of economic power is useful here. We saw that JNL contains no normative requirements regarding how goods should be distributed, but only with regards to how such distributions must emerge. If distributions emerge following rules of private property, they are just distributions with which one should not interfere. Accumulation of wealth, or economic power, does not offend JNL in any manner. Under JNL, individuals are allowed to accumulate as much economic power as they desire, even at the cost of the social system's stability. Such an attitude toward the accumulation of economic power spills over to the field of political power. In a social system under JNL, individuals are allowed to accumulate as much economic power as they desire. There are no normative restraints imposed by JNL in this respect. Without restraints, individuals are then allowed to bring as much economic power as they desire into the social struggle for power. Such possibility offers a risk for the stability of the social system. The more power an individual has in the social struggle for political power, the higher is the likelihood that he succeeds in augmenting his share of political power. In the context of a social system, economic power is one of the critical factors that determine the individuals' capacity in the social struggle. If an individual manages to acquire a considerable amount of economic power, he will bring such power to bear on the social struggle for power.

Individuals are drawn to take part in the social struggle for political power regardless of the amount of power they possess. They are drawn into the struggle through the pernicious anticipation logic that underlies it. Even if they do not desire to exploit the others through corruption and rent-seeking, they are lead to take part in the struggle, since the most efficacious manner of assuring that one will not be a victim of attempts to repatrimonialize the state is to acquire political power for oneself. An individual that has a high amount of economic power is, then, not only an individual with a reason to take part in the struggle. He is also a person with the means to succeed in the struggle. It would be irrational for him not to employ his economic power to acquire a share of political power or to enlarge the share he already owns.

Two factors indicate that destabilizing shifts in the distribution of political power cannot be avoided. The first is that individuals always have reasons to pursue more political power. The
second factor is that the distribution of economic power is always changing. Economic power is one of the fundamental factors in the struggle for political power. The struggle for political power is a zero-sum game. Whereas the struggle remains in a stalemate, no shifts occur. However, the stalemate situation cannot endure, since economic power is one of the most significant factors in the struggle and economic power is subject to constant change. For this reason, individuals' positions in the struggle are also subject to changes, as their economic powers change. Shifts in the distributional baseline of political power will reflect those of economic power at some point when some individuals involved in the struggle acquire an amount of economic power that they did not have before.

A social system under JNL is vulnerable to destabilization due to shifts in the distribution of political power. Such vulnerability lies in the connection between political power and economic power. A social system governed by JNL is not able to contain in any manner the emergence of considerable discrepancies in economic power. Such discrepancies, in their turn, are likely to produce also discrepancies in the political realm. Since economic power is vital in the struggle for political offices, it is reasonable to think that large discrepancies in economic power lead to significant discrepancies in political power. JNL's vulnerability to destabilization in virtue of shifts in the distribution of political power is a consequence of its vulnerability to shifts in the distribution of economic power. JNL is, we can conclude, an unstable conception of justice with regards to its capacity of maintaining a social balance of power.

3 POVERTY, STABILITY, AND JUSTICE

We move now to the examination of the third circumstance of moral cooperation in the context of JNL and stability. The third circumstance states that there must not be widespread poverty in the social system in order for cooperation to be justified for everyone. We examine here two objects. First, the effects that alterations in such circumstance have over the stability a social system informed by JNL. Second, the capacity that a social system informed by JNL has of coping with such effects. Our objective is to investigate whether JNL can be deemed stable in light of changes in the third circumstance.
3.1 Natural Scarcity and Artificial Scarcity

Poverty, we argued in chapter three, is the lack of capacity to command goods and resources necessary for the fulfillment of basic needs. This inability may have two different causes. Each cause refers to one of the forms scarcity that we discussed before in chapter four, namely natural and artificial scarcity. Natural scarcity refers to a case where the overall amount of goods and resources in the system is reduced drastically, and a considerable number of individuals fall into poverty. In the occurrence of natural scarcity, the cause of poverty is the unavailability of the goods the poor require. The unavailability of the goods results from the goods becoming so scarce that the few who have it are not willing to exchange it. Such a situation is brought about by, for example, a natural disaster. Artificial scarcity, in its turn, represents a case in which a considerable number of individuals fall into poverty although there are enough goods and resources in the community that they could use to escape poverty. Despite that, individuals are unable to command such goods for themselves. In the case of artificial scarcity, the inability of commanding goods is due to a lack of economic power.

It depicts the case in which the individual's economic power is so low that he is unable to command in the market the most basic goods that he requires, even though they are available in the market. In both cases, the inability of acquiring the goods he needs has severe consequences for the individual. Without being able to command such goods, the individual is not able to live adequately in any sense, and he is also subject to face suffering and death by not fulfilling his basic needs. The incentive the individual has to do whatever it takes to acquire the goods is vast. Such incentive leads to uncooperative behavior.

In the occurrence of both natural and artificial scarcity in a social system, those that become poor have a huge incentive to become non-compliant. They become poor, and thus unable to command the goods they require. However, the goods are available in the system. What hinders them from acquiring the good is, first, the moral order itself, and, second, the power of the holder of the good, in case he is capable of defending the good against the poor individual. In the case in which the poor are more powerful than the holders of the good they desire, the moral order becomes an explicit hindrance to the promotion of the fundamental interest of the poor, the interest in being free from poverty. The moral order becomes unjustified for the poor person since its normative requirements make her worse off. The stability of the rules becomes entirely dependent on the sanctioning mechanism.

In the case of natural scarcity, the mechanism and the moral order collapse. When natural

289 Economic power considered here as a form of outcome power.
scarcity occurs, a significant majority of individuals become poor. They become poor because the system does not contain enough of the goods and resources they require to be free from poverty. For such expressive majority, the moral order becomes immediately unjustified, a hindrance to their desire to be free from poverty. In such circumstances, the majority would immediately remove any support or investment in the sanctioning mechanism to reinstall a state of nature between them and the minority that holds the remaining goods and resources required to be free from poverty. Although such minority would gladly prefer the maintenance of the moral order, since it protects their rights over the scarce goods, they are not able to maintain the rest of the community inside the moral order with them. The best alternative that those who hold possession over the goods have to defend their goods from the poor is to form a coalition against the majority. In any case, the state of nature is reinstalled in the community, and individuals will compete for the scarce goods until a new balance of power is reached. Natural scarcity is a threat to a moral order informed by JNL. However, it is so for any moral order. Natural scarcity presents a scenario in which moral cooperation between the poor and nonpoor is impossible. Insofar as it is assumed that individuals are vulnerable to poverty and are utility-maximizers, it is hard to see how one can assure the continuity and stability of the moral order in the event of natural scarcity.

3.2 POVERTY, SCARCITY AND STABILITY

The only provision that can be made against the destabilizing effects of natural scarcity is a collective insurance against disasters. The idea is that while the community is free from scarcity, it will accumulate and stock the goods needed for freedom from poverty. With such a stock, the community will be able to maintain its stability in the occurrence of natural scarcity. Of course that such measure works under the assumption that the scarcity is either temporary or that it can be overcome if the community maintains its stability for some time. If the state of natural scarcity is definitive, the deposit will be extinguished at some point and, the moral order will collapse. In any case, the insurance could serve as a measure to prevent the destabilizing effects of natural scarcity.

As it has been the conclusion so far, a moral order informed by JNL would not be able to realize such form of insurance. Such an insurance implies some form of obligation that individuals would have of transferring part of their goods and resources to the insurance. Such obligation runs against JNL. According to the principle of natural liberty, every individual is entitled to his goods if he acquired them in conformity with the rules of private property. The obligation created by the
insurance implies that individuals are not entitled to part of their goods. Instead, it is the insurance that becomes entitled to such goods. The introduction of the insurance brings a new standard of entitlement that is foreign and runs against the principle of natural liberty.

Artificial scarcity appears in a different context than natural scarcity. Artificial scarcity is an economic problem. It is not a problem of lack of goods and resources. When artificial scarcity occurs, there are enough resources and goods available in the market that, if properly distributed, would eliminate the scarcity from the system. Poor individuals appear in the system always by causes over which they have no control. Two causes are the most likely in this respect. The first is bad luck. It may be that the individual is targeted by an unpredictable event through which he loses all his economic power to command goods in the market to fulfill his basic needs. The second cause is connected with an error in utility-maximization. As individuals interact in the market, they do it with the purpose of maximizing their utility. However, what individuals truly have is an expectation of utility maximization. With the information they have, they interact in the market in manners that they deem most conducive to the maximization of their utility. Nonetheless, individuals are subject to what Hayek called epistemological ignorance. They do not fully comprehend nor know all the facts and events that surround them. For this reason, their expectations in the market may be proved to be wrong as time develops. It is from mistakes in evaluations during market interactions that individuals may fall into poverty. A false expectation regarding a chain of market interactions, for example. That is how poor individuals appear in the system.

The problem is that the individuals who fall into poverty, who suffer from the scarcity, are unable to command the goods they require in the market. They lack economic power to do so. In such circumstances, the only manner the individual has of acquiring the goods he needs is by violating the moral order and forcefully taking the goods from other individuals. Such a thing indicates that individual becomes rebels. For they believe that in a situation of natural competition they could take control of a larger share of goods and resources than that that they have inside the moral framework. As individuals interact with one another, exchange goods, and produce new goods, both the distributional baseline of goods changes and also the distribution of power changes. Individuals become less or more powerful and less or more wealthy than they were at the start of moral cooperation. These shifts indicate that in the case of a return to the state of nature, poor individuals could predate those that became less powerful but increased their overall share of goods and thus that the poor can improve his standing in the state of nature. The effects of the poor over the stability are highly dependent on their number. The higher their number, then the more powerful a potential coalition of the poor to destabilize the moral order becomes.

Poverty in the context of artificial scarcity is an economic matter. It is a matter of resource
allocation. The poor individual lacks the goods he needs and also the means to command such goods. The goods are, nonetheless, available in the social system. It is just that the poor lack the economic power to command such goods. A social system has three options to cope with the poor in such circumstances. First, it can prevent the emergence of a pattern that depicts artificial scarcity. Second, in case artificial scarcity appears, it can shift the distribution of goods away from such pattern, eliminating the scarcity. Third, it can annul the destabilizing effects that the poor have over the moral order.

To sanction the poor to guarantee the stability of society is, as we argued many times before, not an appropriate solution. The poor behave as rebels do. They regard the social system and its normative requirements as unjustified. It is because the order puts them under specific normative requirements that they are unable to acquire the goods they require. To confront their non-compliant attitude with the sanctioning mechanism has the typical problems of confronting rebels with such method. The poor's perception that the social system is unjustified does not change, and they will remain non-compliant and will violate the cooperative framework whenever they have a chance to do it. The risk they pose to the stability of the social system is not adequately addressed.

To prevent the emergence of artificial scarcity and to eliminate it are different approaches, but they involve very similar mechanisms. In the case of prevention, what the social system must guarantee is that the vast majority of individuals always remain able to command the goods they require for their fundamental needs. On the one side, that is a guarantee against failures and errors of utility maximization. On the other side, to eliminate artificial scarcity, once it has emerged, amounts to ensure that individuals who are currently unable to command the goods they need become able to do so. In the first case, the idea is to establish a threshold of economic power below which individuals, in general, cannot fall. In the second case, the idea is to restore an individual that finds himself below the threshold to the threshold. What both cases will necessarily involve is some sort of redistribution of wealth. In the case of elimination of scarcity, it is easy to see why. If an individual is below the threshold established, the only manner of lifting him to the threshold is to transfer wealth to him. On the other side, the case of prevention must involve the transference of wealth for one crucial reason. Although it is true that prevention will also involve some form of mandatory savings, we cannot reasonably assume that every individual in the social spectrum, or that the majority of them, has a level of wealth that is big enough to sustain their consumption of goods and the savings required by the social system. Some individuals would fall into poverty immediately if they were required to save a certain amount of their wealth. In virtue of such individuals, the prevention of artificial scarcity must also involve the transference of wealth. For individuals unable to fulfill the mandatory savings, the only manner to prevent their fall into
poverty is to transfer to them wealth so that they become able to save it. In both cases, the threat to the stability of the social system that poverty poses is eliminated.

The failure of JNL to face the problems of stability that arise from poverty is another display of how Hobbes' intuition regarding the harm that views of absolute propriety have over social order. In the context of a social system informed by JNL, the measures required to solve the threat that poverty offers to the stability of social cooperation face the problem that we have already considered. The measures require some transfers of goods. It gives individuals an entitlement to resources by their need for goods. That is a standard of entitlement foreign and opposed to the principle of natural liberty that informs JNL. In a social system under JNL, individuals are entitled to the goods they have if they acquired them in conformity with the rules of private property. Poverty and needs do not give rise to an entitlement to goods. Hence, individuals cannot be forced to transfer goods they are entitled to other individuals. Again, in virtue of JNL, the social system is not able to cope with changes in the third circumstance of moral cooperation. It is an unstable social system, for this reason.

4 Conclusion of the Chapter

We can conclude this chapter with the affirmation that a social system governed by JNL is an unstable social system. Such a social system is 'besieged' by threats against its stability offered by rebels and is unable to address such problems. Regarding economic power, political power, and poverty, a social system under JNL is unable to cope with the threats to its stability posed by shifts in the distribution of such powers or shifts that cause the appearance of poor individuals. JNL's inability is reducible to one single factor, its incapacity of interfering with distributions of wealth and thus of economic power. Such incapacity makes JNL unable to cope with alterations in both the social balance of power and the absence of widespread poverty. JNL fails to adequately address destabilizing shifts in the distribution of the two most significant forms of social power, political power and economic power because it cannot control shifts in the distribution of wealth. It is also for the same reason that it fails to contain shifts that create poor individuals who destabilize the system. Such conclusions warrant us to uphold Hobbes' view that doctrines that concede absolute propriety over one's goods are harmful to the social order.

Regarding our general goal of justifying principles of justice with more substantive redistribution requirements than JNL, this chapter fulfills an important role. We offer here a
negative argument for such principles. Our conclusion in the chapter is that JNL is unstable because it cannot endorse redistributive measures, and that such redistributive measures are a vital factor to ensure the stability of the moral order. What we demonstrate here is that the moral order must include a conception of justice that allows it to interfere with the distribution of wealth. Otherwise, it is entirely incapable of assuring its own stability. In the next chapter, we offer a more positive argument for substantive principles of distributive justice.
OVERVIEW OF THE CHAPTER

In this chapter, our argument reaches its conclusion with the suggestion of a revised form of JNL. In the last chapter, we suggested a negative argument for a conception of justice with more substantive redistributive demands. We verified that Hobbes' intuition is right. When the state cannot interfere with individuals' private property, such inability weakens the social order and can lead to its dissolution. A moral order informed by JNL cannot be considered stable since it is incapable of interfering with the distribution of wealth in the system. JNL forbids any type of interference with individuals' wealth when the wealth was acquired in conformity with the rules of private property. Consequently, individuals are left able to accumulate enough economic power and political power to threaten the stability of the whole moral order. In this chapter, we suggest a positive argument for more substantive principles of distributive justice. In the first part of the chapter, we consider why the instability of JNL is a reason for rational agents to reject it. In this first part, our task is to argue for why rational agents would consider the stability of a moral principle or moral order relevant. The rejection of JNL indicates that the moral order requires a conception of justice that can interfere in the distributional baseline of wealth. We move, then, to the second part of the chapter, where we suggest the idea of an amendment to the moral agreement. The idea here is to argue that JNL's instability can be overcome if we reform it in some respects. The reform does not amount to a complete abandonment of JNL, but instead to the introduction of new principles that constrain its prohibitions to interfere in the distributional baseline of wealth. We depict such reform as an amendment to the moral agreement. The last part of the chapter contains the proposal of two standards of entitlement that the amendment introduces into the moral agreement, the principle of
balance and the principle of charity.

1 REJECTING JUSTICE AS NATURAL LIBERTY

The fact that JNL is an unstable conception of justice offers rational agents a strong reason to reject it. We covered in the last two chapters what stability refers to and why JNL is unstable. JNL is unable to solve the problem to stability that rebels pose. A society governed by JNL cannot adopt any measures to prevent the emergence of rebels or to remove them from the system once they have appeared. The only means that JNL has at its disposition to confront the threat of rebellion is the sanctioning mechanism. The mechanism, however, is not a viable option to contain the destabilizing effects of rebels, as we argued in previous chapters. Different from the cheater, the source of the destabilization that rebels provoke is the fact that they regard the moral order as unjustified, and the mechanism is powerless to affect such view that rebels have. The mechanism can inculcate some level of compliance with the moral order in the rebel, that is correct, but that is not enough to ensure the rebel's compliance. The moral order remains unjustified for the rebel regardless of threats made by the mechanism, and, for this reason, the destabilizing effects over the moral order remain present.

JNL's instability is, we believe, a reason to reject it as a proper conception of justice. Persons in a state of nature deliberating on which conception of justice should frame their mutual relations would reject JNL. Although JNL can be justified for them, such conception of justice creates opportunities to change the circumstances of moral justification, and JNL cannot cope with such changes. It is for these two reasons that we believe that persons would reject JNL in a state of nature. However, to understand such a rejection adequately, we must consider the ties between moral cooperation, social cooperation, and stability, since it is in those ties that the rationale for rejecting JNL is to be found.

1.1 MORAL AND SOCIAL COOPERATION

The stability of moral cooperation refers to, as we argued, to the capacity of a moral principle or moral order of engendering a *modus vivendi*. It refers to how able the principle or order is of eliciting rational compliance from those under its normative requirements. However, our discussion left open so far the subject of why stability is something that rational agents must value or give
importance to. One could argue against the structure of our argument for a more substantive form of
distributive justice by affirming that although JNL is unstable, such characteristic is not critical for
evaluating which moral conceptions should govern social life. In order to dismiss such criticism, we
must show why stability is a critical factor in the selection of moral principles.

The contractarian outlook depicts morality as a system of normative constraints created for
one single reason, the production of benefits and advantages for those under such constraints. The
most fundamental aspect of moral normativity lies in the usefulness of the constraints that it sets, as
we considered in chapter three with the ideas of the circumstances of justice and our interpretation
of the circumstances of moral justification. Persons find it rational to take part in a moral
agreement, i.e., to institute a moral order and also to comply with normative requirements such
order creates insofar as doing both things is advantageous for them. However, although morality
and moral cooperation afford individuals a few benefits, the bulk of the advantages they give
individuals arise from other forms of social cooperation.

Moral cooperation is a form of social cooperation. Individuals cooperate in moral terms
when they coordinate their actions with respect to normative requirements of moral principles. In
this sense, when Mark refrains from killing Joe and Joe refrains from killing Mark, even though
they hate each other mutually, they engage in moral cooperation. Social cooperation, on the other
hand, is a much broader category than moral cooperation. Social cooperation refers to every form of
cooperative interaction between persons with the purpose of creating benefits for those involved in

---

290 As Hume argues, the 'use' and 'tendency' of justice, which is considered as a virtue in Hume's argument, is 'to
procure happiness and security, by preserving order in society.' See Hume, M 3.12, SBN, p. 186. Notice, however,
that there are other factors involved in particular contractarian outlooks. In Gauthier's theory, dispositions also play
an important role. In Hobbes' and Stemmer's views, sanctions appear along normative constraints as central factors
of the approach. Dispositions and sanctions are examples of elements beyond normative constraints that are part of
morality. Such elements, nonetheless, always appear in connection with the constraints. In the case of Gauthier's
dispositions and Hobbes' and Stemmer's sanctions, they are both factors introduced to ensure compliance with the
normative constraints. We discussed such connection in chapter one and chapter four.

291 One could discuss here how active or passive must such coordination be in order for it to count as moral
cooperation. When Mark approaches Joe to buy a good from him, we can interpret the presence of moral
cooperation and social cooperation in their exchange in two different manners. The first is that they are cooperating
socially and morally, at the same time. In this view, since Mark does not grab the good and run, or Joe shoots Mark
and takes his money, the social cooperation they perform – the trade – comes along moral cooperation – refrain
from stealing and killing. The second interpretation is that there is only social cooperation in the matter. Although a
framework of moral rules and expectations is necessary for Mark and Joe to trade, the actual trade they perform
does not involve any form of moral cooperation, since they do not even conceive of violating moral rules in their
cooperative activity. We find the second interpretation more attractive because the first seems to imply that people
are engaged in moral cooperation at every second of their lives when they are near other persons. According to the
first interpretation, whenever Mark and Joe perform any sort of cooperative activity, they are engaging in moral
cooperation at the same time because they could kill, steal, or injury one another but they do not do it. For the first
interpretation, whenever there is mere possibility of acting immorally, even if the agent does not even contemplate
such possibility, and the agent does not acts so, there is a case of moral cooperation. The second interpretation sees
moral cooperation much more as a matter of accepting common principles to guide action and of actively
respecting these principles. For the second interpretation, when agents engage in social cooperation, but do not even
consider violating moral rules during the cooperation, then there is no reason to speak about moral cooperation, but
only social cooperation.

218
the cooperative activity. In this regard, we can speak of social cooperation when, for example, Mark and Joe go together in a hunt after a wolf that has been attacking their farms. They team up to pursue a goal, the death of the wolf, that is desired by both of them. Social cooperation must not involve necessarily immediate benefits for every party. When Mark helps Joe to fix the roof of his barn, doing so provides Mark with no immediate benefits, but it provides him with the expectation that Joe will help him in the future when he needs to do some maintenance work in his own farm. With this broad view of social cooperation, it is true that moral cooperation could be considered a form of social cooperation. After all, Mark refrains from killing Joe for the same reason he helps Joe to fix his roof, the expectation of a benefit in the future. We can differentiate both forms cooperation, nonetheless, by the fact that moral cooperation enables social cooperation broadly considered. Social cooperation on non-moral matters cannot take place adequately as long as moral cooperation is not in place. Interestingly, although moral cooperation is a requirement for social cooperation, it is the possibility of social cooperation that enables moral cooperation, as we argued.

The reason why individuals hold social cooperation in high esteem is that social cooperation makes their lives better. We stressed this factor in chapter three with what we have called the circumstance of unsocial sociability. Persons are such that they can improve their own conditions a lot utilizing by cooperating. Concerning the benefits that cooperation affords, Joseph Heath stresses that social cooperation involves five ‘primary mechanisms’ through which benefits are created for cooperators. The mechanisms are economies of scale, gains from trade, risk pooling, self-binding, and information transmission. Economies of scale refer to the fact that including an additional person in the performance of some activities results in the creation of a massive amount of benefit in comparison with the situation in which just one person performs the activity. Gains from trade refer to the fact that persons have different preferences and tastes as well as different skills and talents. Because they have different tastes, they can exchange goods to give away a good that does not rank high in one's preferences, but does so in the rank of the other person, for a good that ranks high among one's preferences, but ranks not so high in the other person's rank. The fact that persons have different talents and skills indicates that they can acquire profits from some form of division of labor. The idea is that if Mark is a good farmer and Joe a good fisherman, they both gain more when Mark just farms and Joe just fishes, and afterward they exchange vegetables for fishes, rather than

292 Here, we use the word 'adequately' because social cooperation on non-moral matters can happen in the state of nature. People can cooperate in the state of nature as long as the structure of the cooperative activity is such that to cooperate is a dominant strategy for every individual involved in the activity. When such structure obtains, no individual has a reason to defect or refuse to cooperate. Nonetheless, it is hard to imagine such forms of cooperation happening in the state of nature because such structure of cooperation can hardly be realized materially in a state of nature.

each of them engaging in both activities. Risk pooling, in its turn, refers to advantages that insurances offer individuals. Through social cooperation, individuals can establish insurance structures to attenuate the effects of losses they expect to suffer in the future, but cannot be certain that will come to pass. If we assume individuals to be risk-averse, as they often are, insurance affords benefits for all its participants insofar as they know that in the occurrence of the event that they fear, the losses will be reduced. Self-binding refers to using cooperative arrangements to bind oneself toward something in the future. The paradigmatic case of self-binding is that of the alarm clock. We set the alarm clock to ring early in the morning because we have to go to work. At the moment we set the clock, we are resolved to go to work, and we know that in the morning, just after waking up, our resolve to work is much weaker than it is now. We know that in the morning it is tough for us to get up 'by ourselves'. Self-binding offers us the possibility of enlisting other individuals to help us to deal with such inconsistencies and variations in our preferences. The last mechanism that Heath suggests is information transmission. Social cooperation enables us to reduce and economize learning costs. To cooperate, we must immerse ourselves in informative networks, such as language, for example, which allow us to extend our knowledge and information of the world much beyond the immediate surrounds that we can perceive with our senses. Heath's discussion of the mechanisms involved in the production of benefits of social cooperation help us to stress the value that social cooperation has for individuals. It is from such value that we can understand why the stability of moral principles or the moral order is a decisive factor for rational agents.

1.2 The Value of Stability

Morality, we have argued, enables social cooperation. However, in order for morality to enable social cooperation, it is not only important that a moral order is in place, but also that individuals in society are compliant with the normative requirements of the principle or the order. If there are normative requirements in society but individuals often refuse to comply with them, it is hard to see how social cooperation can adequately ensue. Morality enables social cooperation because it provides individuals with a framework of expectations that indicates them that others are willing to cooperate. Inside a moral order, individuals expect that everyone will act in foreseeable manners.

294 James Buchanan even dismisses the topic of stability and argues that it is predictability that is important in the context of changes in the constitutional status quo, i.e., at the level of defining the terms of cooperation. "What is important is not stability but predictability; the constitutional status quo offers the basis upon which individuals may form expectations, about the course of events, expectations which are necessary for rational planning."
Morality provides individuals with this framework of expectations. Without such a framework, social cooperation is not possible. It is regarding expectations that stability is an important matter.

A framework of expectations that enables social cooperation is in place only when individuals have moral rules, and it is known that individuals conform their behavior with the rules. Recurrent compliance with moral rules is a constitutive element of the framework of expectations. Recurrent compliance is nothing else but a reliable indicator of stability. A principle is defined as stable if individuals find it rational to comply with it and if they comply with it through their actions. Then, if we observe recurrent compliance of individuals to a principle, we can safely argue that such a principle is stable. Stability, then, refers to maintaining an expectation framework which, in its turn, enables individuals to engage in social cooperation. An unstable moral order, on the other side, is that which is not capable of sustaining a consistent such a framework. Since the moral order is unstable, that means that it is known in the social scheme that persons tend to violate the normative requirements of the principle. Consequently, the constraints and thus the expectations that enable social cooperation become unable to sustain cooperation. In such a scenario, it becomes irrational for individuals to engage in social cooperation.

Stability of the moral order is something desired by rational agents insofar as it is a requirement for the enjoyment of social cooperation. Without stability, the moral order is unable to provide the framework of expectations that enables cooperative activities. Consequently, individuals cannot properly enjoy the benefits of social cooperation when the moral framework that allegedly enables such benefits is unstable. The best way to conceive the relationship between stability and social cooperation is to think of a spectrum. In one extreme, we have the state of nature. In this domain, there is no moral framework and consequently no relevant framework of expectations that enables social cooperation. Social cooperation is, in the state of nature, either nonexistent or rare. On the other extreme, we can conceive of a perfectly stable moral framework. Here, compliance with the moral framework is always rational for every individual in the system. The framework of expectations here is such that individuals always expect the others to be cooperative and to act under the framework. In a perfectly stable moral framework, individuals enjoy the benefits of social cooperation to their fullest extent. Between the two extremes, we can conceive moral frameworks

---

Buchanan (1975), p. 100.

295 We refer here exclusively to frameworks of expectations that enable social cooperation. Such frameworks are virtually non-existent in the state of nature. Despite that, we can think of frameworks of expectations in the state of nature if we abandon the connection of such frameworks with social cooperation. In the state of nature, individuals expect that the others will be hostile or non-cooperative toward themselves. It is due to this framework of expectations that they are driven into the anticipation logic that Hobbes refers to.

296 As we argued before, social cooperation is possible in the state of nature, but only under specific circumstances. And given the specificity of such circumstances, it is reasonable to assume that few or no opportunities for social cooperation will be present in the state of nature.
with different levels of stability. Different levels of stability determine how well individuals can enjoy the benefits of social cooperation. The more stable a moral order is, the more certain are individuals' expectations about how the others behave. The more certain are those expectations, then the abler are individuals to participate in social cooperation. Contrarily, a low level of stability indicates a framework of expectations that is not that solid. In this situation, individuals are not certain about other individuals' willingness to cooperate. Such uncertainty generates in individuals a skeptical attitude toward social cooperation. Since social cooperation is based on an expectation of reciprocity, uncertainty regarding the other's willingness to be reciprocal can lead to a refusal to cooperate. The stability of the moral order, we can conclude, indicates the amount of benefits that individuals may acquire through social cooperation. The more stable a moral order is, then the more significant is the share of benefits produced by social cooperation that individuals can attain. That is why stability must be considered an essential property of moral principles. In an unstable moral order, it is unclear whether individuals can enjoy the benefits of social cooperation adequately. Because there is no robust framework of expectations, it is likely that individuals will refrain from cooperation in virtue of fear that the other party might not reciprocate.

In chapter five, we stressed a particular problem of stability, namely its tendency to lead the moral order to its collapse. That happens when an individual becomes ascendant through economic power or by becoming a tyrant. In such cases, the moral order collapses and another cooperative framework is instituted in its place. Avoiding such collapse is one of the reasons that rational agents have to prefer stable moral orders to unstable ones. What we stress now is a second reason, namely the quality of social cooperation. Instability indicates not only a risk of 'revolution', but also a decrease in the profits and benefits that individuals can acquire from social cooperation. The more unstable a moral order is, the more individuals expect cooperative partners to double-cross them. The volume of advantages they expect to reap from social cooperation reduces insofar as the stability of the moral order reduces, either because they expect to be double-crossed by other people or because they start to refrain from cooperation out of fear of exploitation.

The fact that stability is essential for the enjoyment of social cooperation allows us to connect it with the idea of fundamental general interests. Stemmer argues that the rationale for several moral rules can be reduced to a specific fundamental general interest. The moral rule that forbids murder, for example, can be reduced to the fundamental general interest in not being killed. It is because individuals are mutually vulnerable to and mutually capable of killing each other, and because they strongly desire not to be killed, that they have a rationale to agree on a moral rule that forbids murder. Fundamental general interests are at the basis of the hypothetical deliberation over the moral agreement. Due to the importance that stability has for the enjoyment of the benefits of
social cooperation, both moral and non-moral, it seems to us reasonable to suppose that there is an interest in the stability of the moral order that follows directly from the others fundamental general interests. Such an interest is not fundamental itself, since it derivatives from the group of fundamental interests. However, it can be considered a general and critical interest, since it is directed to an object without which no fundamental general interest can be promoted adequately. Morality and non-moral forms of social cooperation can be considered primary goods for rational agents. An interest in morality is something that we have already explored in chapter one. The source of such an interest lies in the connection between stability and social cooperation. It is only in a stable moral order that individuals can collect the benefits of moral cooperation fully. In an unstable moral order, although they can promote in some sense their fundamental interests and their other interests, the promotion of such interests is marked by difficulties.

If we are correct in formulating a general interest in stable cooperative frameworks, then we can be sure that the stability of cooperative frameworks is a factor that individuals are concerned with when deliberating over the moral agreement. In other words, the stability of moral order or moral principle must play an essential role in the contractarian method. If we imagine individuals in a state of nature deliberating over a moral agreement, we must presuppose that they are concerned not only with the justification of the principles that compose the agreement but also to their stability. In the thought-experiment, individuals abandon the state of nature with the single purpose of acquiring the benefits of social and moral cooperation that are unavailable for them in the state of nature. In this respect, if they were to select a moral order that is not stable, it would be the case that they are not fulfilling their objective correctly. An unstable moral order indicates that, as we have argued, individuals will often fail to attain the benefits of cooperation due to the lack of a robust framework of expectations. To chose, then, a moral order that is not stable in the state of nature is irrational.

What JNL offers individuals are uncertain prospects regarding the enjoyment of the benefits created by cooperation. Its inability to handling the threat to stability that rebels offer indicates that as soon as substantive renegotiation expectations appear, the benefits of cooperation available will start to decrease. Agreeing on JNL from the state of nature might provide individuals benefits of cooperation immediately. However, as individuals interact and the distribution of power changes or individuals fall into poverty, society enters a downward slope regarding its stability that cannot be averted nor reverted by JNL. The uncertainty that revolves around the enjoyment of the benefits of cooperation we associate with JNL is, we believe, considerable evidence to affirm that JNL fails to promote the general interest in stability. With JNL, individuals cannot be sure that they will remain enjoying the benefits of cooperation throughout time. For this reason, we think that persons in a
state of nature would reject JNL as a suitable conception of justice for their society. The maintenance of stability requires a conception of justice more substantive than JNL with regards to redistributive normative demands.

2 AN AMENDMENT TO THE MORAL AGREEMENT

In light of the problem of stability, we want to suggest a new conception of justice. As we examined in previous chapters, the stability of social cooperation is intrinsically connected with the circumstances of moral justification. The presence of the circumstances is the crucial factor for the justifiability of the cooperative framework and its normative requirements. Such circumstances are, nonetheless, not permanent. They change with time due to interactions that individuals conduct with each other inside the cooperative framework. As the circumstances change, so does the attitude of some individuals toward the moral order. If a change is drastic enough, individuals may start to regard the framework as unjustified, and thus become rebels. In chapter five, we concluded that a moral order governed by JNL is entirely unable to respond to the challenge offered by rebels. As a conception of justice, then, JNL must be deemed unstable. What JNL's failure indicates is that the moral order requires a principle that allows interferences in the distribution of wealth, since such interferences are required to tackle the stability issue associated with the rebels. We want to suggest an approach to distributive justice different from JNL that is focused on solving the problems we diagnosed in the last chapter.

In our analysis in chapter two, we considered Aristotle's view of distributive justice and argued that the central problem of justice lies in determining the standard according to which goods should be divided. In this sense, to propose a new conception of justice involves proposing a new standard of entitlement. In the case of JNL, we argued that the principle of natural liberty is such standard. Such principle determines that an individual is entitled to a good if he was able to acquire it and the acquisition was made under the rules of private property. The principle of natural liberty is, as we argued before, extremely minimalistic. And in our discussion of stability, we discovered that such minimalism is highly problematic. It is due to the principle of natural liberty that JNL is unable to cope with the threat of destabilization offered by rebels, and is thus an unstable conception of justice. To overcome the difficulty posed by rebels, we introduce two standards of entitlement in this chapter, and thus a new conception of justice. However, despite the new factors that we introduce, we preserve some aspects of JNL and the principle of natural liberty. For this and
some other reasons, we formulate the conception of justice that we suggest here as an amendment to the moral agreement.

An amendment is a consensual alteration of a contract's clauses and provisions that is binding on the participants of the contract. It introduces changes while it maintains some aspects of the contract. The moral agreement that we have in mind here is a moral agreement that includes JNL, i.e., an unstable moral agreement. The amendment has the objective of solving the problem of stability. It aims at doing so by reforming the conception of justice that the contract contains, JNL. Our choice of referring to an amendment of the moral agreement instead of only suggesting a new moral agreement is not arbitrary. As we argued, we preserve some aspects of JNL in the revised moral agreement. The normative demands of JNL remain, in some sense, present in the amended version of the contract. In truth, the principle of natural liberty remains a critical standard of entitlement governing the idea of justice. What the amendment introduces are restrictions to the normative requirements of JNL and a few new normative requirements based on standards of entitlement different from the principle of natural liberty. Such changes are necessary to assure the stability of the moral order that springs from the moral agreement.

2.1 AMENDMENT AND PROVISO

Referring to the conception of justice that we suggest as an amendment has another relevant reason beyond indicating that some aspects of JNL and the principle of natural liberty are preserved. Formulating the conception as an amendment helps us to differentiate our view of distributive justice from that suggested by Gauthier in the form of the Lockean proviso. Both Gauthier's proviso and the amendment we suggest here are views of distributive justice grounded on the Hobbesian method of moral justification. However, the amendment has a different relationship with the moral agreement than that of the proviso. A proviso is a clause whose fulfillment is necessary for the validity of the agreement. When Locke speaks of his famous proviso, for example, he is referring to a condition that must be fulfilled in order for the acquisition of land to be valid. If land is acquired in disrespect to the proviso, no valid act of acquisition has happened. In this sense, the proviso to the moral agreement is a necessary condition for the validity of the contract. If the proviso is disrespected, the contract loses its validity. The amendment, on the other hand, has a different relationship with the agreement. It is not a necessary condition for the contract. The validity of the contract is not conditioned by the amendment, as it is in the case of the proviso. The amendment is an alteration of the contract that is binding for all parts of the contract. It does not condition in any
sense the validity of the contract. Instead, it is the validity of the amendment that is conditioned by the contract. To see that, we can consider how both the amendment and the proviso are connected with the rationale for social cooperation.

In Gauthier's argument, rational agents find it rational to take part in the moral agreement only if the proviso has been observed by everyone previously. If someone has disrespected the proviso, there are necessarily individuals who regard moral cooperation as an irrational enterprise. In this sense, the rationale of moral cooperation is conditioned to respect for the proviso. It is rational to cooperate only if the proviso is in place. On the other side, the amendment does not condition the rationale for moral cooperation in this manner. A proposal of moral cooperation that does not include our amendment is not straightforwardly irrational. The rationality of the contract is independent of the presence or absence of the amendment. The rationale for amending the moral agreement is but an extension of the rationale that drives rational agents toward moral cooperation. The purpose of the amendment is ensuring the stability of the moral agreement and moral cooperation. Its objective is but the maintenance of the contract that individuals find it rational to sign. Whereas the rationale for taking part in the contract is the enjoyment of the benefits and advantages that it affords, the rationale for the amendment is the maintenance of these benefits and advantages. The rationale for the amendment is grounded in the general interest in stability that we suggested above. Consider the example below for further clarification of the difference between the amendment and the proviso.

Consider two scenarios. In the first scenario, rational agents must choose either to remain in the state of nature or to sign a moral agreement that does not include the amendment. In the second scenario, they must choose between signing a moral agreement without the amendment and one that has been amended. In the first situation, to sign the moral agreement is the rational thing to do. Even if the moral order that ensues from the contract is an unstable one, it still affords individuals benefits and advantages that they could not acquire if they had remained in the state of nature. An unstable moral order is still more advantageous than the Hobbesian state of nature. In the second case, the rational option is, we believe, to amend the contract and refuse its unamended version. With the amended version, rational agents ensure for themselves not only benefits and advantages from social cooperation but also that these benefits and advantages perdure throughout time. In the

297 In chapter two, one of the arguments we raised against Gauthier's idea of the proviso refers to this point. Gauthier argues that taking part in the moral agreement without the proviso would be irrational for specific individuals. It is not clear, however, whether that is the case. Although a moral order founded from an initial bargaining position constrained by the proviso would be better for some than the natural distribution, it is reasonable to believe that a moral order based on the natural distribution is still better for everyone than the state of nature. Gauthier does not seem to take into account this factor. Cooperation that starts from some form of exploitation may be more profitable than permanence in the state of nature for those that were exploited. If that is true, Gauthier's argument for the proviso is undermined.
second case, refusing the amendment is the irrational option, since choosing it amounts to picking the option that affords less utility for oneself. The first scenario offers an excellent case to compare the amendment with the proviso. If the first case involved an agreement without Gauthier's proviso instead of our amendment, the rational option would be to remain in the state of nature and to refuse social cooperation. As argued, in the case of the proviso, the rationale of cooperation is conditioned to the presence of the proviso. Otherwise, cooperation is irrational.

Gauthier also refers to stability as one of the reasons that rational agents have to accept the proviso. He argues that a moral order whose initial positions are not determined by the proviso is inherently unstable because it is an unfair moral order. For Gauthier, an unfair moral order is a *modus vivendi* in the sense considered in Rawls' discussion of stability, which we considered in chapter four. It is a group of rules that are accepted as long as they are regarded as advantageous for those they constrain. As the rules cease to be advantageous, it becomes irrational to comply with them. Fairness, Gauthier believes, solves this problem. When the moral agreement is fair, i.e., when it takes into account the proviso, individuals remain committed to it, unless it becomes too costly for oneself to do so. Gauthier's appeal to stability is different from ours. The fair moral order is in Gauthier's argument something similar to what a social order founded on an *overlapping consensus* is in Rawls'. The fair moral order is such that individuals are committed with it even if it ceases to be advantageous for them. Our view refuses from the outset the possibility of such form of stability that is anchored on commitments that transcend utility-maximization. Stability always refers to a *modus vivendi*. The question of stability is the question of maintaining a *modus vivendi*, whereas in Gauthier's case it seems to be the question of transforming a *modus vivendi* into something more akin to a social order based on an overlapping consensus.

Another difference between Gauthier's treatment of the question of stability and ours is that a normative requirement of producing a stable moral order can be justified in our case, whereas it is unclear whether the same can be said of Gauthier's. Gauthier's view connects stability with fairness. For him, the stability of a moral order is the consequence of its fairness. But fairness in Gauthier's context is an obscure concept. It appears to be a normative concept that is not submitted to the contractarian methodology. In Gauthier's view, rational agents prefer fair outcomes to unfair outcomes, even if a fair outcome is inferior to an unfair one regarding the advantages one obtains.

---

299 We refer here to his appeal to stability done in the final part of chapter VII in *Morals by Agreement*. The discussion in this chapter is different from that we considered previously in chapter four, which referred to Gauthier's article *The Social Contract as Ideology*.
300 We discussed this matter thoroughly in chapter two. Gauthier refers to the proviso as the element to differentiate fair from unfair outcomes in the state of nature. However, the proviso clearly favors some individuals over others in terms of advantages. The adoption of the proviso makes a group of individuals, which we have called the 'strong' individuals, worse off. For them, the fair outcome is less advantageous than the unfair outcome. For this reason,
Stability seems to be a mere welcoming consequence of the agent's inclination toward fairness. For this reason, if we were to assume that agents have no commitment to fairness, Gauthier's arguments for any normative requirements originated from the search of stability would collapse.

Referring to the new conception of justice as an amendment to the moral agreement and not as a proviso free us from obscurities from which Gauthier's argument suffers. As we considered in chapter two, the idea of a proviso to the moral agreement, as Gauthier formulates it, is highly problematic. The proviso is suggested to protect 'weaker' individuals that, without it, would be coerced by 'strong' individuals to perform unproductive transfers. The purpose of the proviso is to nullify the effect of such transfers and to plain the field from which individuals negotiate, assuring that the negotiation is fair. Gauthier argues that if such proviso is not respected, the weak would refrain from negotiating and thus of taking part in the agreement. In Gauthier's view, the weak would regard a bargain for the moral agreement in which the strong held the fruits of unproductive transfer as unfair. Due to the unfairness, the weak would refuse to take part in the moral agreement. The refusal of the weak would, Gauthier believes, prompt the strong to accept the proviso, since otherwise they, the strong, would not be able to enjoy the benefits that social cooperation affords since the weak would trap them in the state of nature. Gauthier's argument is obscure in many aspects, as we considered in chapter two. It is not clear why the weak would refuse to cooperate. The idea of fairness that Gauthier brings into the discussion seems somewhat arbitrary and unjustified from the Hobbesian perspective. Lastly, it is unclear why the strong would bend to the weak's demand for the proviso, assuming that the weak would indeed refuse to cooperate if the proviso is not in place.

The central matter that the proviso's argument refers to is the moralization of the state of nature. The argument for the proviso is a demonstration that rational agents would accept and respect certain moral-like constraints on their actions in a state of nature. From the viewpoint of utility-maximization, it is hard to see how such a task can be accomplished without an appeal to altruism. An altruistic factor seems to be present in Gauthier's argument, represented by his appeal to fairness, but such factor lacks proper justification. The contractarian framework in which Gauthier's discussion is located does not allow the introduction of such altruistic factors to justify rules, for the altruistic factor itself requires some justification. An amendment, in its turn, is free of such problems. It does not require an argument for the introduction of constraints in the state of nature, as the proviso requires. Different from the proviso, the rationale for the amendment

Gauthier's view of fairness cannot be endorsed by the contractarian methodology. The 'strong' would never agree to an agreement on such a conception of fairness. However, if the strong are willing to accept the proviso, then it must be the case that they have some prior commitment to such a view of fairness, a commitment that would not require justification through the contractarian method.
'parasitizes' the rationale for the moral agreement itself. The discussion of whether rational agents should amend or not the moral agreement is posterior to the discussion of whether signing the contract is rational. And the result of the discussion does not affect the rationale to cooperate.

It is true, nonetheless, that the amendment is much more conservative in its objectives than the proviso. Both of them seek to argue for a rule of distributive justice. However, whereas the amendment defines the rule as a consequence of the willingness to cooperate, the proviso defines it as a precondition of such willingness. If successful, the proviso demonstrates that moral cooperation is only possible under the terms stipulated by the rule of justice it contains. Otherwise, the benefits of cooperation are out of reach for everyone. The strength of the proviso's argument lies in the fact that it conditionalizes the enjoyment of the benefits of moral cooperation to the acceptance and respect for the rule of justice it upholds. If it is successfully demonstrated that there is such relation between the proviso and the benefits of cooperation, the rationale for accepting the proviso becomes ironclad. The amendment, on the other side, offers a more conservative outlook. The argument for the amendment does not affirm that the benefits of moral cooperation can only be enjoyed if the amendment is accepted. Instead, the argument is that individuals can only guarantee that they will continue to enjoy the benefits as times goes by if they introduce the amendment.

3 THE AMENDMENT

We come now to the proposal of the amendment to the moral agreement. The proposal is divided into two parts. We saw in the last chapter that JNL's instability arises from its inability to deal with shifts in the balance of power and the emergence of poor individuals in the system. The amendment is introduced as a solution to these two problems. We divide the amendment into two parts. We consider first the part that must be introduced to handle the stability problems connected with the balance of power. In this part, our discussion concerns mainly wealth inequality. Secondly, we treat the issues to stability that appear from poverty. In this second step, we consider what changes the amendment must perform in the moral agreement in order to assure its stability in the face of natural and artificial scarcity. The introduction of the amendment involves the introduction of two new standards of entitlement and, consequently, in the alteration of JNL. The two new standards, the principle of balance and the principle of charity, constrain the principle of natural liberty. They both have priority over the principle of natural liberty in specific cases concerning the production of normative demands of individuals. The constraints represent the reformation of JNL into a different
conception of justice, which we refer to simply as *justice as constrained natural liberty*, or JCNL.

### 3.1 The Principle of Balance

What the idea of the social balance of power indicates is that power must be distributed in the social system in a balanced manner in order for the system to be stable. It indicates, first, that the system does not require that power be equally distributed in order to be stable. As long as the power of every unit is in check by that of another unit or group of units, stability ensues. Inequality in the distribution of power is, then, not a threat to stability *per se*. However, inequalities in power can become a threat to the stability when they are such that some individuals start to regard the moral order more as a hindrance to their interests rather than something that helps their promotion. Considerable gaps in power may lead the powerful to think that their ability to exact their interests over the weak is much higher than the capacity of the weak to resist or to confront them in return. That is when substantive renegotiation expectations appear and create rebels, which, in their turn, destabilize the system. Notice, however, that the emergence of the rebel is connected with an increase in the level of inequality in power. It is not, however, any level of inequality that creates rebels. They appear only when the inequality reaches or surpasses a specific magnitude. The preservation of stability, then, demands the maintenance of inequalities in power inside what we call from now on a 'safe margin'.

The safe margin refers to the maximum magnitude that a society's level of inequality can reach without suffering considerable problems of stability. It is not necessarily a margin that represents a scenario in which there are no rebels in the system, but, instead, it is a margin where there is no individual with renegotiation expectations substantive enough to cause problems of stability. Now, to empirically define the safe margin is a task that we cannot pursue here. Such a task is highly empirical. Each society has its own safe margin. The margin is defined by empirical aspects such as the number of individuals in society and the total availability of empirical factors of power such as wealth. Societies with a high number of individuals might be able to tolerate larger margins of inequalities in power between persons than societies with a smaller number of individuals, for example. In any case, determining with exactitude the safe margin or examining with precision the factors involved in its determination are tasks that our discussion here does not need to pursue. For our discussion, the observation that there is always such a margin is already enough.

In the last chapter, we argued that the dynamics of power inside a moral order are governed
primarily by the distribution of wealth. Shifts in the distribution of wealth strongly influence shifts in economic power and shifts in political power, which we considered to be the key types of power involved in the social balance of power. It is because of wealth accumulation that the balance of power can be overturned and individuals converted into rebels. If wealth is the central factor that governs the balance of power inside the moral order, then the inequalities that threaten the stability of the order are wealth inequalities, primarily.\[301\] To handle the occurrence of such form of inequalities is, we think, the path to solving the problem of stability that we stressed in our examination of JNL.

In order to tackle the matter of substantive wealth inequality, the principle of natural liberty must be changed. The principle of natural liberty is JNL's entitlement standard. With the principle of natural liberty, a given distribution of goods is just if every individual involved in it has acquired the goods he possesses in conformity with the rules of private property. A direct consequence of the principle of natural liberty is that interferences with the distribution of wealth are forbidden. If what entitles individuals to a good is that the good was acquired under the rules of private property, then any form of interference with a possessed good acquired under the rules of private property amounts to an injustice.\[302\] Such restriction is problematic because it entitles individuals to shares of wealth that destabilize society. If an individual holds an amount of wealth that gives him tremendous economic power and, consequently, substantive renegotiation expectations, there is nothing that can be done to contain him, who becomes a threat to the moral order. Individuals cannot have absolute ownership rights over their goods, as Hobbes argues.\[303\] Consequently, the principle of natural liberty cannot determine entitlement to wealth entirely if the moral order is to be stable. That was our conclusion in chapter five. Entitlement must also be defined following the requirements of the moral order to be stable. That is the central factor that the amendment introduces.

The moral order must allow for the possibility of interference with the distribution of wealth to preserve its stability. Whenever the level of inequality in the system reaches or surpasses what we have termed the 'safe margin' of inequality, there are rebels in the system who destabilize the moral order. Eliminating the rationale to rebellion requires from the moral order the power to interfere with the distribution of wealth. Such interference has the goal of restoring the inequality in the

\[301\] There are, naturally, other forms of inequalities in power that threaten the stability of the moral order. If we imagine two social groups with extremely unequal levels of military power, for example, their moral order will be unstable. But, as we argued in chapter five, economic and political power are the basic forms of social power. In this sense, discrepancies in military power can be reduced to either political or economic power in some sense. For this reason, we discuss the social balance of power from the perspective of these two types of power exclusively.

\[302\] Except when the interference is done with the consensus of the owner of the good.

social system to the safe margin. Keep in mind, nonetheless, that the critical factor for the
discussion is not the amount of wealth that a particular individual has. Instead, it is the gap in
economic power between individuals. It is the magnitude of inequality that is the primary factor that
destabilizes the system. For this reason, a shift in which an individual loses a considerable amount
of wealth can also be a destabilizing shift. In this case, the size of the gap in wealth might surpass
the safe margin of inequality even though the amount of wealth that the most powerful possess
remains unchanged. Although the powerful do not acquire more power in absolute terms in this
case, the gap that appears serves to inculcate in them a rationale toward rebellion.

What the amendment must introduce, then, is a principle to allow the moral order to cope
with the occurrence of shifts that create inequalities in wealth whose magnitude surpass the safe
margin. In order to cope with such shifts, there are two measures that the moral order can adopt. It
can either control directly individuals' capacity to use their own wealth or to interfere with
individuals' shares. We believe that the second option is the most suitable, and the one after which
the amendment should be shaped.

Controlling individuals' capacity to use their wealth directly does not seem a viable
alternative in our view due to some problems. A first and clear problem with such approach derives
from Friedrich Hayek's epistemological pessimism that we discussed in chapter three. As we argued
in chapter five, shifts in the distribution of economic power are not intentional, and due to the sheer
magnitude of information they involve, it is inconceivable to control or direct such shifts
effectively. To detect and prevent shifts that create rebels from occurring, one would require a social
planner capable of accessing and processing the vast information involved in the occurrence of
shifts in economic power. Society would require planners with almost omniscient capacities to
succeed in this task. For this reason, we think that skepticism regarding the real viability of such
option seems warranted. However, even if we assume that such an option is viable, it still faces
particular problems.

In order to control individuals' capacity to use their own wealth directly, the moral order
must be ready to prevent an already powerful individual from acquiring more wealth and also
preventing lesser powerful individuals from losing their wealth. As we argued before, both the
powerful becoming more affluent and the less powerful becoming poorer are situations that can
create gaps in power that surpass the safe margin.

In order to prevent the powerful from acquiring more wealth, the moral order must isolate
them from the market. One can only augment one's amount of wealth and thus of economic power
employing trade or productive activities. To prevent an increase in wealth, then, the moral order
must prohibit the powerful individual from engaging in both these activities. Such prohibition,
however, imposes on the powerful what we consider to be a restriction that is too draconian. Curtailing the powerful from trade and productive activities means curtailing them from a considerable share of the benefits of social cooperation. The restriction on trade is particularly damaging. If the powerful are not allowed to trade with the others, the range of possible interactions in which they can engage with other members of the social scheme is drastically reduced. We think that such restriction imposes a too drastic reduction of the advantages and benefits that the powerful can reap from social cooperation. Such a drastic reduction would likely lead the powerful to refuse any principle that has as a normative consequence the employment of such measure. Although they are interested in the stability of the moral order, they are also equally interested in enjoying the benefits that the cooperative framework affords. To curtail their chances of enjoying such benefits seems to imply a reduction in their willingness to cooperate which could be translated as a refusal to accept standards of entitlement that would impose such restrictions.

Regarding the less powerful members of society, adopting measures to hinder the less powerful from losing part of their wealth is likely to have a similar result for the non-wealthy than the restriction we examined for the powerful. To preserve the gap in power inside the safe margin in this context, one must employ measures to prevent the decrease of the wealth of the less powerful individuals in society. Such a measure implies hindering less powerful individuals from spending or using part of their wealth. Preventing this involves preventing the owner of the wealth from spending his own wealth. Such prevention can be understood as a form of mandatory or forced saving. Individuals would be required to save and to not spend part of their wealth in order to ensure that they always maintain a certain level of economic power. With such savings in place, the magnitude of wealth inequality in the society can be kept under control in the safe margin. The problem with this approach is that it also involves curtailing benefits of social cooperation, but this time for the less powerful. Depending on the amount of wealth that an individual has, the saving he will be required to make can encompass a considerable proportion of his available wealth. The less wealthy he is, the larger is the share of his own wealth that he will be forced to save, and thus the lower becomes his economic power to command the goods he desires through his own wealth. Such restriction also involves a restriction on the individuals' capacity to enjoy the benefits of social cooperation, the benefits created by economic interactions. It is questionable whether non-wealthy individuals would prefer to accept such restriction instead of facing the risks of instability since the restriction imposes a reduction of the benefits from social cooperation that they can enjoy. The second problem with this approach is that depending on the level of wealth of some individuals, the savings they are required to make may encompass a share of their wealth that, once incapable of being used, results in the individual falling into poverty. If an individual's wealth is not that large,
maintaining his overall level of economic power requires hindering the individual from using a considerable part of his wealth, what can create poverty for the individual. That is not an acceptable measure to tackle the issue of instability because the poor, as we argued in previous chapters, are also a destabilizing factor.

Controlling individuals' capacity to use their own wealth directly is not, we can conclude, a viable alternative due to the problems it involves. We turn now to the other option we mentioned before, i.e., addressing the destabilizing effects of inequalities in wealth through interferences with individuals shares of wealth. The idea here is that one must interfere in the distributional baseline of wealth to reallocate wealth to a different baseline in order to address such inequalities. The general goal is to interfere with a distribution and to shift it to another distribution that is less unequal and thus more stable. Although such type of measure immediately evokes the idea of redistribution, it is not associated with redistribution necessarily. The purpose here is to restore inequality to the safe margin. Whether such a task is accomplished by confiscating wealth and removing it from the system, redistributing it, or allocating it to the state, is a matter that the principle does not refer to. With all the mentioned measures, the goal can be achieved.

We can divide interferences in the baseline of wealth for containing the magnitude of wealth inequality inside the safe margin in two groups, maintenance of the safe margin or reduction of inequality. The first option involves an interference that has as a goal the maintenance of the wealth inequality inside the safe margin. The second option involves creating a distributional baseline with a gap in power that assures a higher degree of stability than the safe margin. In this case, the interference with the distributional baseline of wealth does not aim exclusively into restoring the level of inequality to acceptable margins, but also to reduce inequality in general.

In both cases, what the amendment must introduce is the possibility of interference in the distributional baseline by the sanctioning mechanism in order to tackle threats to the stability of the moral order that arise from wealth inequalities. Either for the goal of preserving the level of inequality inside the safe margin or of creating less unequal distributions, the entitlement standard that governs the distribution of wealth, the principle of natural liberty, must be constrained. Both cases involve direct interference with the distributional baseline of wealth and, consequently, with individuals' shares of wealth. If the principle of natural liberty is the sole standard of entitlement that defines distributive justice, then such interference cannot be performed without disrespecting the moral order. The role of the principle of natural liberty in determining entitlements, then, must be reduced and, consequently, the normative requirements that JNL creates, altered.

What creates stability problems is the presence of substantive wealth inequality. To solve such problem, we saw that we must interfere with the wealth of the powerful, since they are the
ones that become rebels when the balance of power is uneven. Such interference has the goal of reducing the level of inequality in economic power present in society. Concerning entitlements, then, what we must do is to alter the entitlement to wealth that those who threaten the stability of the moral order have. The powerful cannot be entitled to the share of their wealth responsible for making the level of inequality in the social scheme surpass the safe margin. It is the entitlement to this share in particular that lies at the basis of the stability problem that the principle of natural liberty creates in the case of inequality. It is because of such share that individuals acquire substantive renegotiation expectations and become rebels. In order to solve this problem, the amendment must cancel the normative requirements of the principle of natural liberty concerning such shares.\textsuperscript{304} Performing this task amounts to introducing a new standard of entitlement, one that defines that certain individuals, but not the powerful, are entitled to a share of the wealth of the powerful. Since the ultimate goal of the introduction of this new standard of entitlement is the preservation of the balance of power, we shall refer to it as the \textit{principle of balance}.

The principle of balance determines that shares of wealth that threaten the stability of the moral order must be reallocated in the system so that the wealth inequality be brought back into the safe margin. In a different manner from the principle of natural liberty, the principle of balance does not stipulate who is entitled to the wealth with clarity. As we argued, the principle of balance does not imply redistribution necessarily. The share of wealth over which the powerful lose control due to the principle of balance is not transferred to the weak necessarily. To do so is one of the means of applying the principle of balance. Nonetheless, there are other means. The entitlement could also be transferred to the state, or the wealth simply eliminated from the system. In essence, what guides the reallocation of such entitlement is how helpful the reallocation is to restore the inequality of wealth back to the safe margin. What determines whether a given person, the state, or no one is entitled to the fruits of the interference is whether the reallocation of such fruits or part of them is effective to eliminate the destabilizing inequality present.\textsuperscript{305}

There is a strong temptation here to depict the interference happening as a redistribution between the 'powerful' and the 'less powerful' members of society. Such depiction is wrong. The

\textsuperscript{304} Notice that renegotiation expectations are diminished not by giving wealth to individuals that have such expectations directly. Instead, such goal is attained by reducing the inequality in power between those with such expectations and those that are weaker than them. With a smaller inequality, their difference in power becomes such that the powerful individual cannot expect anymore that he would fare in a state of nature as well as we would before the redistribution.

\textsuperscript{305} In some sense, the principle of balance bears some structural similarity with John Rawls' difference principle. The difference principle states that social and economic inequalities are to be arranged so that they are reasonably expected to be in the advantage of every person. The idea is that inequalities are only authorized to be present in the system when their presence is beneficial for every individual there. The principle of balance, in its turn, does not refer to social inequalities, but only to wealth, or economic, inequalities. It states, in essence, that economic inequalities are to be tolerated insofar as they pose no threat to the stability of the system. See Rawls (1999), p. 53.
principle of balance removes the entitlement of the powerful over a share of their wealth. However, to whom it gives such entitlement is a question that the principle in its pure abstract form leaves open. Such entitlement is to be distributed with the goal of reducing the level of inequality in the social system. Such a goal can only be attained on an empirical level, in which we evaluate the actual distribution of wealth and the actual levels of inequalities in the system. In a particular society, it may be the case that the interference demanded by the principle does not contemplate the worst-off members of society. For this reason, we see the need for suggesting an additional principle, the principle of charity, which we examine in the next section.

The principle of balance takes precedence over the principle of natural liberty in determining entitlements to wealth with regards to a share of the wealth of the powerful. However, the relation between the two principles becomes more complicated when we consider the ultimate goal of the principle of balance. If the sole purpose of the principle of balance is to restore inequality to acceptable levels (a restrictive interpretation), then the relation is clear. The principle of natural liberty determines entitlements to wealth whenever levels of inequality are within the acceptable margin and the principle of balance gains priority when such levels extrapolate the margin. However, we can also conceive that the principle of balance has a more ambitious goal. It can be interpreted to include a goal of actively pursuing the reduction of inequality, even when its level remains inside the safe margin (an egalitarian interpretation). In this case, the fact that inequalities rest within the safe margin is not sufficient to guarantee that the principle of natural liberty will always determine individuals' entitlements to wealth. To actively pursue a reduction of wealth inequalities, the moral order must interfere with the distribution of wealth even if the level of inequality is inside the safe margin. Nonetheless, in both interpretations, the demands for interference with the distributional baseline of the principle of balance are more substantive than those produced by the principle of natural liberty. They involve requiring some individuals to transfer part of their wealth to others, something that the principle of natural liberty cannot endorse. On the other side, under the egalitarian interpretation, the principle affects the entitlement of the powerful over a larger share of their wealth. The idea here is to decrease the gap in power present in society continuously. For this purpose, it is necessary to extract a considerable amount of wealth from the powerful.

We do not pursue here thoroughly this interpretative question. The general goal of our argument is to show that contractarianism can support a conception of distributive justice that the principle of natural liberty can endorse transfers of wealth from the powerful to the less powerful insofar as the powerful have acquired goods from the less powerful without observing the rules of private property. In this case, the less powerful are entitled to certain goods held by the powerful and thus the powerful are under the normative requirement of transferring such goods to the less powerful. However, that is a case that pertains to corrective justice, and not distributive justice, as we discussing here.
produces substantive redistributive requirements. The principle of balance is standard of entitlement that in both of its interpretations serves to ground such a conception. Either if under the restrictive interpretation or the egalitarian interpretation, the principle can create substantive redistributive demands. In both cases, the principle entitles someone to a share of the wealth of the powerful. Consequently, the powerful find themselves under the normative requirement of transferring part of their wealth to others. Nonetheless, we believe that there are some empirical difficulties associated with one of the interpretations that could indicate that the other should be favored.

The more restrictive interpretation of the principle of balance is that in which the principle demands the maintenance of the level of inequality in wealth around the safe margin. This interpretation is problematic. It is unclear whether it is even possible to calculate such a margin. Individuals know who are the powerful members of their society, who are the weak members of their societies, and they also have some grasp of how unequal is their society concerning wealth distribution. Despite that, they are incapable of having precise knowledge regarding the power dynamics that operate in their society. They know that if the gap in economic power grows too large, society will become unstable. But they are utterly ignorant about how much must the inequality grow in order for problems to appear. The restrictive interpretation seems to require precise knowledge of the power dynamics in society in order for the principle of balance to fulfill its goal appropriately. Given individuals' lack of precise knowledge in this respect, it seems that the principle would not be able to function correctly and it would, therefore, fail to fulfill its goal.

The egalitarian interpretation of the principle of balance seems to avoid the problem we considered above. Whereas the restrictive interpretation demands only 'surgical' interventions to maintain the level of inequality somewhat near the safe margin, the egalitarian interpretation considers the reduction of inequality as a constant effort. Under the restrictive interpretation, the priority of the principle of balance over that of natural liberty ceases when the safe margin is achieved. The egalitarian interpretation maintains such a priority indefinitely. Such an approach seems more efficacious than that of the restrictive interpretation because it does not require individuals to have accurate knowledge of the dynamics of power in society. It does not matter which is the exact safe margin of inequality for the egalitarian interpretation since redistributive efforts are carried out even if the levels of inequality are within the safe margin. Hence, if individuals are committed to preserving the stability of their society, it seems to be the case that the adoption of the principle of balance under the egalitarian interpretation is more suitable for them.

307 That the weak are entitled to a share of the wealth of the powerful does not mean that they can take away such a share whenever they see fit. The redistributive demand that entails from the entitlement of the weak does not give them a right to seize the part of the wealth of the powerful. The procedure through which the normative demand is to be processed is also defined by the terms of cooperation or the legal framework of the social scheme.
With such an interpretation, they avoid the practical problem to which the restrictive interpretation is subject to.

Notice, nonetheless, that the preference for the egalitarian interpretation over the restrictive interpretation does not arise nor imply a commitment to equality of wealth. What stands at the center of the principle of balance is a commitment to stability. It is because less unequal distributions of wealth offer fewer dangers to the stability of the moral order than more unequal distributions that we can justify the principle. Even under the egalitarian interpretation, the goal of the principle of balance is not to create perfect equality of wealth, but simply to strive for less unequal distributions, and thus to more stable distributions.

The rationale for the principle of balance should be evident given the discussion regarding stability and social cooperation that we performed at the start of the chapter and also our discussion in the last chapter. It is only when the moral order can maintain inequality inside the safe margin that it can preserve its stability. The preservation of stability is something that both the powerful and the weak have strong reasons to desire. The advantages they acquire from both moral and social cooperation can only be sustained and be considered assured when the moral order is stable. Otherwise, the enjoyment of such benefits is harmed. For the more unstable the moral order is, the more uncertain is the enjoyment of the benefits of cooperation. Our discussion in chapter five indicated that preserving the stability of the moral order is only possible if the moral order is capable of interfering in the distribution of wealth. Otherwise, shifts in the distribution are bound to create inequalities that create rebels with substantive renegotiation expectations and who act to destabilize the moral order. If individuals want to preserve the stability of their moral order, as we assume they do in virtue of their general interest in stability, then they must introduce the principle of balance in their moral order.

Notice, however, that the principle of natural liberty must not be wholly abandoned. Although the principle of natural liberty is responsible for the instability of JNL, we do not need to remove it entirely from the moral order. As we argue, instability does not proceed from inequality per se, but instead from inequality that surpasses the safe margin. The principle reveals itself problematic when the safe margin is surpassed and the moral order is prohibited from interfering with the distribution of wealth to ensure its own stability. Despite this problem, the principle of natural liberty remains a useful principle for the entire social system when the inequalities in power remain inside the safe margin. The reason for that is that as a principle that determines how goods are allocated, the principle of natural liberty is likely to lead to efficient or Pareto optimal allocations. As we argue in chapter two, the principle of natural liberty conducts the social scheme.
to what Rawls has called a system of natural liberty. In the system of natural liberty, goods and careers are open to individuals' talents and skills. The distribution of talents and skills governs the emergence of distributions in the system. Insofar as the distributions of wealth remains inside the safe margin of inequality, the efficiency of the principle of natural liberty remains useful for individuals. Nonetheless, when shifts in the distribution of wealth threaten the stability of the moral order, the principle of balance must take the priority in differentiating just from unjust distributions.

One could argue against our approach that the principle of balance – especially under the egalitarian interpretation – is subject to what Derek Parfit has called the *Leveling Down Objection*. Derek Parfit advances this critique against some forms of egalitarianism. The critique is that if one deems inequality to be something bad, and thus that its removal is something good, then one must endorse changes in the system that make some individuals worse off than they are with the purpose of promoting equality in the system, even if doing so makes no individual in the system better off. Consider this example. Assume that an equal distribution of candies for children is something morally praiseworthy. Consider now that we have a system with four children, one of which has a candy, and the others have none. That is an unequal distribution of candies and, consequently, it stands against our principle of equal distribution of candies between children. A normative requirement of the principle would be that we must take away the only candy in the system. In this manner, equality is restored to the system. However, equality is restored by making one child worse off, the child who had the candy, despite the fact no other child is made better-off. Nonetheless, according to the principle, a distribution in which no child has a candy is morally superior to the one in which just one has the candy. Parfit deems such a conclusion absurd, and we agree with him.

The principle of balance could be targeted by Parfit's objection because it can endorse measures that involve making individuals worse off exclusively with the purpose of reducing inequality. It is true that, concerning wealth, the principle of balance would endorse a measure to, for example, confiscate part of the wealth of a powerful individual and instead of redistributing the wealth, simply eliminating it from the system to reduce the level of inequality. In this sense, the principle could endorse making the powerful worse off with regards to wealth without making anyone in the system better off regarding wealth, since no redistribution must necessarily take place. That is a form of leveling down. However, to criticize the principle on such grounds seems, in our view, to miss the point of the principle. The principle of balance is not an egalitarian principle in essence. The principle's hostility toward inequality does not arise from a commitment to equality.

309 See Parfit (1997), pp. 210-211.
Instead, the goal of the principle is stability. It serves to protect and maintain stability in the moral order. Concerning what the principle is committed to, the leveling down objection cannot be raised. By interfering with the distribution of wealth in the system, the principle is thought to improve the standing of every individual in the system. Every individual is supposed to have an interest in stability. The interferences that the principle of balance endorses are such that the interest in stability that every individual has is promoted, and thus advantages are created for everyone, even for the powerful individual who is made worse off in terms of wealth by the interference.

The principle of balance is a standard of entitlement that entitles certain individuals to a share of the wealth of the powerful. Determining who the individuals that receive such an entitlement are is a task that can be answered only on the empirical level. Such individuals are not necessarily the worst-off individuals in the social scheme, as one could expect. They could be contemplated by the principle of balance, if the empirical circumstances make it so that the worst-off being entitled to a share of the wealth of the powerful helps to diminish the level of inequality in the social scheme. In any case, the principle of balance allows the sanctioning mechanism to interfere in the distributional baseline of wealth, what, in its turn, enables the possibility of redistribution. The principle composes the first part of the amendment to the moral agreement. Since a new standard of entitlement is included in the moral order, JNL is necessarily altered to become JCNL. Distributive justice is not defined by the principle of natural liberty exclusively, but also for the principle of balance. Such a principle composes the first part of the amendment. The second part contains the principle of charity, which we discuss next.

3.2 The Principle of Charity

We move now to examine the amendment to the moral agreement in the context of the second source of instability for the moral order, the appearance of poor individuals. Poor individuals behave, as we have argued, as rebels. They regard the moral order as a hindrance to the promotion of their fundamental general interest in freedom from poverty. In the context of poverty, we saw in chapter five that what the moral order must do to contain the destabilizing effects of the poor over itself is to transfer them the goods they require to escape from poverty. However, in a moral order under JNL, such transfer is forbidden, since it interferes with the distribution of wealth in the social scheme. Hence, to ensure stability in the face of the threats posed by the poor, the amendment must introduce a new principle, a new standard of entitlement, to constrain the principle of natural liberty
and assure the stability of the moral order. We shall refer to this principle as the *principle of charity*.

The principle of charity we suggest is an elaboration of an intuition from Hobbes. In a passage of *Leviathan*, he argues that he who attempts to hold for himself those things that are superfluous for him, but necessary for others, should be either left outside of society or driven out of society. In Hobbes' view, such individual is guilty of creating a conflict with the needy, those that are in dire need of what is superfluous to him. He is a source of instability for the social order. The attitude of the poor is justifiable insofar as they seek goods out of natural necessity, i.e., irrespectively of the context in which they find themselves, they will always act in the same manner. In this sense, to locate the problem of stability that poverty creates on the poor's non-compliant attitude is to create a problem that cannot be solved. To address the problem of stability that springs from poverty, it is necessary, then, to focus on the non-poor refusal of helping the poor with goods that are superfluous for them. In another passage of *Leviathan*, Hobbes approaches the matter of public charity. He argues that the state must take charge of assisting to those who become unable to sustain themselves. Hobbes himself does not provide a justification for the inclusion of such provisions among the laws that govern society. Gregory Kavka, however, in a commentary to Hobbes' views on economic welfare, suggests two general reasons that individuals in a Hobbesian state of nature would have to include provisions of economic welfare in the social contract. The first reason is that the inclusion of such provisions works as a form of insurance for the whole society. In this sense, the non-poor can be said to have some gains with its inclusion because it provides them with an assurance that they will never fall into complete destitution. The second reason is the reason that we explored in our discussion since chapter five, instability. The poor are, as we have argued before, a strong source of instability. If there are no reliable means available for them to escape poverty, they turn against the moral order in an attempt to obtain the goods they desperately need. The inclusion of provisions for economic welfare is an efficient means to curtail this source of instability. The principle of charity that we introduce in this section is a principle that enables the inclusion of such provisions.

One could argue that the principle of balance is already sufficient to handle cases where poverty is a threat to the stability of moral cooperation. Such a belief is false. Although it is true that the principle of balance might be apt to solve problems of stability posed by the poor in *some* cases, that is not necessarily so. We can distinguish the two principles in four different aspects.

The first difference concerns with which entitlements under the principle of natural liberty

---

each principle interferes. The principle of balance affects the entitlement of the powerful over a share of their wealth. It constrains the principle of natural liberty by entitling other individuals to this share of wealth, instead of the powerful. That is not the case with the principle of charity. The principle of charity affects the entitlements over wealth of every non-poor individual in the system. In the category of 'non-poor' individuals, we have a much wider range of individuals than in the category of 'powerful' individuals.

The second difference refers to which individuals each principle attempts to contain. In the case of the principle of balance, it attempts to contain the economically powerful individuals in society. These are the individuals who can become rebels by reason of their economic power. The principle of charity, in its turn, is not concerned with economic power in any sense that is relevant for the principle of balance. The principle of charity is directed at containing the destabilizing effects that the poor have over the moral order. The poor do not rebel because they are economically powerful, but because they are unable to fulfill their basic needs inside the moral order. It is true that power is an important factor in the rationale of the poor to rebel. Likewise the economically powerful individual, the poor believe that they are powerful enough to assure for themselves a better outcome in the state of nature in comparison to how they are in the status quo. However, the power that leads the poor to rebel is not economic power, but other forms of power such as physical power or power obtained through a coalition of poor individuals.

The third difference refers to the goal each principle has. The principle of balance aims at containing economically powerful individuals from destabilizing society. For this end, it must maintain the level of inequality in wealth inside what he has termed a safe margin. Insofar as the safe margin is not surpassed, no powerful individuals are inclined to rebel against the moral order. The principle of charity, on the other hand, is not concerned with inequality in any sense. It is poverty, i.e., the reduction of poverty, that is the goal of the principle of charity. Whether the level of inequality in a society is inside the safe margin or not is not a relevant factor for the operation of the principle of charity.

A fourth difference concerns redistribution. As we have argued, the principle of balance is not committed with redistribution necessarily. Instead, it enables the sanctioning mechanism to interfere with the distributional baseline of wealth to control the levels of inequality. The exact measure the mechanism adopts to perform such control is determined by each society. Redistribution from the powerful to the weak is a viable alternative for this goal, but so are the allocation of part of the wealth of the powerful to the sanctioning mechanism itself or the removal of such wealth from the system. The principle of charity, on the other hand, is connected with redistribution necessarily. It is a principle to contain the harm the poor cause over the stability of the
moral order. For this purpose, it must necessarily provide the poor the goods they require to escape poverty. Such a result can only be achieved through redistribution.

In light of the four differences, we can conceive cases in which the principle of balance is not applicable, but the principle of charity is. Consider a society in which the level of inequality in power is much below the safe margin. In such a society, the powerful are very far from having a reason to rebel. They are satisfied with the status quo. In such circumstances, the principle of balance does not operate, i.e., the principle of natural liberty takes priority over the principle of balance in every case. The entitlement of the most powerful over their wealth is not affected by the principle of balance here since they pose no risk to the stability of the moral order. Despite the non-threatening levels of inequality present in such a society, it may contain a considerable number of poor individuals. And such individuals may themselves offer a risk to the stability of the moral order. The threat they offer, however, is not covered by the principle of balance, since it is not a threat that arises from the economically powerful individuals and from the presence of substantive inequality in wealth. Dealing with such a threat requires the introduction of a new principle, i.e., a new standard of entitlement. That is the principle of charity.

Before considering the entitlement standard that the amendment must introduce to solve the problem of instability that poverty creates, we must recall an important aspect of the matter of stability connected with poverty. Poverty can appear in the moral order in two different manners, either through artificial scarcity or natural scarcity. The distinction is important because the measures that can be employed to overcome the instability that such events create are necessarily different. Redistributive measures can address the instability created by artificial scarcity. However, such measures are not suitable in the case of natural scarcity. Natural scarcity refers to the scenario in which the number of goods available in the system is such that no matter how the goods are organized there always is a considerable number of poor individuals. In this sense, it does not matter which redistribute measure is introduced, no amount of redistributing can ensure the stability of the moral order. Due to this difference, we consider the two scenarios in different moments. We first address the case of artificial scarcity and then consider briefly that of natural scarcity.

### 3.2.1 Artificial Scarcity and the Principle of Charity

To contain the effects of the poor over the moral order, there is but one alternative, namely to transfer to the poor the wealth they require to be free from poverty. Insofar as they remain poor, they regard the moral order as unjustified and thus behave as rebels. What the amendment must
achieve, then, is to introduce a change that allows for the moral order to create normative requirements for individuals to transfer wealth to the poor. Concerning entitlement, such change amounts to canceling non-poor individuals' entitlements over part of their own wealth and entitling the poor to such share of the wealth. That is the task that the principle of charity fulfills. The principle of charity states that poverty is a factor that determines entitlement to wealth. Consequently, the principle of charity conflicts with the principle of natural liberty. For the principle of charity, individuals' basic needs entitle them to wealth that was acquired or produced by other individuals in conformity with the rules of private property. Consequently, the principle of charity gives the poor entitlement over wealth that is under the possession of individuals who are already entitled to such wealth under the principle of natural liberty.

To allow for redistribution, the principle of charity must take precedence over the principle of natural liberty in at least some circumstances. Despite the importance of the principle of charity, however, its priority over the principle of natural liberty cannot be absolute. If that is the case, entitlements are defined first by needs and then by acquisition in observance of private property. That is the case of a principle of justice that we can extract from Saint Thomas of Aquinas' discussion of almsgiving. Saint Thomas suggests a principle to govern the normativity of almsgiving that can be formulated as follows: *When disposing of one's own resources, one should do it in the following serial order: First to fulfill one's necessities; second to fulfill the necessities of those under one's guard; and third to fulfill the needs of the poor.* The principle demands that alms be given only with resources that are not essential for the fulfillment of one's needs and of those under one's protection, i.e., resources that can be said to superfluous. If the resources available are only sufficient to fulfill one's own needs and of those under one's protection, then the principle entitles the owner of such goods to them. Saint Thomas view is that the principle of charity has priority over the principle of natural liberty. It is just when there are no poor individuals in the moral order that the principle of natural liberty comes into place. Otherwise, if there is someone whose basic needs are not fulfilled, every resource or good available ought to be transferred for this person until her basic needs are fulfilled.

An interpretation of the relation between natural liberty and charity that goes in the direction of what Saint Thomas argues is, in our view, tough to sustain from a contractarian perspective. The normative requirements that such priority generate are too demanding and too strong for the non-

---

314 As it is the case with the principle of balance, the fact that the poor are entitled to a share of the wealth of another individual does imply that the poor can simply take away the wealth they are entitled to. The transference, we imagine, must fulfill certain procedural requirements.

315 See Saint Thomas Aquinas (1975), pp. 251-253 (ST II-II. Q 32. A 5 co.)

316 The discussion also assumes that one's own needs have priorities over those of other persons, as well as the needs of those under one's guard.
poor. The priority determines that poverty must be eradicated from the moral order before any individual can acquire an entitlement to a particular superfluous good or resource through the principle of natural liberty. The problem with such strong normative requirements is that the stability of the moral order does not require the complete eradication of poverty. What causes a problem for the stability of cooperation is widespread poverty. Poverty can subsist in the system without affecting its stability. Since the central rationale for the principle of charity is solving the problem of stability, and that the eradication of poverty does not constitute a requirement for solving such a problem, it is reasonable to assume that accepting the principle is irrational when it has absolute priority over the principle of natural liberty.

The priority of the principle of charity must be restricted to particular cases. Otherwise, its inclusion in the moral order becomes irrational, as we argued above. However, to define such cases is a challenging task. Although it is clear that the introduction of the principle must be rational in some circumstances, the actual circumstances vary strongly following empirical and historical contingencies that determine society. Societies with different levels of affluence will be differently welcoming to the priority of charity depending on the circumstances to which it applies. Poorer societies will, we believe, tend to be more open to prioritizing charity over natural liberty since it is easier for poverty to become widespread in them. More affluent societies, on the other hand, might be less open to prioritize the principle of charity priority over the principle of natural liberty in a wide range of circumstances. The reason for that would be that due to its level of affluence, it is harder for poverty to become widespread. In any case, we prefer to leave open the exact relation that the principle of charity and the principle of natural liberty have in the moral order. It is clear that the introduction of the principle of charity is rational and must be done. The principle of charity must be introduced in order to assure the stability of the moral order, and thus to promote the interest in stability that rational agents have. Although the principle of balance may be able to contain the destabilizing effects of the poor in some measure, there are situations in which the principle of balance is unable to require transfers for the poor. The principle of charity is necessary to ensure that the destabilizing effects of the poor will be under control. The incorporation of the principle of charity into the moral order composes the second part of the amendment, and thus the reformulation of JNL into JCNL.

3.2.2 A NOTE ON JUSTICE AND NATURAL SCARCITY

We move now to examine the case of natural scarcity. Our examination of this topic is brief since it
is much more complex and distinct from the others we examined before. The stability problem that arises from natural scarcity is different from that produced by artificial scarcity. Natural scarcity is about a lack of resources in the system and not of about bad allocations of resources. With natural scarcity a considerable number of individuals becomes poor, and there is no manner to lift them out of poverty and restore the stability of the moral order. The lack of goods caused by natural scarcity blocks the possibility of any form of redistribution to restore stability. In order to handle the threat to stability that such form of scarcity poses, the moral order must necessarily introduce measures before it occurs. In a different manner from inequalities in power or artificial scarcity, the destabilizing effects that natural scarcity has cannot be contained after they appear. It is a scheme of prevention that must be installed.

Preventing artificial scarcity is, in essence, about preventing the occurrence of severe environmental disasters. The cause of this sort of scarcity is a drastic environmental alteration that results in a drastic reduction of goods and resources available. But it is not because natural scarcity is connected with the environment that its causal sources are independent of human action. Natural scarcity can also be a product of human action. The paradigmatic cases of natural scarcity are those of the drought or the flood. Nonetheless, individuals themselves can create scenarios of natural scarcity by means of situations such as the tragedy of the commons. The tragedy depicts a case of overconsumption of fundamental goods that leads to scenarios of natural scarcity. Overconsumption depicts a case in which individuals consume a specific good at a pace that is faster than the rate of renewal or regeneration of the good. If overconsumption targets a good fundamental for social existence, such as water, air, or food sources, society is setting itself in course for the appearance of natural scarcity.

The investigation of principles of justice suitable to address cases of natural scarcity drives us to matters utterly different from those we have been examining so far. Such discussions refer much more to how individuals should interact with their environment than to how they should interact with each other. The threat to stability here comes more from the natural world than from the social world. Such analysis is far beyond the scope of our present inquiry. For this reason, we do not pursue it here.

4 CONCLUSION OF THE CHAPTER

317 See Hardin (1968).
We come now to the conclusion of the chapter, and of our whole argument. We have argued that JNL is an unstable conception of justice. The source of its instability is its incapacity to endorse interferences with individuals' shares of goods. Without such interferences, we argued, there is no way to tackle the destabilizing effects that rebels have over the moral order. A *modus vivendi* requires a moral order that is justified for everyone and stable. Stability refers to the likelihood that individuals will remain enjoying the benefits of social cooperation which are allowed by the moral framework through the future. We saw that JNL is unfit to, by itself, sustain a *modus vivendi*. In this chapter, however, we amended JNL with the introduction of two principles to solve its problem of stability, the principle of balance and the principle of charity. The amendment turns JNL into JCNL. Both principles are standards of entitlement. They define the entitlement that individuals have over particular external goods. In the case of the principle of balance, it works to maintain the inequality in economic power within levels that do not threaten the stability of the moral order. It does so by affecting the entitlement of the powerful over a share of their own wealth and reallocating such entitlement – and the wealth along with it – with the purpose of reducing the level of inequality to non-dangerous levels. In this manner, the principle of balance creates a normative requirement of redistributing economic power from the powerful to other individuals. Such measure, we argued, solves the problem of stability that JNL has due to individuals with considerable renegotiation expectations by reason of massive gaps in power. The principle of charity, on the other hand, introduces poverty as an element of entitlement. It determines that to be poor entitles individuals in some measure to wealth that they do not possess. A delicate matter, however, is how this principle is supposed to relate with the principle of natural liberty. To determine such a relationship is difficult and we left the question open, for one main reason. Such a relationship is best determined in light of the social and historical circumstances that each society has. What our argument concludes, nonetheless, is that every society must contain a principle of charity that gives poor individuals entitlement to wealth simply by the fact that they are poor. With the balance principle and principle of charity amending the moral agreement that included JNL, JNL is transformed into JCNL, and the contract becomes able to sustain a *modus vivendi*. 
We conclude now our work with a few remarks. In chapter two, we suggested that the part of the social contract tradition called contractarianism faces a particular problem. It faces considerable difficulties to approach the matter of distributive justice. Contractarianism is capable of justifying moral norms insofar as they are mutually beneficial for the individuals under the authority of the norm. And rules of distributive justice with substantive redistributive demands are often one-sided with regards the benefits they produce. They are rules that favor a particular group – those that receive the fruits of redistribution – at the expenses of another group – those who lose goods and resources in the redistributive effort. Due to this asymmetry, contractarianism faces difficulties to justify such rules.

Despite the difficulty, however, contractarianism can justify a specific view of distributive justice. That is the view of justice as natural liberty. Contractarianism manages to justify such a conception because it is highly minimalistic regarding redistributive demands. It is a conception of justice that strongly resembles Robert Nozick's libertarian view of justice. According to JNL, a distribution of goods is just if every share involved in the distribution was acquired in accordance with the rules of private property. We deemed JNL minimalistic because it has virtually no redistributive requirements. Since the standard of entitlement that underlies the conception affirms that entitlement is defined by acquiring a good under the rules of private property, JNL demands redistribution only in cases where the rules of private property are violated. Otherwise, individuals ought not to interfere with the goods of one another.

The problem with such minimalism regarding redistribution is that it stands against widespread intuitions about morality and our own daily experience. We not only often believe that some form of substantive redistribution of goods is morally mandatory, but we also live – and have lived throughout history – in societies that include such moral requirements. Contractarianism's
apparent failure to justify a conception of distributive justice with redistributive requirements more substantive than JNL seems to indicate that it is a theory that cannot fully account for human moral intuitions and moral experiences. Our task in the four chapters that followed chapter two was to argue for a more substantive conception of justice from a contractarian perspective, and thus to show that contractarianism can account for our common intuitions and experiences regarding the morality of redistribution.

Our argument consists in demonstrating that a robust conception of distributive justice is connected with the stability of the moral order. Stability, we argued, can be understood in two different senses. There is the moral conception of stability, which is endorsed by John Rawls, and also the rational conception of stability. We favored the latter over the former. Endorsing moral stability require assumptions that we cannot have due to our contractarian viewpoint. Rational stability, its turn, suits the Hobbesian outlook perfectly. For such a conception, a moral order is stable insofar as it can sustain a modus vivendi for those that live under its normative authority. A modus vivendi is sustained insofar as the moral order can elicit the compliance of every individual in it. In order for the moral order to achieve this goal, it must be the case that individuals find it rational to comply with the moral order. We saw that non-compliance can appear from two different sources, namely cheating and rebellion. The cheater is the individual whom Hobbes depicts as the Fool and Hume as the Sensible Knave. He violates the moral order in the expectation of an immediate benefit. But despite that, the cheater remains committed with the moral order, i.e., he still regards the moral order as justified. The rebel, on the other hand, is the individual who violates the moral order because he considers it unjustified. His objective is not the acquisition of an immediate benefit with the violation, but instead the destabilization of the moral order. The rebel seeks to renegotiate the moral agreement and, as we argued in chapter four, it is highly likely that the other individuals in the moral order will refuse to renegotiate, since doing so necessarily results in the worsening of their condition. Given their refusal, the individual who wants to renegotiate the terms of cooperation becomes a rebel. He starts to be not only non-compliant with the moral order but also to actively work to undermine it.

The rebel, we have argued, indicates a rationale toward non-compliance that the contractarian tradition has overlooked. When they approach the problem of stability, contractarians often have in mind the cheater. For this reason, they discuss the matter of stability as the matter of introducing appropriate sanctions to contain non-compliant behavior. The problem with such approach is that although it is efficient to contain the cheater, it fails to do so for the rebel. When we consider the sanctions and the sanctioning mechanism from a less formal and more realistic perspective, we observe that the mechanism is inherently inefficient regarding its sanctioning
activity. It always leaves some margin for individuals to be non-compliant. When we have the cheater in mind, the inefficiency of the mechanism is not that big of a problem. The cheater does not want the breakdown of the moral order or a substantive increase in its instability. Since he regards it as justified, his non-compliant attitude will never go as far as causing serious instability problems for the moral order. However, once we take into account that not only cheaters, but also rebels have a margin to be non-compliant, it is possible to observe that reducing the matter of stability to the choice of sanctions is wrong. Rebels want to destabilize the moral order. And if the sanctioning mechanism leaves them room to do so, the problem of stability is not solved, since they remain operating to decrease the level of instability of the moral order, and the sanctioning mechanism cannot hinder them from doing so. If contractarians want to tackle the problem of stability, they must go beyond the matter of sanctions. The rebel indicates that this move is necessary. Otherwise, the contractarian method cannot ensure the stability of the moral orders it justifies. Tackling the issue of stability involves, primarily, dealing with the challenge that the rebel offers.

Rebels appear in the social system due to changes in what we have called the circumstances of moral justification. Such circumstances are the result of a contractarian interpretation of David Hume's idea of the circumstances of justice that we suggest. The circumstances of moral justification indicate the conditions that must in place for the justification of moral rules for the whole social system. It is only when such conditions obtain that moral justification is possible. We have suggested three critical circumstances: unsocial sociability, the social balance of power, and the absence of widespread poverty. It is mainly with respect to alterations in these three circumstances that rebels appear in the social system. We examined how rebels appear in the context of two circumstances, the social balance of power and the absence of widespread poverty. In the case of the social balance of power, rebels appear when the distribution of power in the social scheme is considerably unequal. When considerable inequalities in power appear, powerful individuals start to regard the moral order more as a hindrance to their interests than a help to their promotion. Such perception leads them to behave as rebels. Poverty, on the other side, is also a factor that drives individuals to become rebels. Poor individuals are those unable to command the goods they require to fulfill their basic needs. In such a condition, individuals lose immediately every reason to comply with the moral order, for the order starts to work as a hindrance to the promotion of the fundamental general interest in freedom from poverty that they have.

Tackling the issue of stability involves considering how to prevent or deal with alterations in circumstances of moral justification that we considered above. We performed this analysis in chapter five. Our analysis involved an examination of JNL in particular. We examined how stable is
JNL by considering JNL’s relationship with the circumstances of moral justification. JNL is a conception of justice that is very similar to one of the doctrines that Hobbes lists as those that weaken society and lead to its dissolution, namely an absolute propriety over one's own goods. Such a conception of justice forbids the sanctioning mechanism from interfering with individuals properties in almost every case. Chapter five's analysis is an elaboration of Hobbes argument. We verify that JNL, by reason of its minimalism regarding redistributive demands, is unstable for two reasons. The first reason is that JNL allows for changes in the circumstances of moral justification to happen. JNL's normative requirements thwart in no sense the appearance of substantive inequalities in power in the social scheme, nor does it hinder the appearance of considerable numbers of poor individuals. In other words, JNL is unable to prevent the emergence of rebels in the social system. The second reason is that after rebels have appeared in the system, JNL is not only incapable of handling the destabilizing effects that they produce, but it also hinders the use of the measures that would be efficient to contain the rebels. Here JNL's similarity with the doctrine that Hobbes' criticizes is important. JNL forbids the moral order and the sanctioning mechanism from interfering with the distribution of goods in place in society. From this prohibition, it follows that no redistributive measures can be employed to contain the rebels. And such measures, we have argued, are essential to ensuring stability against rebels.

What our examination of JNL's stability indicates is that the issue of stability, in contractarianism, must involve more than a discussion of sanctions. It must involve a consideration of the actual moral rules that compose the moral order under analysis. JNL's failure indicates that dealing with rebels requires the presence of particular moral rules that enable the use of specific measures against the rebels. In chapter five, we conclude that stability can only be assured if the sanctioning mechanism is allowed to interfere with the distributional baseline of goods. If it is thwarted from doing so by a moral principle, such as JNL does, the stability of the moral order cannot be assured. The argument we develop in chapter five is a negative argument. It demonstrates that a moral order that cannot interfere with the distribution of goods is necessarily unstable. It shows, in particular, that the moral order must be able to perform redistribution of goods to assure its own stability. Nonetheless, we do not discuss in chapter five which principles should guide the redistributive effort required to ensure the stability of the moral order. We develop such principles in chapter six.

We argue that there two principles that must be included in every moral order to ensure that they are stable, namely the principle of balance and the principle of charity. Both principles are standards of entitlement in the same manner that the principle of natural liberty is, they define what entitles individuals to external goods. The introduction of these two principles serves to constrain
the principle of natural liberty in some cases and allow for the sanctioning mechanism to interfere with the distribution of goods, which is something that the principle of natural liberty does not allow. They do not, however, supplant the principle of natural liberty. It is only with respect to particular cases that the two principles take priority over the principle of natural liberty to define individuals' entitlements. Conjoined with JNL, these two principles turn the conception of justice into JCNL, i.e., justice as constrained natural liberty.

The principle of balance is introduced to contain the effects of individuals who become rebels due to substantive inequalities in power. It is a principle that demands the sanctioning mechanism to interfere in the distribution of goods to maintain the level of inequality in power inside a margin that does not prompt the appearance of rebels. It does so by removing the entitlement of the powerful over part of their goods and conferring such entitlement either to other individuals or to the sanctioning mechanism itself. The principle, however, does not stipulates who becomes entitled to such goods. Such entitlement is conferred insofar as the new allocation of the entitlement promotes the stability of the social scheme. Determining which is such allocation is, however, an empirical matter that each society must decide for itself. By removing part of the goods of the powerful, and consequently part of their power, the principle of balance assures that they never grow powerful enough to become threats to the stability of the moral order.

The principle of charity is introduced to contain the destabilizing effects that poverty has over the social order. Different from the principle of balance, the principle of charity is necessarily a redistributive principle. Whereas the principle of balance merely refers to altering how power is allocated in the social scheme to an allocation that does not destabilize the social order, the principle of charity refers directly to redistribution. It is a principle that confers to the poor entitlement over part of the wealth of the non-poor. In this manner, it assures that the sanctioning mechanism can interfere with the societal distribution of goods to redistribute from the non-poor to the poor they goods they need to be free from poverty. When the poor are delivered the goods they require, their rationale to rebel disappears immediately, and the problem of stability they pose is, thus, solved.

There is a tough question concerning the two principles and their relationship with the principle of natural liberty that we leave open in our discussion. It is the matter of in which situations the principles have priority over the principle of natural liberty. The broader the range of situations, the more frequent interferences with the societal distribution of goods is to happen. We prefer to leave this question open because the relationship between the principles is best determined at the empirical level. Although every society requires the two principles in some form in order for their moral orders to be stable, the exact relationships between the principles and the principle of natural liberty will vary from society to society. The variation follows the exact empirical
circumstances in each society. In some societies, it may be necessary that the principles have priority over natural liberty in a wide range of situations, whereas in others societies, it may be the case that priority only in a small range over situations can suffice to guarantee the stability of the moral order.

In matters of justification, we refer to two factors to show that rational agents would accept the principles. The first factor refers to the risk of rebellion, i.e., of rebels overturning the moral order and imposing terms of cooperation that they prefer. Without the inclusion of a conception of justice that requires substantive redistribution of goods, the sole method that the moral order has of constraining rebels is the sanctioning mechanism. However, the mechanism is not wholly efficient for this purpose. Rebels always have a margin to be non-compliant and to undermine the stability of the moral order. Consequently, individuals should expect that, in time, the rebels will manage to either undermine the system of sanctions or overpower it. To prevent the occurrence of such a drastic change in the cooperative framework, it is rational for individuals to endorse transforming JNL into JCNL. The second factor is that the presence of rebels in the social system harms the prospects for social cooperation drastically. Rebels reduce the level of stability of society. The more unstable a society is, the more individuals there are who desire to exploit the others. Instability indicates that individuals will be uncertain, and thus skeptical, about the intentions of others in cooperating with them. Either because the number of individuals who want to double-cross the others increases, or because individuals start to refuse to cooperate out of fear that they will be exploited, the outcome is that the profits from social cooperation available in the cooperative framework are reduced. Given these two factors, we can infer that the stability of the moral order is something that rational agents will be interested in. Referring to this interest, we can justify JCNL along with its the two new standards of entitlement, the principle of balance and the principle of charity.

In sum, our argument appeals to an interest in the stability of the moral order to justify a conception of distributive justice that creates substantive redistributive requirements. Insofar as such the conception is a crucial factor for the stability of the moral order, individuals must include it in their moral order. Otherwise, they cannot ensure the stability of their moral order. What the conception of distributive justice must include is the two standards of entitlement that we considered before, the principle of balance and the principle of charity. Both principles enable the sanctioning mechanism to interfere with the societal baseline of goods and to apply the necessary measures to contain the destabilizing effects that rebels have over the moral order. In this manner, the problem of stability posed by the rebels is solved, and it is demonstrated that the Hobbesian method of moral justification can justify substantive conceptions of distributive justice.


_____ Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht (AA): Akademie Ausgabe, Band VIII, 1794.

_____ Die Metaphysik der Sitten (AA): Akademie Ausgabe, Band VI, 1797.


